Multivariate analysis to discriminate top quark pair production channels at LHC

Relatore:
Prof. Maximiliano Sioli

Presentata da:
Elena Del Gratta

Correlatori:
Prof. Rita Fioresi
Dott. Claudio Severi

"The only truly real things are the clicks in detectors."

Abstract

Il quark top è una delle particelle fondamentali del Modello Standard, ed è osservato a LHC nelle collisioni a più elevata energia. In particolare, la coppia top-antitop ($t\bar{t}$) è prodotta tramite interazione forte da eventi gluone-gluone (gg) oppure collisioni di quark e antiquark ($q\bar{q}$). I diversi meccanismi di produzione portano ad avere coppie con proprietà diverse: un esempio è lo stato di spin di $t\bar{t}$, che vicino alla soglia di produzione è maggiormente correlato nel caso di un evento gg. Uno studio che voglia misurare l'entità di tali correlazioni risulta quindi essere significativamente facilitato da un metodo di discriminazione delle coppie risultanti sulla base del loro canale di produzione.

Il lavoro qui presentato ha quindi lo scopo di ottenere uno strumento per effettuare tale differenziazione, attraverso l'uso di tecniche di analisi multivariata. Tali metodi sono spesso applicati per separare un segnale da un fondo che ostacola l'analisi, in questo caso rispettivamente gli eventi gg e $q\bar{q}$. Si dice che si ha a che fare con un problema di *classificazione*.

Si è quindi studiata la prestazione di diversi algoritmi di analisi, prendendo in esame le distribuzioni di numerose variabili associate al processo di produzione di coppie $t\bar{t}$. Si è poi selezionato il migliore in base all'efficienza di riconoscimento degli eventi di segnale e alla reiezione degli eventi di fondo. Per questo elaborato l'algoritmo più performante è il *Boosted Decision Trees*, che permette di ottenere da un campione con purezza iniziale 0.81 una purezza finale di 0.92, al costo di un'efficienza ridotta a 0.74.

Abstract

The top quark is one of the fundamental fermions of the Standard Model, and is observed in the highest energy collisions. Our focus is the $t\bar{t}$ pair, which is produced through strong interaction in two cases: from gluon fusion (gg) or quark-antiquark annihilation ($q\bar{q}$). Different production channels lead to pairs with different characteristics: one example is the $t\bar{t}$ spin state, which near the production threshold presents higher correlations in the case of a gg event. A study that proposes to study the entity of such correlations can thus benefit from a way to discriminate pairs on the basis of their production channels.

This work has therefore the purpose of obtaining, through the use of multivariate analysis methods, a tool to select events in such way. Multivariate algorithms are often used to separate a signal from a background that pollutes the sample; in this case for the signal we choose gg events, while for the background we select to $q\bar{q}$ events. Such a problem is called a *classification* problem.

We thus studied the performance of some classifiers, using the distributions of some variables associated to the $t\bar{t}$ production process. Then we selected the best performing algorithm evaluating its efficiency in selecting signal events and rejecting background ones. The chosen classifier turns out to be the *Boosted Decision Trees*, which allows to obtain a sample of purity 0.92, starting from an initial purity of 0.81, at the cost of a reduced efficiency of 0.74.

Contents

Introduction

In the Standard Model the top is one of six quarks, and its peculiar phenomenology is driven by its mass. It is the only quark that decays semi-weakly, producing an on-shell W boson and a bottom quark; this results in a lifetime that is shorter than the typical hadronization time, and thus it plays a special role in the Standard Model and many of its extentions.

In proton-proton collisions top quarks are mostly produced in pairs, along with their antiparticle ($t\bar{t}$ pairs), and at LO this happens through the processes $gg \rightarrow t\bar{t}$ and $q\bar{q} \rightarrow t\bar{t}$ (gluon fusion and quark-antiquark annihilation). The higher the energy in these collisions, the more gg events dominate over $q\bar{q}$ ones: at the Tevatron, with $\sqrt{s} = 1.96$ TeV, about 85% of the production cross section comes from quark-antiquark annihilation, whereas at the LHC ($\sqrt{s} = 13$ TeV) approximately 90% of $t\bar{t}$ pairs production is due to gluon fusion [1].

These production channels cause the resulting pairs to have different properties. One example of this is found in [2], where correlations in the $t\bar{t}$ spin state are studied in order to look for violations of Bell Inequalities. In this analysis two complementary regimes are important: at threshold, i.e., when the top quarks are slow in the pair's rest frame, and when their transverse momentum is large. These correspond to two separate regions in the $m_{t\bar{t}} - \theta$ plane (where $m_{t\bar{t}}$ is the invariant mass of the system and θ is the angle of production in the center of mass reference frame); they are defined respectively by having $m_{t\bar{t}} \approx 2m_t$ and $m_{t\bar{t}} \gg 2m_t$ with $\theta \approx \pi/2$. At threshold gluon fusion leads to an entangled spin-0 state, whereas quark-antiquark annihilation to a spin-1 state. The latter, although subdominant at the LHC, acts as an irreducible background, making an analysis in such a regime impossible. This is an example of a case that would benefit from the ability to discriminate events on the basis of their production channel.

Moreover, such a result would work as an additional confirmation of what is already investigated in [3] and [4], evaluating if the cross sections ratio obtained working with PDFs is accurate. These works have inspired what is here presented, and some of the same discriminating variables were used in the process. However, a multivariate approach to the problem is proposed here, employing different algorithms thanks to the ROOT-integrated environment TMVA, *Toolkit for Multivariate Analysis*.

Multivariate techniques have often been used in high-energy physics, as it studies

phenomena characterized by a high number of parameters, all associated to the same event, and is sometimes lacking a definite mathematical model to describe the data. This warrants the use of multivariate algorithms, which are able to work with a large number of variables.

A typical area of application of such methods is background suppression, also known as classification, where one divides the data in two categories on some basis, assigning to one the label of "signal", to be isolated, and identifies the other one as "background", to be eliminated because it pollutes the data. This is precisely what we wish to achieve, providing a tool for classification of single events.

The thesis is organized as follows. The first chapter contains an introduction to the generalities of the Standard Model with a focus on the physics of the top quark and the production of $t\bar{t}$ pairs. The second chapter presents the methods of multivariate analysis, describing how each one of those used approaches the classification problem. Finally, the third chapter describes the analysis carried out and its results.

Chapter 1

The Standard Model and the physics of $t\bar{t}$ pairs

1.1 The Standard Model

As of today, the study of the fundamental constituents of matter is reduced to the theory known as *The Standard Model* (SM). It is a theoretical framework developed between the 1960s and 70s describing all particle physics phenomena, based on a combination of quantum mechanics and relativistic theory: *quantum field theory* (QFT). QFT treats particles as excited states (also called quanta) of their underlying quantum fields, which are more fundamental than the particles.

It is also a *gauge theory*, that is a field theory based on the existence of certain local symmetries; it thus requires the relative Lagrangian to be invariant under the action of certain symmetry groups. Such transformations are called *gauge transformations*. The symmetry group involved is $SU_C(3) \times SU_L(2) \times U_Y(1)$, where:

- $SU_C(3)$ is the color charge symmetry group, for the strong interaction, and it is the basis for its quantum field description, QCD (Quantum Chromodynamics).

- $SU_L(2) \times U_Y(1)$ is the symmetry group on which is built the electroweak theory, the unified description of electromagnetism (Quantum Electrodynamics, or QED) and the weak interaction (Quantum Flavordynamics or QFD).

1.1.1 Elementary particles

According to the Standard Model there are two types of fundamental particles, which are fermions with spin $\frac{1}{2}$, leptons and quarks, and mediating bosons with spin 0, 1.

The leptons are six, divided in three generations with two species each: we have the electron e paired with the electron neutrino ν_e, the muon μ and its neutrino ν_μ, and the

tau τ along with the tau neutrino ν_τ. e, μ and τ have -1 electrical unit charge, while neutrinos are neutral.

Similarly, the quarks are also six and divided in three pairs, one *up-like* quark paired with a *down-like* quark. They are respectively the up u and the down d, the charm c with the strange s, and the top t and the bottom b. However, their charge differs from leptons, up-like quarks having $+\frac{2}{3}$ charge and down-like one having $-\frac{1}{3}$, and they have color charge as well. More properly one refers not to different species of elementary particles but to *flavors*.

Along with these 12 particles there are 12 *antiparticles*, which are identical to those just described (same mass and spin) except for charges, which are opposite. For example, the electron's antiparticle is the positron, often written as e^+ to highlight the opposite electric charge [5].
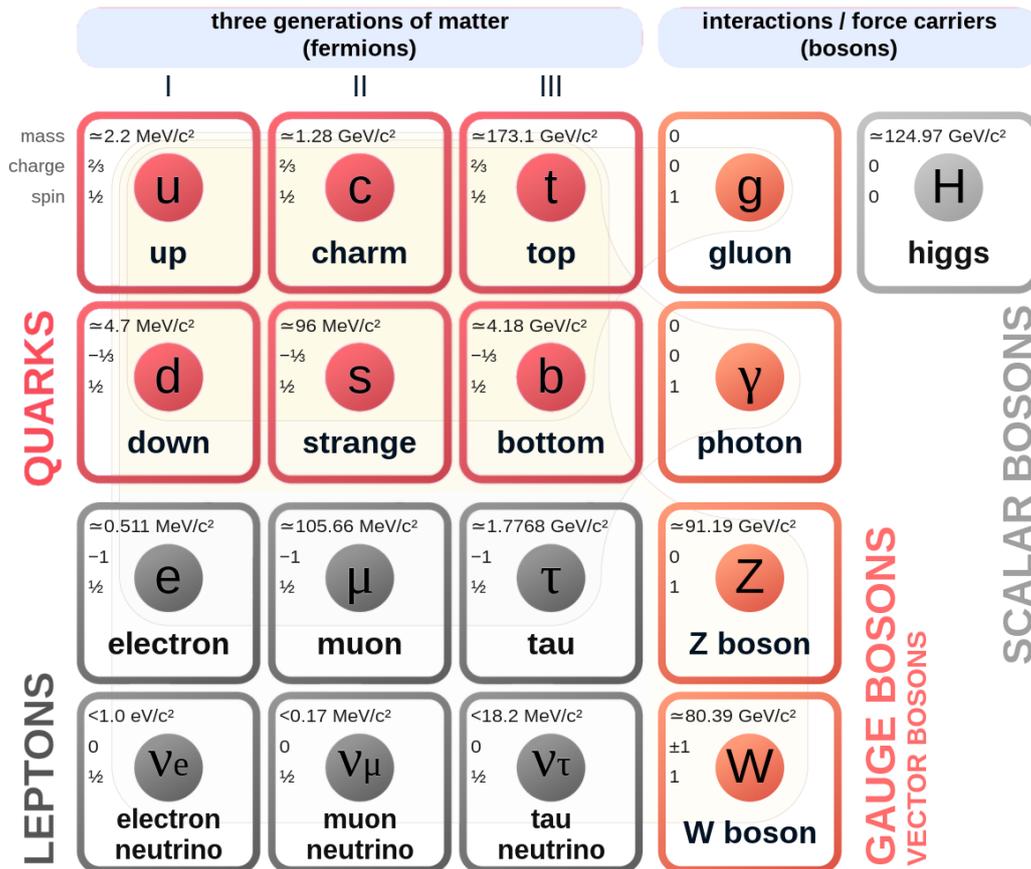

Figure 1.1: *Elementary particles of the Standard Model: quark, lepton and gauge fields.*

Lastly there are 12 gauge bosons, with spin one, associated with the three fundamental interactions: the electromagnetic force, the strong interaction and the weak interaction. Gravity, the fourth fundamental force, still lacks a consistent quantum theory; however the masses involved in subnuclear processes are so small that the gravitational field can easily be neglected, allowing for a complete description of most phenomena observed even without a complete theory.

To each fundamental interaction is associated a charge, which allows for fermion fields to couple with the force ones, represented by the mediating bosons. All particles that are electrically charged take part in electromagnetic interaction, carried by the photon γ, which is massless and neutral. For the strong interaction we have what is called the color charge and the propagators are 8 colored gluons g, massless and not electrically charged. Finally, the weak interaction corresponds to the *weak isospin* and three mediators, W^+, W^- and Z^0; they have both mass and electric charge. A 13th scalar boson is required for the theory to be consistent, also known as Higgs boson H; it represents the Higgs field, a scalar field that couples with most fermions causing them to acquire mass.

The gauge bosons are $1 + 3 + 8 = 12$, coherent with the dimension of the symmetry group, as they each represent one of the generators. Different characteristics of the fundamental fermions and bosons are summarized in Fig. 1.1.

1.1.2 The electromagnetic interaction and QED

The electromagnetic interaction was already very well described by classical theories, and its behavior summed up by Maxwell's equations. Its study however also highlighted some of the behaviors that lead to the birth of quantum field theories, and the reformulation of electromagnetism as *Quantum Electrodynamics* (QED) is the starting point for the description of almost all fundamental interactions.
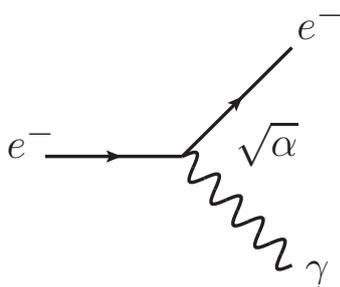
Figure 1.2: *Fundamental QED vertex. The electron e^- can be exchanged with any fermion.*

Associated with the electromagnetic interaction is the symmetry group $U(1) = \{e^{i\theta}, \theta \in [0, 2\pi]\}$. The charge associated with the electromagnetic field is the electric charge, and the corresponding propagator is the photon γ. Coherently with its null mass it has infinite range, and its intensity is parameterized by the fine structure constant α_{em}:

$$\alpha_{em} = \frac{e^2}{4\pi\epsilon_0 \hbar c} \approx \frac{1}{137} \ . \qquad (1.1.1)$$

The fine structure constant is one of the three coupling constants; each interaction has its own, an adimensional parameter not actually constant but depending from the energies in the process.

Generally speaking the fundamental electromagnetic interaction can be of two kinds: the creation of a particle and its antiparticle from a photon or the emission of one from a fermion. In Fig. 1.2 is depicted the fundamental vertex of the interaction.

1.1.3 The weak interaction and QFD

The weak interaction was at first observed in radioactive decays of atomic nuclei. β^- emission is due to the decay of the neutron thanks to the weak force:

$$n \rightarrow p \ e^- \ \bar{\nu}_e \ . \tag{1.1.2}$$

This highlights one of the defining characteristics of the weak interaction: it is not constricted by laws of flavor conservation. It is also less intense than the others, as can be seen evaluating its coupling constant, using the mass of the proton m_p and the Fermi constant G_F:

$$\alpha_w = \frac{G_F m_p c}{\hbar^3} \approx 10^{-5} \ . \tag{1.1.3}$$

Due to the fact that it is less intense its timescales are longer, which means that it is often masked by the other interactions. It has been generally observed in decays that show the violation of flavor conservation.

The weak force is universal, and both quarks and leptons take part in its processes. Its mediators are the $W^\pm$ and Z^0 bosons, and it has finite range due to their mass. It is associated with the isospin symmetry group $SU(2)$. The processes coupling fermions with the $W^\pm$ bosons can be described arranging the 12 fundamental particles in weak isospin doubles, also called generations:

$$\begin{bmatrix} \nu_e \\ e^- \end{bmatrix} \begin{bmatrix} \nu_\mu \\ \mu^- \end{bmatrix} \begin{bmatrix} \nu_\tau \\ \tau^- \end{bmatrix} \quad \begin{bmatrix} u \\ d' \end{bmatrix} \begin{bmatrix} c \\ s' \end{bmatrix} \begin{bmatrix} t \\ b' \end{bmatrix}$$

The process can be described as having one of the bottom particles transforming in one of the top ones emitting a W^- boson, or vice versa with the emission of a W^+. To be more precise in the case of quarks there is also the mixing of flavors of the interacting fermions. In fact the $W^\pm$ bosons do not couple directly with the mass eigenstates d, s and b but with a linear combination of those, d', s' and b', as described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

Lastly there are neutral weak processes, where a fermion can emit Z^0 boson without changing its nature. Some of the fundamental vertices of the weak interaction can be seen in Fig. 1.3.

Weak isospin T_3 serves as an additive quantum number that restricts how the particle can interact with the $W^\pm$ and Z^0 bosons; however it is also correlated with the electric
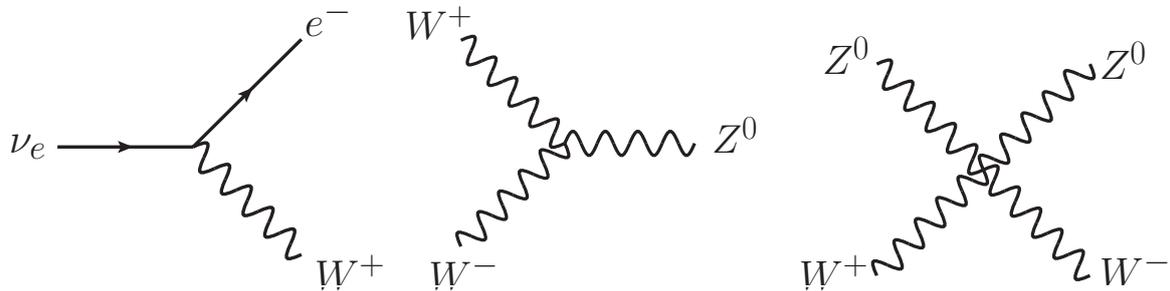
Figure 1.3: *Some of the fundamental vertices of QFD, showing the emission of a W^+ boson as well as the interaction between propagators.*

charge Q. Both electromagnetic and weak processes conserve the weak hypercharge, defined as $Y_W = 2(Q - T_3)$, and they are in fact described through a unified theory, the *Electroweak field theory*.

Another defining feature of QFD is that it is a chiral theory, acting only on the left-handed components of the fermion fields. The weak force in fact completely violates parity symmetry as well as charge-parity symmetry, and it is the only one that does so.

1.1.4 The strong interaction and QCD

The strong force is an interaction between quarks and the associated propagators, the gluons. The associated charge is the color charge, which can take three values, red, green and blue (r, g and b), as well as their opposite, antired, antigreen and antiblue ($\bar{r}$, $\bar{g}$ and $\bar{b}$); the gluons are massless and not electrically charged, but have color, each carrying one color charge and one anticolor charge, so we can distinguish 8 of them. A possible choice of basis is $r\bar{b}$, $r\bar{g}$, $b\bar{r}$, $b\bar{g}$, $g\bar{r}$, $g\bar{b}$, $(r\bar{r} - g\bar{g})/\sqrt{2}$ and $(r\bar{r} + g\bar{g} - 2b\bar{b})/\sqrt{6}$. Due to the fact that they are colored gluons are self-interacting, as one can see in Fig. 1.4, where the fundamental vertices of QCD are presented.

Strong interactions were first introduced in the study of the forces between nucleons, which are very strong at lengths about 1 fm but rapidly decreases at just double that distance. The complicated profile of the potential acting in the nucleus is now explained as a residual of the interaction between quarks in the internal structure of nucleons, just as London dispersion forces are due to the electromagnetic force that holds together the atom. Both protons and neutrons are in fact *baryons*, bound states of three quarks. Together with *mesons*, bound states of two quarks, they make up the subatomic particles known as *hadrons*.

One of the key properties of the strong interaction is *color confinment*, which means that no free states of quarks can have non-null color: all known hadrons, in fact, are colorless. The phenomenon can be understood by noting that gluons have color charge
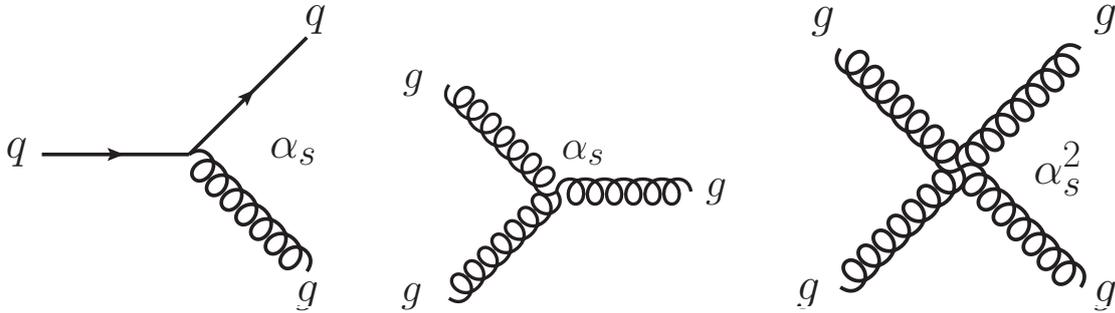
7

Figure 1.4: *Fundamental vertices of QCD. Two of them show the interaction among gluons.*

and are self-interacting, and this feature causes the force between quarks to be constant regardless of their separation; therefore, as two color charges are separated, at some point it becomes energetically favorable for a new quark–antiquark pair to appear, so only neutral-color states of quarks are observed by detectors.

The strong interaction conserves flavor and electric charge, but changes color. The conservation of flavor can be more precisely expressed introducing quantum number associated with the flavor of each quark. However, the behavior of strong processes involving the u and d quarks highlights that they are invariant for rotations in the bidimensional space of u and d flavor, called isospin space. This (approximate) symmetry is expressed as invariance under transformation of the $SU(2)$ group. The interaction is also generally invariant for transformations of the $SU(3)$ group.

The coupling constant, which expresses the intensity of the force, can be evaluated assuming a typical length of the interaction of 1 fm and a typical mass of 938 GeV, the mass of the proton. We thus have:

$$\alpha_s = \sqrt{\frac{m_p l \hbar}{c}} \approx 1 \ . \tag{1.1.4}$$

Compared to the other coupling parameters α_s is much larger, in fact particles that decay due to the strong force have a very short lifetime.

1.2 Some useful concepts

At the LHC one studies collisions from two proton beams (pp collisions). In this setting it is common to distinguish three different levels of analysis - *parton level*, *particle level* and *detector level*.

A parton level analysis involves only the incoming (so part of the protons' internal structure) and outgoing partons. In this work this means taking into account only the initial state particles (gluons or quarks), the tops and their kinematics.

The outgoing partons are however unstable products and will quickly hadronize, due to the confinement of the strong interaction, or decay weakly. This means that the parton level products are not directly observed, but reconstructed through the resulting stable hadrons and leptons. This is what is known as the particle level analysis.

Finally, we have the detector level. This refers to the work of inferring the dynamics of the collision from the response of the detectors, which happens when the final state particles interact with them. In an ideal situation, if detectors were completely efficient, the detector level would be identical to the particle level, however that is not possible due to imperfections of the apparatus or finite resolution effects.

Another key concept is that of Leading Order (LO) and Next-to-Leading Order (NLO) calculations. This refers to the process of computing the transition amplitudes for a given process, which is generally done through the use of Feynman diagrams.

The amplitude for scattering is the sum of each possible interaction history over all possible intermediate particle states. This can be done through a perturbative series expansion, where each term can be represented as a Feynman diagram. The expression is written in terms of a coupling constant, usually called α, which depends on the type of interaction in the scattering. In this work we are focusing on the production to $t\bar{t}$ pairs through the strong interaction, so we consider the strong coupling constant α_s.

If the coupling constant is small, the terms of the expansion that are proportional to higher powers of α will be suppressed, and have small contributions to the sum; depending on the level of accuracy wanted one can thus ignore those less significant contributions. So we can distinguish Leading Order and Next-to-Leading Order calculations: at LO one keeps only the most important terms, that are proportional to the lowest non-zero power of α, while at NLO the next largest terms are considered as well. This translates to having some diagrams of the process that do not contribute at LO, but do at NLO.

In QCD the coupling constant is generally not small, and this complicates matters as it does not always allow perturbation techniques. This is however not always the case, and in high energy events or small distance interactions α_s is small ($\alpha_s \approx 0.1$), and perturbative calculations can be used.

1.3 The top quark and $t\bar{t}$ pairs production

The top quark is the heaviest known fundamental particle, with a mass of 172.56 ± 0.4 GeV. Due to its mass it is produced on-shell only in very high energy events, and it was discovered relatively late, in 1995 at the Tevatron collider at FNAL by the CDF and

DØ collaborations [6, 7]. It takes part in all interactions having both electric and color charge.

1.3.1 Production

In hadron colliders top quarks can be produced along with their antiparticle, in what are usually called $t\bar{t}$ pairs, through the strong interaction, or as singles from weak interaction. Some Feynman diagrams of these processes are shown in Fig. 1.5.
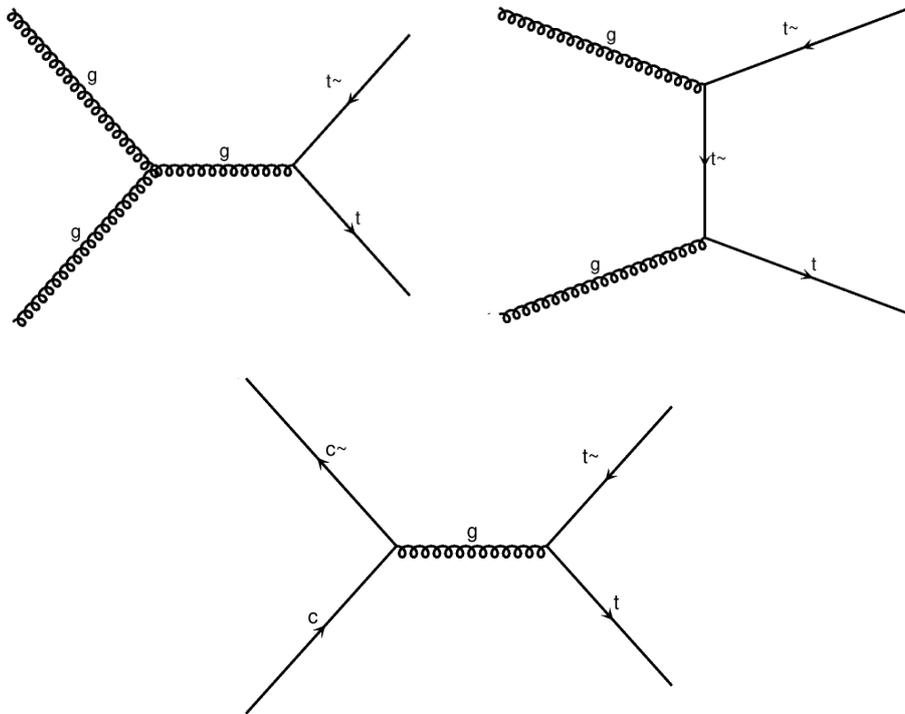

Figure 1.5: *Some diagrams for the different $t\bar{t}$ production processes in pp collisions at LO QCD. The quark charm can be exchanged for any other quark. Diagrams related by mirroring or taking the charge conjugate are identified and only one is drawn.*

Our focus is the production of top-antitop pairs. At LO these come from either gluon fusion or quark-antiquark annihilation, distinguishing two different *production channels*:

$$g\,g \;\rightarrow\; t\,\bar{t} \tag{1.3.1}$$

$$q\,\bar{q} \;\rightarrow\; t\,\bar{t}. \tag{1.3.2}$$

Note that $q\bar{q} \rightarrow t\bar{t}$ can happen only through an s-channel, while $gg \rightarrow t\bar{t}$ it is possible in both s and t-channels.

The fully differential top cross sections for $gg/q\bar{q} \to t\bar{t}$ are (at LO in QCD and neglecting EW corrections) [8]:

$$\frac{d^2\sigma}{d\beta d\rho}(s)_{gg} = \frac{4\pi\alpha_s^2}{12s}\left[\left(1+\rho+\frac{\rho^2}{16}\right)\ln\frac{1+\beta}{1-\beta} - \beta\left(\frac{7}{4}+\frac{31}{16}\rho\right)\right] \qquad (1.3.3)$$

$$\frac{d^2\sigma}{d\beta d\rho}(s)_{q\bar{q}} = \frac{8\pi\alpha_s^2}{27s}\beta\left[1+\frac{\rho}{2}\right] \qquad (1.3.4)$$

with $\rho = 4m_t^2/s$ and $\beta = \sqrt{1-\rho}$ is the velocity of the top in the $t\bar{t}$ center of mass frame of reference. In order to obtain the total $p\bar{p}$ (or pp) cross section one should sum the partonic gg and $q\bar{q}$ cross section over the respective luminosities in hadronic collisions.

At NLO in addition to these production channels we have one more contribution:

$$g\,q \;\to\; t\,\bar{t}\,g\,(q) \qquad (1.3.5)$$
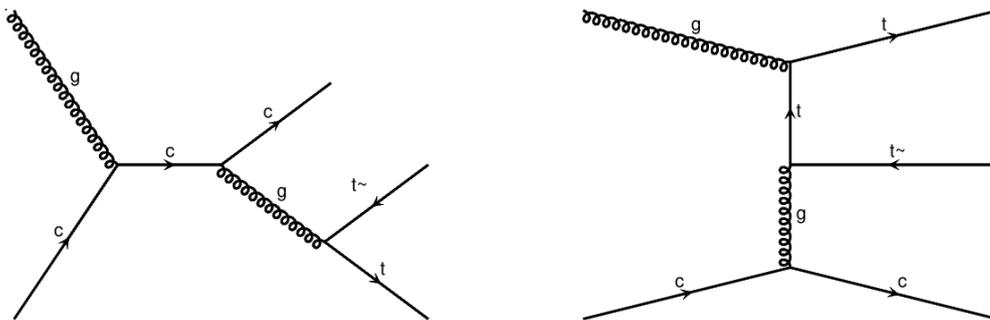

Figure 1.6: *Some diagrams for the $gq \to t\bar{t}\,q$ production processes in pp collisions at NLO QCD. The quark charm can be exchanged for any other quark. Diagrams related by mirroring or taking the charge conjugate are identified and only one is drawn.*

1.3.2 Decay

As mentioned before, due to its mass the top has a mean life of approximately $1.4\,\mathrm{GeV}^{-1}$, over ten times smaller than the typical QCD hadronization timescale. It therefore usually decays before it has time to hadronize, giving the unique opportunity to study the bare structure of a quark.

Both the electromagnetic and the strong interaction respect flavor conservation, which means the top decays semi-weakly, almost exclusively to a W boson and a bottom quark (it can also decay to a strange or a down quark, however the $W\,b$ branching ratio is by far dominant at $R = 1.014 \pm 0.003\,(stat.) \pm 0.032\,(syst.)$ [9]). Thus we have:

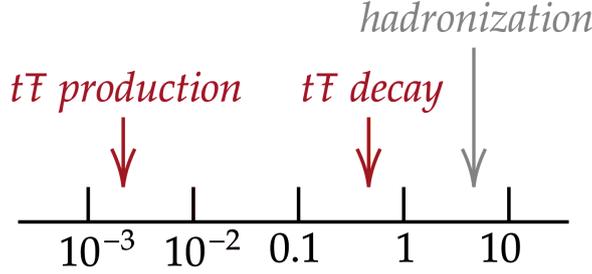$$t\,\bar{t} \;\to\; b\,W^+\,\bar{b}\,W^- . \qquad (1.3.6)$$

Figure 1.7: *Timescales for the top, in GeV^{-1}.*

The W boson can then decay either to a lepton and the corrisponding neutrino, or to light quarks giving two hadroninc jets. These different final states characterize what are called the different *decay channels*: the former is the leptonic final state, while the latter is known has hadronic.

Due to the fact that there are two tops that decay there are three possible overall final states. If both W bosons decay leptonically we obtain what is called a dilepton final state, and we have

$$t\,\bar{t}\ \rightarrow\ b\,\ell^+\,\nu\ \ \bar{b}\,\ell^-\,\nu \tag{1.3.7}$$

In this work, ℓ is limited to be either an electron or a muon as the τ lepton generally decays too quickly to be seen in a final state. If one decays leptonically while the other does so hadronically the result is a mixed lepton and jets state, which can be either of the type

Figure 1.8: *Representation of the decay of a top quark that leads to the emission of a lepton.*
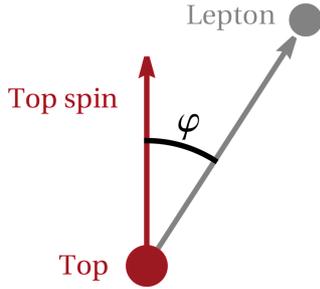
$$t\,\bar{t}\ \rightarrow\ b\,\ell^+\,\nu\ \ \bar{b}\,q\,\bar{q}\quad\text{or} \tag{1.3.8}$$

$$t\,\bar{t}\ \rightarrow\ b\,q\,\bar{q}\ \ \bar{b}\,\ell^-\,\bar{\nu}\ . \tag{1.3.9}$$

Lastly, one can have a fully hadronic decay:

$$t\,\bar{t}\ \rightarrow\ b\,q\,\bar{q}\ \ \bar{b}\,q'\,\bar{q}'\ . \tag{1.3.10}$$

It is a general feature of weak decays that the angular distribution of decay products is correlated with the parent particle's spin. A weakly decaying fermion at LO has differential width:

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\varphi_i} = \frac{1 + \alpha_i\cos\varphi_i}{2} \tag{1.3.11}$$

12

where Γ is the total width of the mother (the top), φ_i is the angle between the i-th decay product and the top quark spin axis in the latter's rest frame (see Fig. 1.8), and the parameter α is known as the spin analyzing power, and its value depends on the decay products [10].
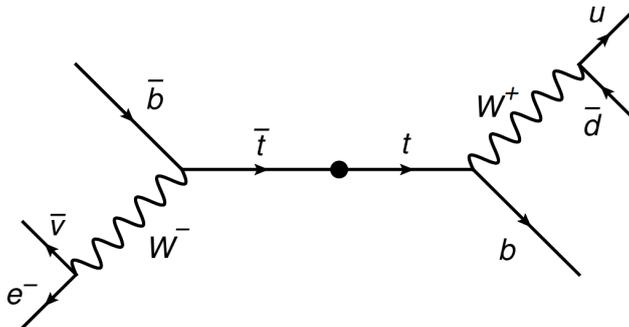

Figure 1.9: *Decay of a $t\bar{t}$ pair, resulting in a mixed lepton and jet final state.*

1.3.3 Kinematics

At LO the kinematics of the pair is determined by the invariant mass of the system $m_{t\bar{t}}$ and the production angle in the zero momentum reference frame θ, both of which can be calculated from the components of the top's momentum p_t; while the former is a relativistic invariant quantity, the latter is not. We thus have:

$$m_{t\bar{t}}^2 = (p_t + p_{\bar{t}})^\mu (p_t + p_{\bar{t}})_\mu = (E_t + E_{\bar{t}})^2 - (\vec{p}_t + \vec{p}_{\bar{t}})^2 \tag{1.3.12}$$

$$\cos\theta = \frac{p_z}{|\vec{p}|} \tag{1.3.13}$$

where $\hat{z}$ is taken to be the direction of the beam. In this work we identify θ and $\pi - \theta$.

Most observables of interest arise from the fact that in colliders particles' velocities are (up to corrections of the order of $\Lambda_{QCD}/\sqrt{s}$) along the beam axis. This leads to the definition of various quantities that, when boosting to frames of references at different velocities along the beam axis, are either invariant or have transformation properties that make them easy to handle and useful for analysis.

One example is the transverse momentum of both t and $\bar{t}$, which is the component of momentum perpendicular to the beam line, and can be cast as a function of $m_{t\bar{t}}$ and θ:

$$p_T = \sin\theta \, |\vec{p}| = \sin\theta \sqrt{\frac{m_{t\bar{t}}^2}{4} - m_t^2} \, . \tag{1.3.14}$$
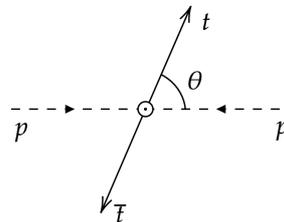

Figure 1.10: *Schematic of a pp collision resulting in a $t\bar{t}$ pair, in the center of mass reference frame.*

13

Note that this quantity is invariant under boosting along the $\hat{z}$ axis.

Two quantities of interest are also the top's velocity v (also called β) in the zero momentum frame of reference, and the $t\bar{t}$ center of mass' rapidity, $y_{t\bar{t}}$. In such reference frame the velocity is related to the invariant mass of the system through:

$$v = \frac{|\vec{p}|}{E} = \sqrt{1 - \frac{4\, m_t^2}{m_{t\bar{t}}^2}}\,, \tag{1.3.15}$$

while for the rapidity we use the definition most commonly used in particle physics, which is relative to the beam axis. A particle's rapidity is defined as:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}\,. \tag{1.3.16}$$

y is zero when a particle is emitted transverse to the beam, so $\theta = \pi/2$, and tends to $\pm\infty$ when the particle is moving close to the $\hat{z}$ axis in either direction. Under a boost in the z direction to a frame with velocity β, $y \to y - \tanh^{-1}\beta$, therefore the difference between the rapidities of two particles is invariant under $\hat{z}$ boosts.

The rapidity is related to the pseudorapidity η, defined as

$$\eta = -\ln\tan\frac{\theta}{2} = \frac{1}{2} \ln \frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\,. \tag{1.3.17}$$

because when a particle travels close to the speed of light (or, equivalently, in the approximation that the mass of the particle is negligible) one can make the substitution $E \approx |\vec{p}|$ and thus $\eta \approx y$. Pseudorapidity is often useful because it can be measured when the mass and momentum of the particle are unknown [1].

Further quantities, relative to the kinematics of the decay products, can also be studied. In order to do so, it is best to fix a basis in the zero momentum frame of reference, the *helicity basis*. It is given by:

$$\hat{k} = \text{top direction}\,, \tag{1.3.18}$$

$$\hat{r} = \frac{\hat{p} - \hat{k}\cos\theta}{\sin\theta}\,, \tag{1.3.19}$$

$$\hat{n} = \hat{k} \times \hat{r} \tag{1.3.20}$$

where $\hat{p}$ is the beam axis, and θ the production angle. A schematic representation is in Figure 1.11.
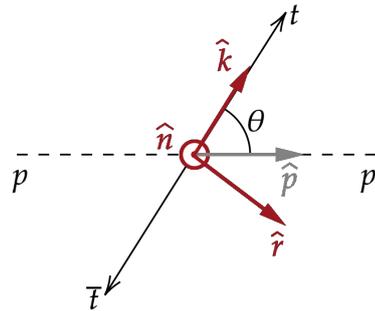

Figure 1.11: *Helicity basis $\{\hat{k}, \hat{r}, \hat{n}\}$, $\hat{n}$ is into the page.*

Chapter 2

Multivariate analysis methods

Multivariate analysis (MVA) is a type of stochastic analysis that allows to work with large numbers of variables at the same time. A typical area of application of multivariate techniques is background suppression, also known as *classification*. The reason to apply statistical training multivariate methods is, in most cases, simply the lack of knowledge about the mathematical dependence of the quantity of interest on the relevant measured variables, as it is our case.

In this work the algorithms have been implemented using the *Toolkit for Multivariate Data Analysis* (TMVA), a ROOT-integrated environment offering different tools for this analysis, thus allowing comparisons between different methods based on their performance.

All multivariate techniques in TMVA belong to the family of "supervised learning" algorithms: they work in two phases, the *training* phase and the *testing* phase. During the training, events for which the desired output is known are used to determine the mapping function that describes a decision boundary in parameter space, and later the algorithm is tested on an independent sample of data.

2.1   TMVA methods

Among the algorithms offered by TMVA three where selected for the current work: *Fisher Linear Discriminant* (FDL), *Boosted Decision Trees* (BDT) and *Multilayer Perceptron* (MLP).

2.1.1   *Fisher Linear Discriminant*

The *Fisher Linear Discriminant*, as the name suggests, belongs to the family of linear classifiers, algorithms that use a linear function of the input data. If the input is a real

vector $\mathbf{x}$, the output (related to the probability of the event being part of the signal) will be:

$$y = \mathbf{w} \cdot \mathbf{x} = \sum_j w_j x_j. \tag{2.1.1}$$

$\mathbf{w}$ is usually called weights vector. This process of classification can be imagined at the act of splitting the input space with a hyperplane, one side identified as belonging to the signal and the other side to the background.

The design of a good classifier becomes rapidly more difficult as the dimensionality of the input space increases. Fisher discriminants solve this problem performing event selection in a transformed one-dimensional variable space, projecting the data on an axis such that the distance between events of the same type (signal or background) is minimized, while distance between different types of events is maximized. The classification then is reduced to the act of placing a cut on the output, known also as the *Fisher discriminant* [11].

Let $\bar{x}_{S(B),k}$ with $k = 1, ..., n$ be the class-specific sample means and C the covariance matrix. The covariance matrix can be decomposed into the sum of two terms, one which describes the dispersion of events relative to the means of their own class (*within-class matrix, W*), and one relative to the overall sample means (*between-class matrix, B*). We can define the *Fisher coefficients* F_k as:

$$F_k = \frac{\sqrt{N_S N_B}}{N_S + N_B} \sum_{j=1}^{n} W_{kj}^{-1}(\bar{x}_{S,j} - \bar{x}_{B,j}), \tag{2.1.2}$$

where $N_{S(B)}$ is the number of signal (background) events used in the training. The Fisher discriminant for the i-th event will then be:

$$y_F(i) = F_0 + \sum_{k=1}^{n} F_k x_k(i). \tag{2.1.3}$$

F_0 is an offset that centers the mean $\bar{y}_F$ for all the events to zero [12].

In spite of their simplicity, Fisher discriminants can deliver a surprisingly good performance, and can be in certain cases competitive with nonlinear classifiers (which generally present better results). One can prove that are optimal when working with Gaussian distributed variables that present linear correlations. On the other hand, no discrimination is achieved when a variables has the same sample mean for both signal and background events. This is why a classification process using a Fisher discriminant can benefit from a suitable transformation of the input variables, changing the shape of the signal and background distributions to obtain a different sample mean.

2.1.2 Boosted Decision Trees

A *decision tree* is a binary tree structured classifier. It is made of different nodes where repeated left/right (yes/no) decisions are taken on one single variable at a time until a stop criterion is fulfilled and a *leaf* is reached. In this way the phase space is split into many regions that are eventually classified as signal or background, depending on the type of the majority of training events that end up in the final leaf. The process is similar to placing rectangular cuts. However, whereas a cut-based analysis is able to select only one region of phase pace, the decision tree is able to split the phase space into a large number of hypercubes, each of which is identified as either "signal-like" or "background-like".
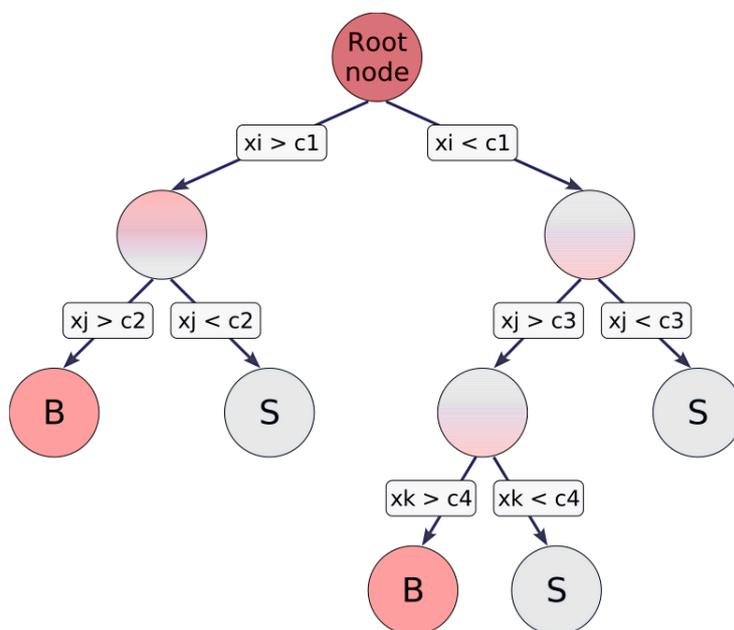

Figure 2.1: *Figure taken from [12], schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is applied to the data. Each split uses the variable that at this node gives the best separation between signal and background when being cut on (the same variable may thus be used at several nodes, while others might not be used at all). The leaf nodes at the bottom end of the tree are labeled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes..*

A shortcoming of this algorithm is its instability with respect to statistical fluctuations in the training sample. For example, if two input variables exhibit similar separation

power, these fluctuations may cause the tree growing algorithm to decide to split on one variable instead of the other. In such a case the whole tree structure is altered below this node, possibly resulting also in a substantially different classifier response. The problem is overcome thanks to the *boosting* of decision trees.

The boosting of decision trees combines many trees into a *forest*. These are all derived from the same training sample and they are later combined into a single classifier once a weight is assigned to each tree; an event is then classified as part of the signal or background on a majority vote of the classifications done by each tree in the forest, taking into account its weight. Boosting increases the statistical stability of the classifier and is able to drastically improve the separation performance compared to a single decision tree. There are different boosting algorithms available for TMVA, and for this work the one chosen was AdaBoost.

A ranking of the BDT input variables is derived by counting how often they are used to split decision tree nodes, and by weighting each split occurrence by the separation it has achieved and by the number of events in the node.

BDTs are often considered as among the best classifiers available. This is because, given the semplicity of their structure, little tuning is required in order to obtain reasonably good result; moreover they are also insensitive to the inclusion of poorly discriminating input variables, as they are essentially ignored during the process of constructing the tree. However, the simplicity of BDTs has the drawback that their theoretically best performance on a given problem is generally inferior to other techniques like neural networks [12].

2.1.3 Multilayer Perceptron

The *Multilayer Perceptron*, MLP, is a type of Artificial Neural Network (ANN). The development of ANNs was inspired by the research on the central nervous system and the neurons; however, currently the approach stimulated by biological research has been abandoned for an approach based on statistics, mathematics and optimization theory.

In its most abstract form a neural network is a nonlinear function $f : \mathbf{x} \rightarrow y$, where $\mathbf{x}$ is the input vector and y the output determining the result of the classification. The term network arises because the function f is a composition of other functions g_i, which are themselves compositions of other functions h_i and so on (see Fig.
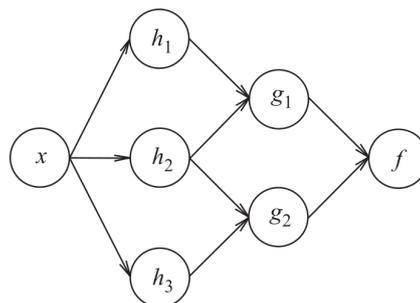

Figure 2.2: *ANN dependence graph. It highlights the relations between the different nodes, having the output f be a composition of the g_i, which are themselves compositions of the h_i.*

18

[2.2](#)). This structure can be represented in a graph where each function corresponds to a node (or *neuron* as they are called), linked to the others (thus showing dependence between functions), and we can see how this approach highlights the similarities with the central nervous system's structure.
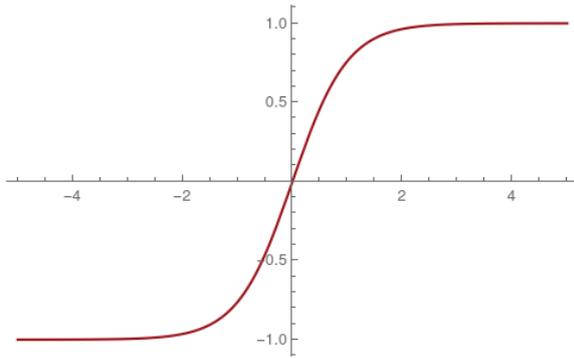

Figure 2.3: *Plot of the hyperbolic tangent,* $\tanh(x)$. *As one can see, it is monotonically increasing and the output ranges for -1 to 1. For $x \gg 1$ $\tanh(x) = 1$, and for $x \ll -1$ the output is -1.*

For each neuron i, w_i is the weight of the corresponding link, and $g_i(\mathbf{x}')$ is its output, which is itself a composition of functions. The activation function K has the purpose of reducing the neuron's output from a real number to either "on" or "off"; this is why it is generally a sigmoid, for example in this application the one used was the hyperbolic tangent [11]. The non linearity of the sigmoid can be crucial in improving the performance of a classifier, as it allows to select the region in input phase space corresponding to the signal using a non linear boundary, that is generally more effective.

In most applications the functions f is a weighted sum of the different neurons' outputs, with the addition of a nonlinear activation function. We can thus write:

$$f(\mathbf{x}) = K\left(\sum_i w_i g_i(\mathbf{x}')\right). \tag{2.1.4}$$

In the case of the *Multilayer Perceptron*, the complexity can be reduced by organizing the neurons in layers and only allowing direct connections from a given layer to the following one. For a classification problem with n input variables one has the *input layer* with n neurons, taking each a variable's value, at least one *hidden layer* with an arbitrary number of nodes, and the *output layer*, consisting of a single neuron holding the output value y.

The MLP neural network implements a variable ranking that uses the sum of the weights-squared of the connections between the variable's neuron in the input layer and the first hidden layer. The importance I_i of the input variable i is given by

$$I_i = \bar{x}_i^2 \sum_{j=1}^{N} (w_{ij}^{(1)})^2 \tag{2.1.5}$$

where x_i is the sample mean of input variable i [12].

Chapter 3

Results

Events were simulated in two contexts, first at Leading Order only at Parton level, focusing solely on the "true" kinematics of top and antitop pairs, then at Next to Leading Order at Parton and Particle level, taking into account decay products as well. They were generated in in collaboration with the authors of [2], for more details see the Appendix.

The events were used as input for classifiers implemented in the ROOT framework, with the purpose of selecting events from gluon fusion (gg events), and separating them from the background ($t\bar{t}$ pairs coming from all other production channels). At LO the background corresponds to just $q\bar{q}$ events, while at NLO the gq production channel needs to be taken into account as well. From this point on when mentioning the signal we will refer to gg events, and the background will correspond to $q\bar{q}$ and gq ones.

The data was then analyzed with the TMVA package, after being divided into two sets, one to train the algorithms and one to test them. We now present the results of the analysis.

3.1 Analysis at Leading Order

At Leading Order (LO) the generated events take into account only the kinematics of the top quark pair. Some variables in the process where isolated examining their distribution in gg and $q\bar{q}$ events. The ones deemed useful were later used by the TMVA algorithms in the classification process: they are the invariant mass of the system, $m_{t\bar{t}}$, the absolute value of the rapidity of the $t\bar{t}$ pair, $|y_{t\bar{t}}|$, and the angle at which the top and the antitop come out in the centre of mass reference frame, θ. Distributions of the three variables are shown in Fig. 3.1.

These parameters are not the only ones that differentiate the two production channels. Any observable can potentially be used as a discriminant, as long as its distributions in signal and background events differ significantly. While the variables above are not the only ones with this characteristic, when working with multivariate algorithms it is also
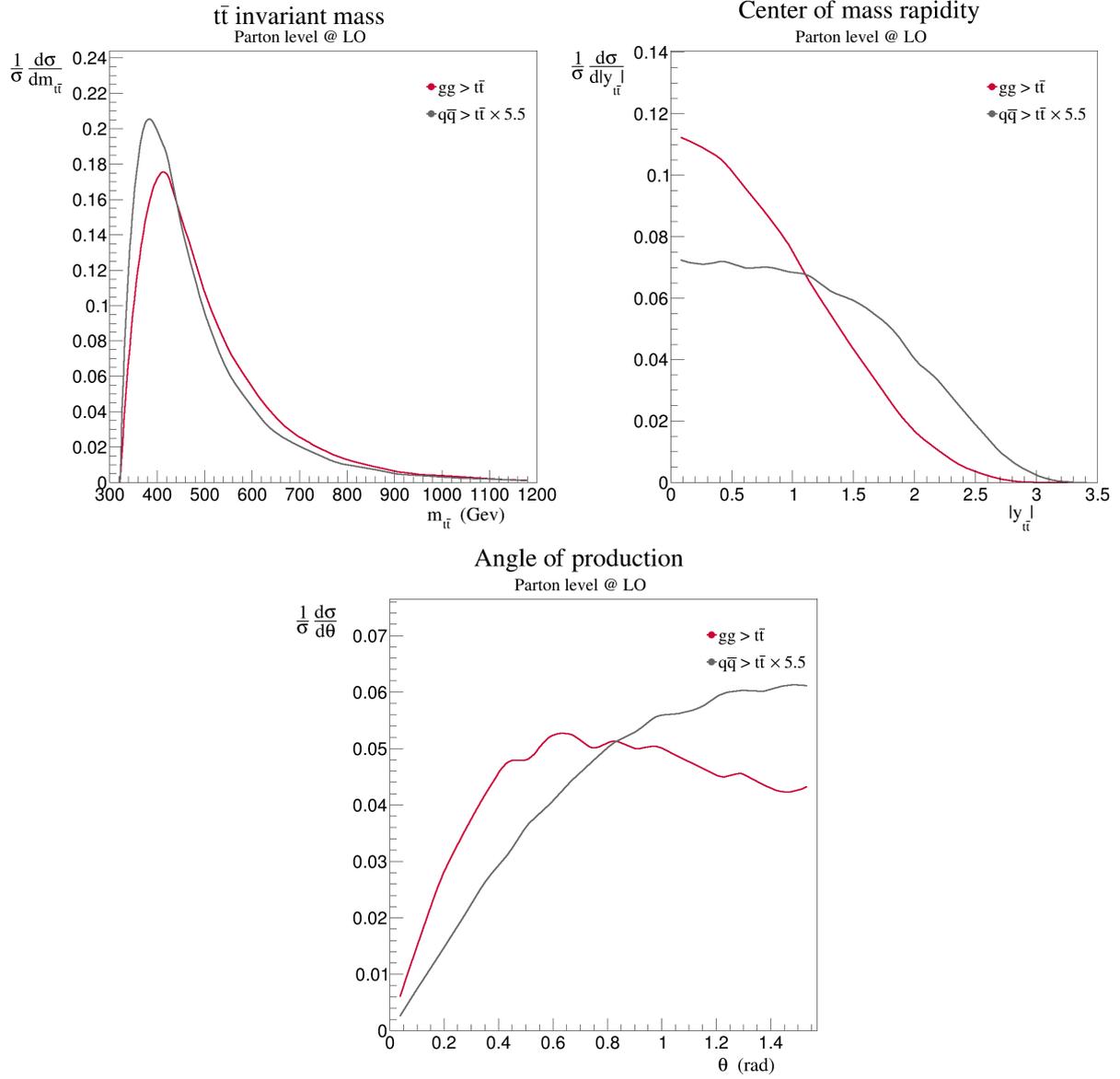
Figure 3.1: *Normalized differential cross section $\frac{1}{\sigma}\frac{d\sigma}{dO}$, where O represents each of the selected observables, in signal and background events at Parton Level at LO QCD. In order from left to right, up to down: the invariant mass of the system $m_{t\bar{t}}$, the absolute value of the rapidity of the pair, $|y_{t\bar{t}}|$ and the production angle θ. The $q\bar{q}$ curves were multiplied by 5.5 to allow a better comparison of the distribution shapes.*

important to make sure the variables are as much uncorrelated as possible. Referencing the description of the kinematics of the pair, it is easy to see that quantities such as the top's velocity in the zero momentum frame of reference or its transverse momentum

cannot be employed, as they can be expressed as functions of $m_{t\bar{t}}$ and θ. Correlation plots for the variables used, both in gg and $q\bar{q}$ events were produced, as is shown in Fig. 3.2 and 3.3.
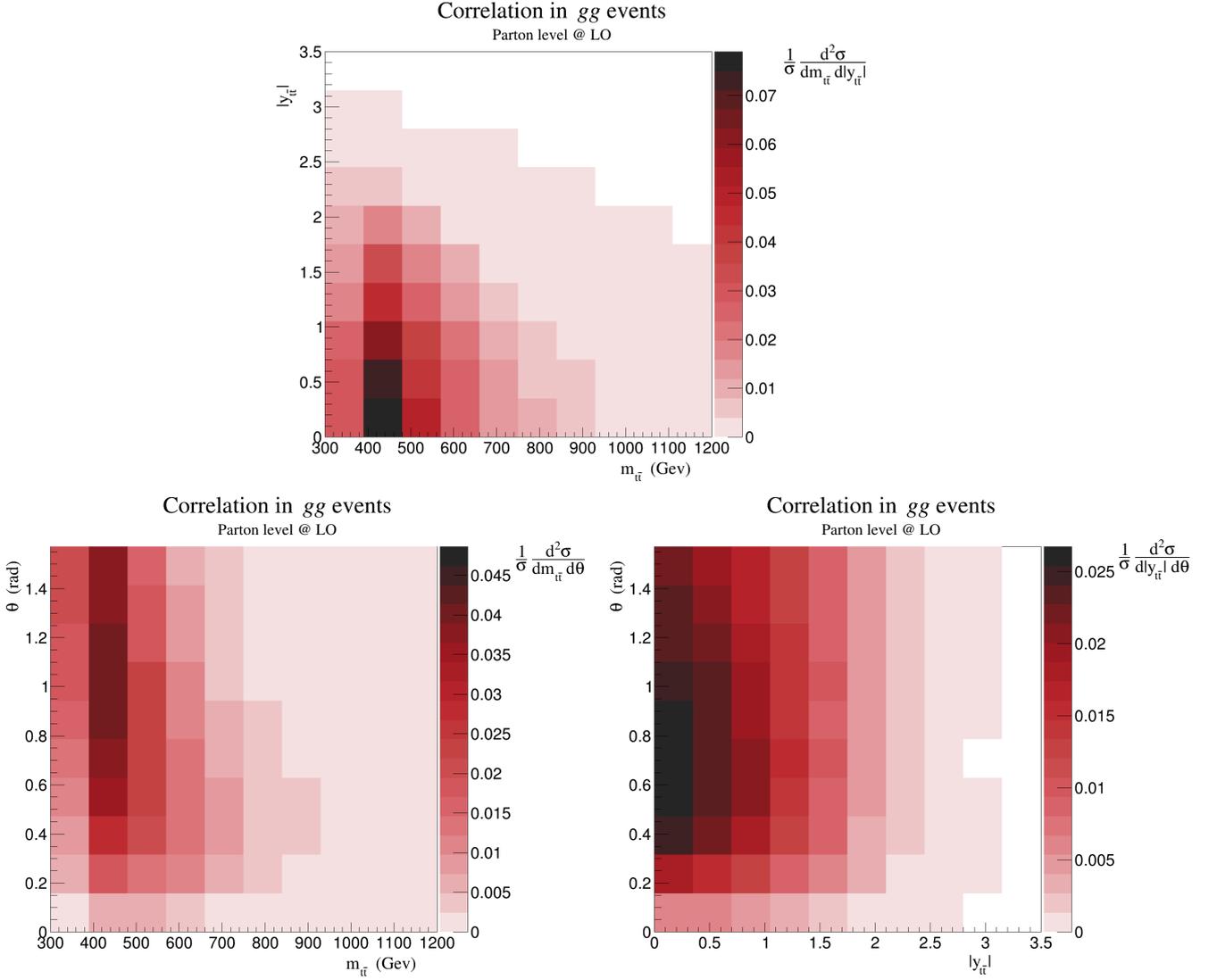

Figure 3.2: *Normalized differential cross section $\frac{1}{\sigma}\frac{d^2\sigma}{dXdY}$, where X and Y are each of the selected observables, in signal (gg) events at Parton Level at LO QCD. In order from up to down, left to right: distribution in $m_{t\bar{t}} - |y_{t\bar{t}}|$ plane, in $m_{t\bar{t}} - \theta$ plane, and in $|y_{t\bar{t}}| - \theta$ plane.*
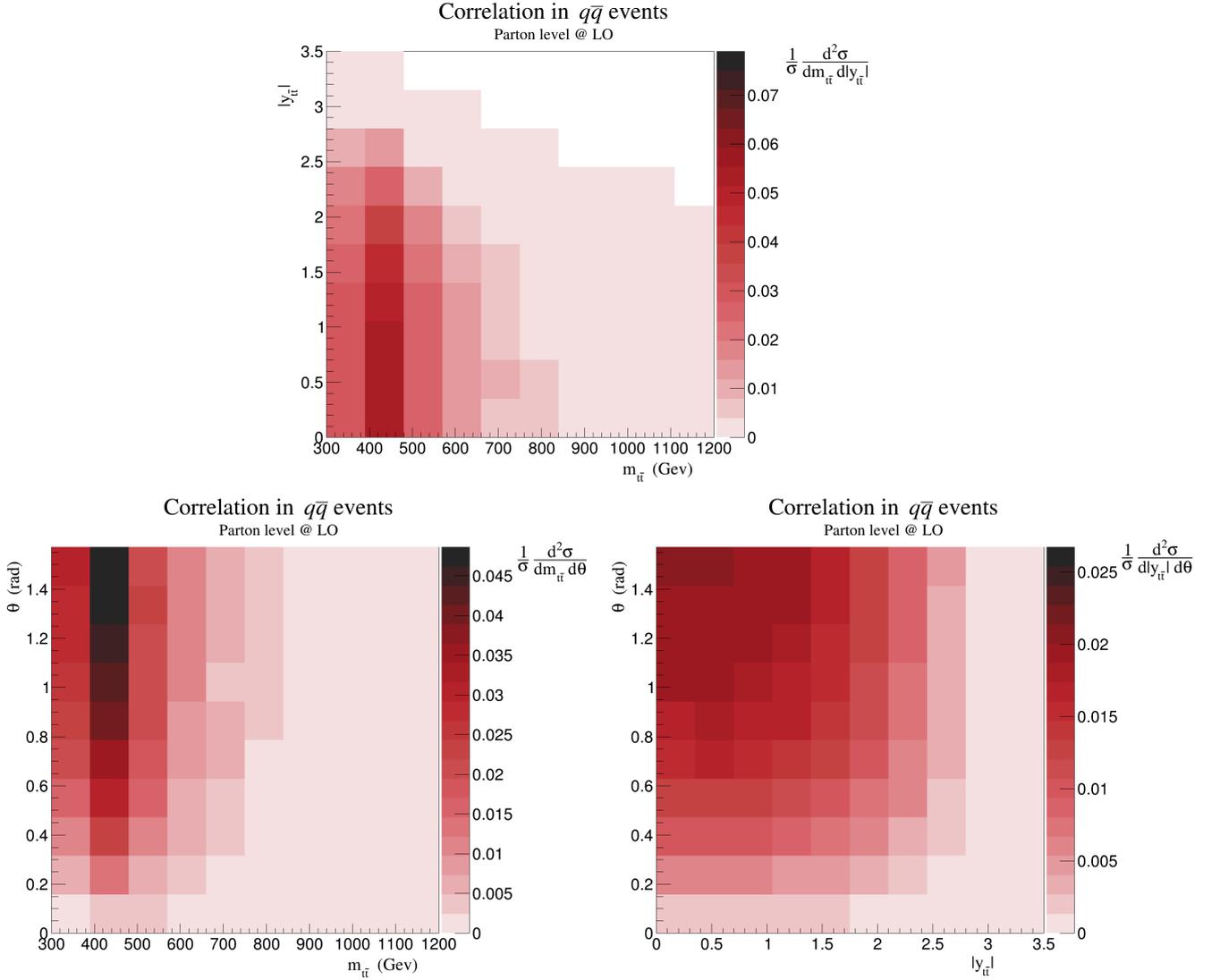
Figure 3.3: *Normalized differential cross section $\frac{1}{\sigma}\frac{d^2\sigma}{dXdY}$, where X and Y are each of the selected observables, in background ($q\bar{q}$) events at Parton Level at LO QCD, multiplied by a 5.5 factor. In order from up to down, left to right: distribution in $m_{t\bar{t}} - |y_{t\bar{t}}|$ plane, in $m_{t\bar{t}} - \theta$ plane, and in $|y_{t\bar{t}}| - \theta$ plane.*

3.1.1 Variable ranking and correlation

TMVA performs a preliminary evaluation of the methods used, and one of the things it produces are the correlation matrices both for the signal and the background. These are shown in Fig. 3.4.
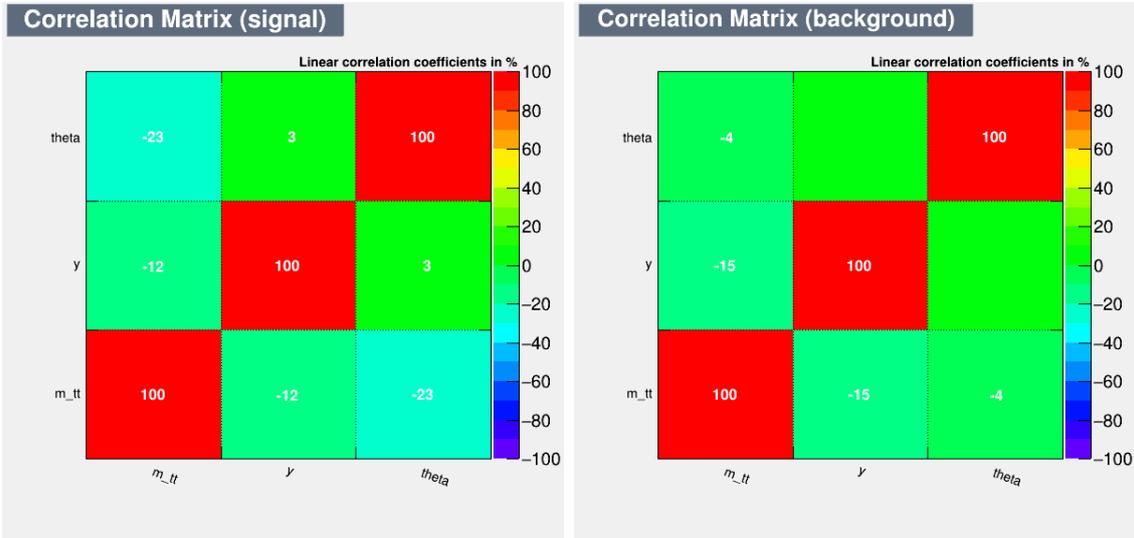
Figure 3.4: *Correlation matrices for signal and background events at Parton Level at LO QCD.*

TMVA also ranks the variables according to their performance in the classification process. Specifically, for each variable the separation $\langle S^2 \rangle$ is calculated. It is defined as the integral:

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(f_S(x) - f_B(x))^2}{f_S(x) + f_B(x)} \, dx \tag{3.1.1}$$

where $f_S(x)$ and $f_B(x)$ are the signal and background PDFs for the variable x respectively. $\langle S^2 \rangle$ is null for identical distributions of the variable in signal and background events, and $\langle S^2 \rangle = 1$ if there is no overlap. The value assigned to each variable is indipendent from the algorithms used.

Variables ranked according to their separation are shown in Table 3.1.

Variables	$\langle S^2 \rangle$
$|y_{t\bar{t}}|$	5.969×10^{-2}
θ	4.196×10^{-2}
$m_{t\bar{t}}$	1.751×10^{-2}

Table 3.1: *Variables ranked according to their separation $\langle S^2 \rangle$.*

Next, TMVA computes the importance I of a classifier. I is algorithm-specific and defined differently each time, and allows to quantify a variable's significance in the training

phase [12]. Each classifier approaches the problem in a distinct way, so different algorithms will favor different variables, depending on whether their distributions present the right characteristics. For a more detailed description on how the variables are evaluated by each algorithm we reference the second chapter of this thesis. Again, a ranking in decreasing order of importance is shown in Table 3.2.

Fisher Linear Discriminant	Boosted Decision Trees	Multilayer Perceptron
$\lvert y_{t\bar{t}} \rvert$	$\lvert y_{t\bar{t}} \rvert$	$m_{t\bar{t}}$
θ	$m_{t\bar{t}}$	$\lvert y_{t\bar{t}} \rvert$
$m_{t\bar{t}}$	θ	θ

Table 3.2: *Variables ranked according to their importance I for each of the algorithms used.*

3.1.2 Algorithm performance

For each algorithm we construct the COD, *Classifier Output Distribution*, which represents the normalized distribution of events recognized as part of the signal and background, as a function of a response variable t_{cut}. These are shown in Figure 3.5.

A frequent problem related to the use of multivariate methods is *overtraining*: it occurs when a machine learning problem has too few degrees of freedom because too many parameters were adjusted to not enough data points. Overtraining leads to increase in the classification performance over what is realistically achievable during the training phase, but poor performance during the testing one, because the algorithm has adapted itself too much to the sample given for training [12]. The problem can be therefore detected comparing performance results in both phases. This comparison is done thanks to the Kolmgorov-Smirnov test, which warns of excessive bias towards the training sample if the result is a very small number, below 0.01. For each algorithm one can see the result of such test in Fig. 3.5, above the CODs.

The goal of each classifier is to produce distributions for signal and background events with very little overlap, accumulating by TMVA convention signal (background) events at large (small) classifier output values. Hence, cutting on the output and retaining the events with t_{cut} larger than the prefixed value selects signal samples with efficiencies and purities that respectively decrease and increase with the cut value. The higher the cut value the less background events will pollute the final sample, however this will have the downside of having to exclude a higher number of events.

CODs thus allow to qualitatively choose the best performing algorithm, by looking at the shape of the distributions. For a more quantitative approach one can look at *Receiver Operating Characteristic* (ROC) curves and integrals for each classification method. A
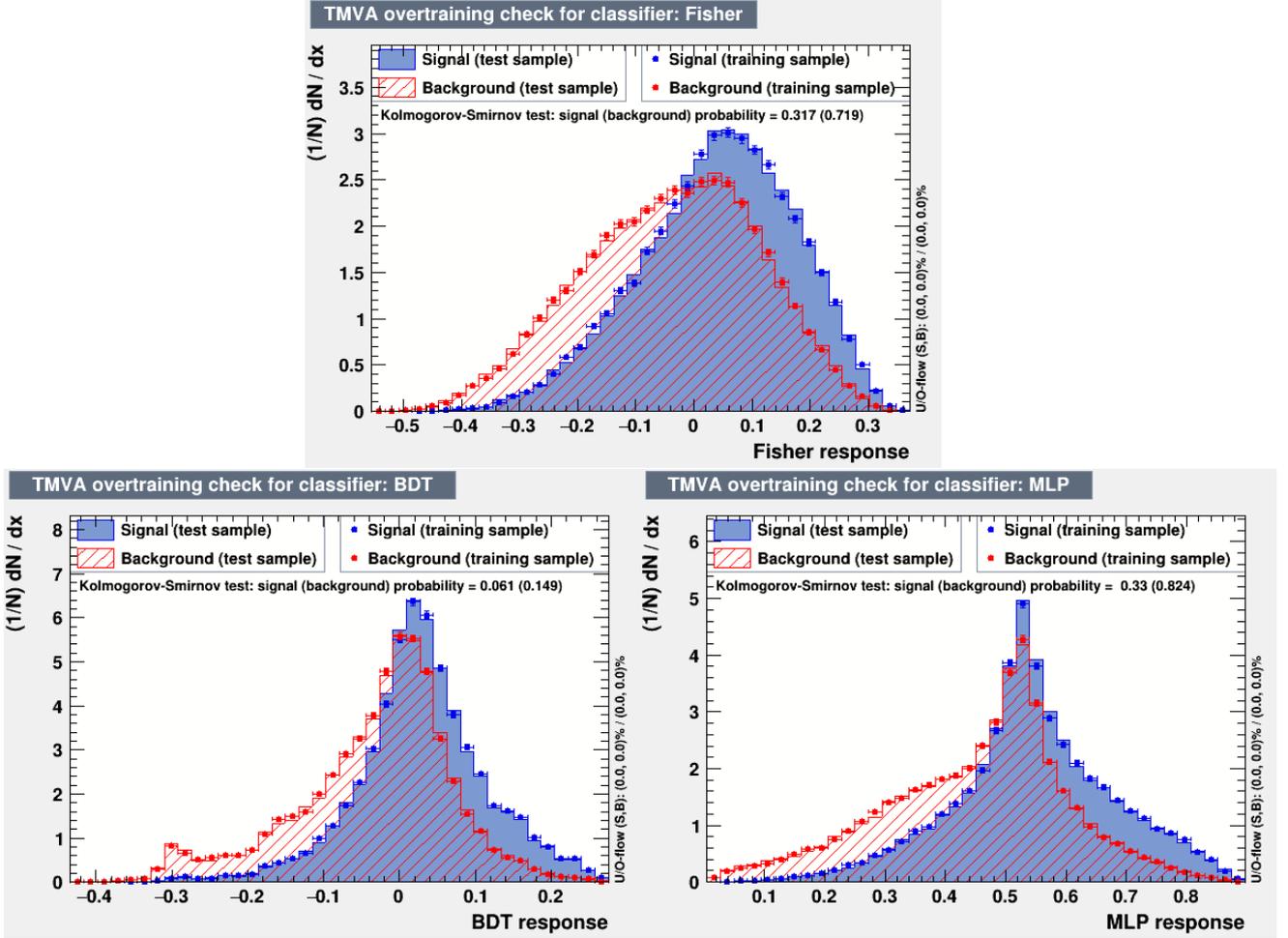
Figure 3.5: *Classifier Output Distributions for the three algorithms used: FLD, BDT and MLP.*

ROC curve is created plotting the efficiency of the signal ϵ_S and the background rejection r_B for each value of t_{cut}. For a fixed value c of t_{cut}, these are defined as:

$$\epsilon_S = \frac{N_S(t_{cut} > c)}{N_S^{tot}} \tag{3.1.2}$$

$$r_B = 1 - \epsilon_B = 1 - \frac{N_B(t_{cut} > c)}{N_B^{tot}} \tag{3.1.3}$$

where $N_{S(B)}$ is the number of events recognized as part of the signal (background) as a function of t_{cut}. Both ϵ_S and r_B have value between 0 and 1, the better the classification algorithm performs the closer they will be to 1. Thus, in efficiency-rejection phase space

the ideal ROC curve will be a step function valued always 1 for both variables and having integral of exactly 1. Of course no real classification process will achieve that, but this approach allows to rank algorithms on their performance. This is shown in Table 3.3, while ROC curves are presented in Figure 3.6.

Algorithm	Integral
Multilayer Perceptron	0.669
Boosted Decision Trees	0.667
Fisher Linear Discriminant	0.654

Table 3.3: *Algorithms ranked according to their performance, thanks to the ROC integral (right).*
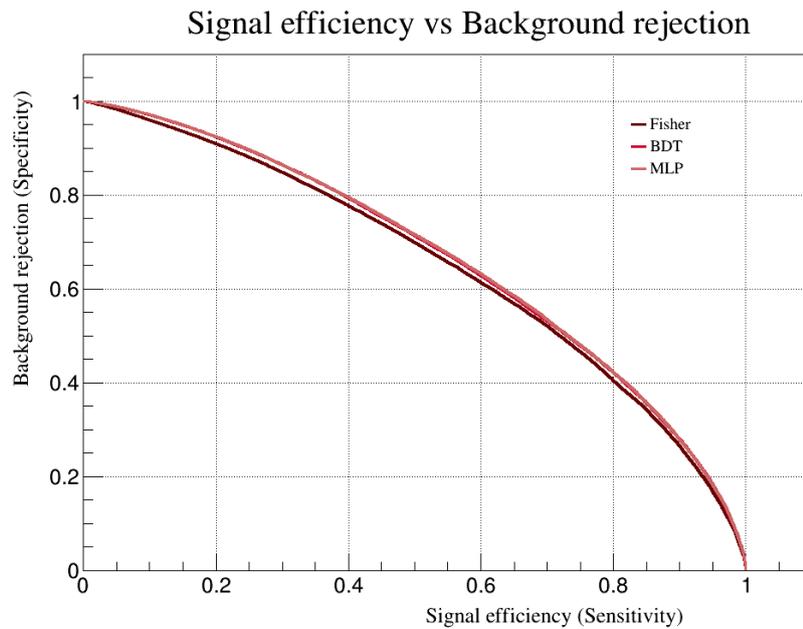
Figure 3.6: *Plots of the three ROC curves.*

As a general comment, we can conclude that the use of LO variables leads to a poor classification, where all the classifiers perform very similarly. Therefore, we extended our analysis to include also NLO-related variables, as reported in the next section.

3.2 Analysis at Next to Leading Order

3.2.1 Validation

As a first step the same observables which were selected at Leading Order were evaluated at Next to Leading Order (NLO), and are show in Fig. 3.7. In addition to the *gg* and $q\bar{q}$ production channel we now need to take into account the *gq* process as well.
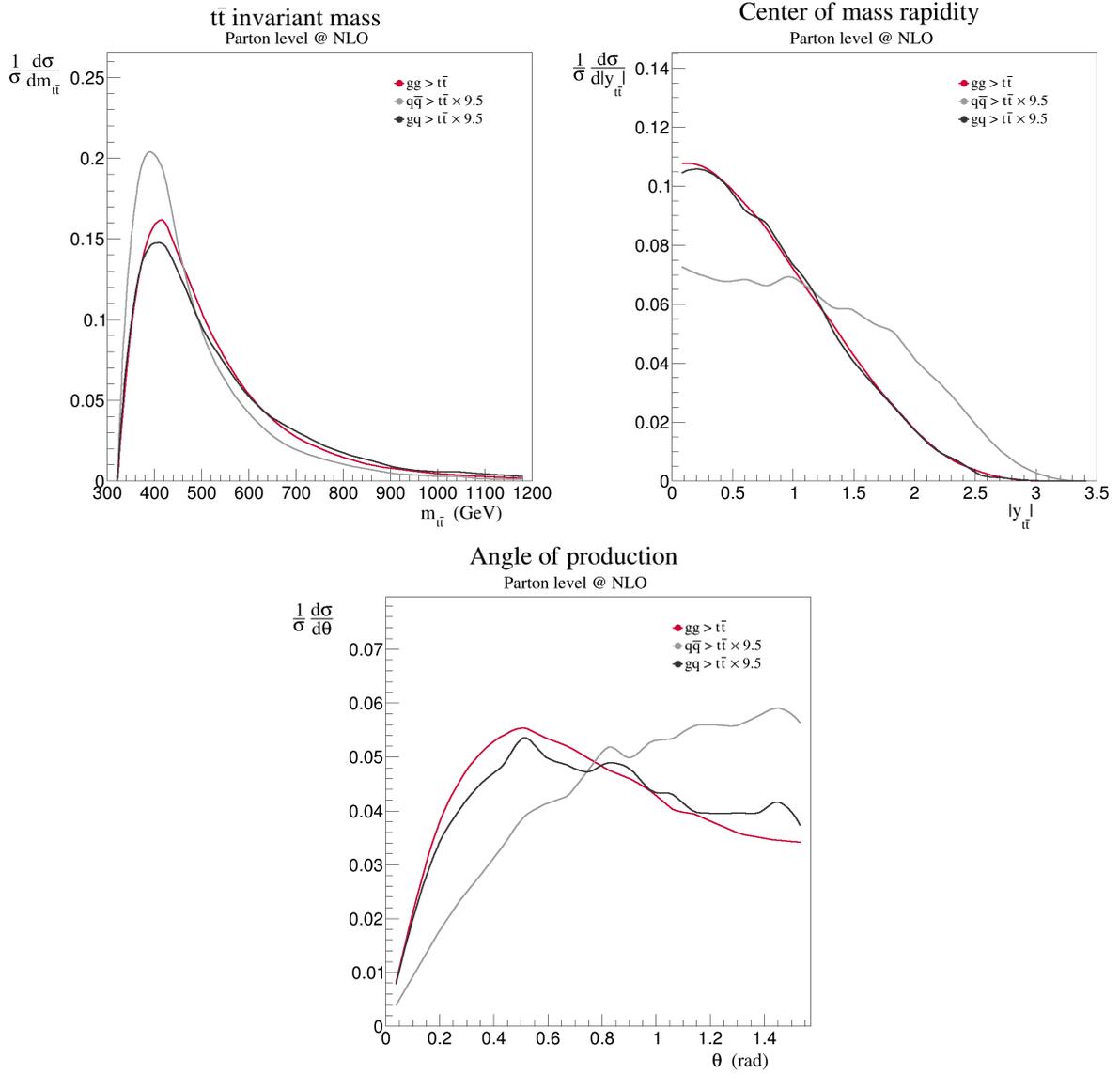

Figure 3.7: *Normalized differential cross section $\frac{1}{\sigma}\frac{d\sigma}{dO}$, where O represents each of the selected observables, in signal and background events at Parton level at NLO. The variables are the same mentioned above. The $q\bar{q}$ curves were multiplied by 9.5 to allow a better comparison of the distribution shapes.*

As one can see the curves are very similar to their counterpart at LO, with the addition of gq events having distributions close to identical to gg ones. This presents an additional challenge to the classification process, as can be see from Fig. 3.7, where one can also compare the two ROC curves, at LO and NLO. The performance at this stage is worse than what previously obtained.
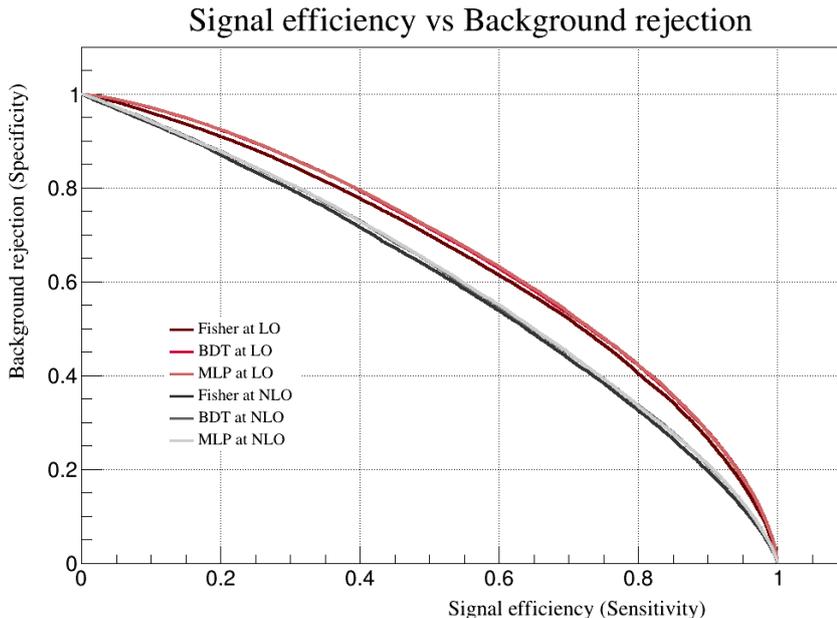

Figure 3.8: *Plots of the three ROC curves using the same variables both at LO and at NLO.*

3.2.2 Additional variables

At NLO there is a further kinematic quantity that helps the discriminating process: it is the transverse momentum of the $t\bar{t}$ pair, p_T. This is because a gluon is more likely to radiate additional gluons carrying a small fraction of its energy compared to a quark, so one expects a higher gluon radiation with low p_T in gg events than in scattering events between quarks and antiquarks [4]. This provides higher transverse momentum for the pair produced through gluon fusion as one can see in Fig. 3.9. Note that the very same variable, at LO, is exactly zero and therefore does not help in the classification process.

In addition, as was already noted in [3] and [4], the number of low transverse momentum charged particles N_{trk} is of interest. Because of the larger probability for low transverse momentum gluon radiation in gg events, one will have a larger number of low p_T charged particles in processes involving more gluons. Taking inspiration from this approach, we used the total number of jets N_j resulting from the event as another discriminating variable. The plot is also shown in Fig. 3.9.
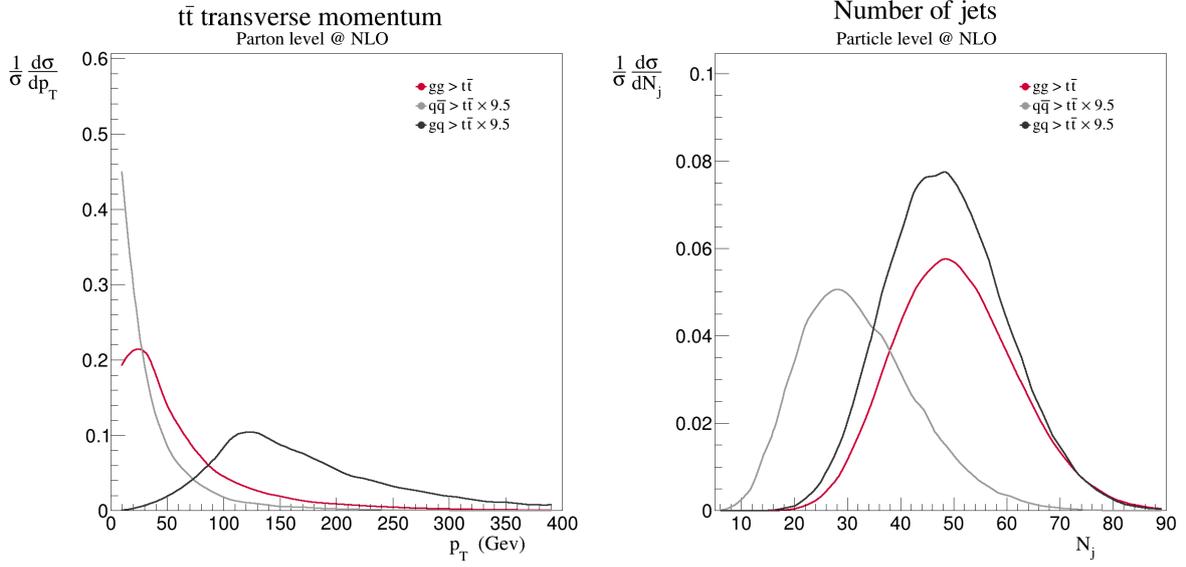
Figure 3.9: *Normalized differential cross section $\frac{1}{\sigma}\frac{d\sigma}{dO}$, where O represents each of the selected observables, in signal and background events at both Parton and Particle level at NLO. Here we see the transverse mometum of the $t\bar{t}$ system p_T and the number of jets in the final state of each event, N_j.*
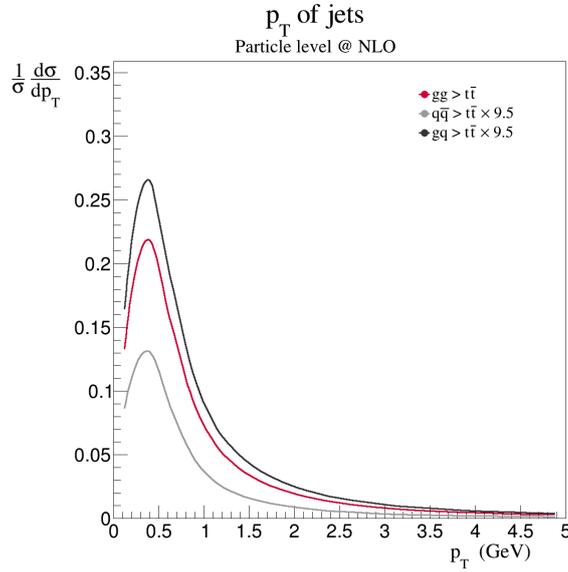

Figure 3.10: *Normalized differential cross section $\frac{1}{\sigma}\frac{d\sigma}{dp_T}$ in signal and background events at Particle level at NLO, displaying the distribution of the transverse momentum of the jets in the final state.*

No constrains were used on the p_T of the jet but we required that they have $|\eta| < 5$. In fact, as one can see from the distribution of p_T of jets in Fig. 3.10, most of them have low transverse momentum, so placing a cut on a high value of p_T (i.e., 10 GeV) does not impact the performance of the classification task.

We then performed the classification processes with these two new variables, and we present here the results comparing them with was previously obtained.

3.2.3 Algorithm performance

In the classification process both $q\bar{q}$ and gg events are considered part of the background, and their combined distributions do not achieve the same level of separation from the signal compared to the analysis at Parton level. Some plots are shown in Fig. 3.11.
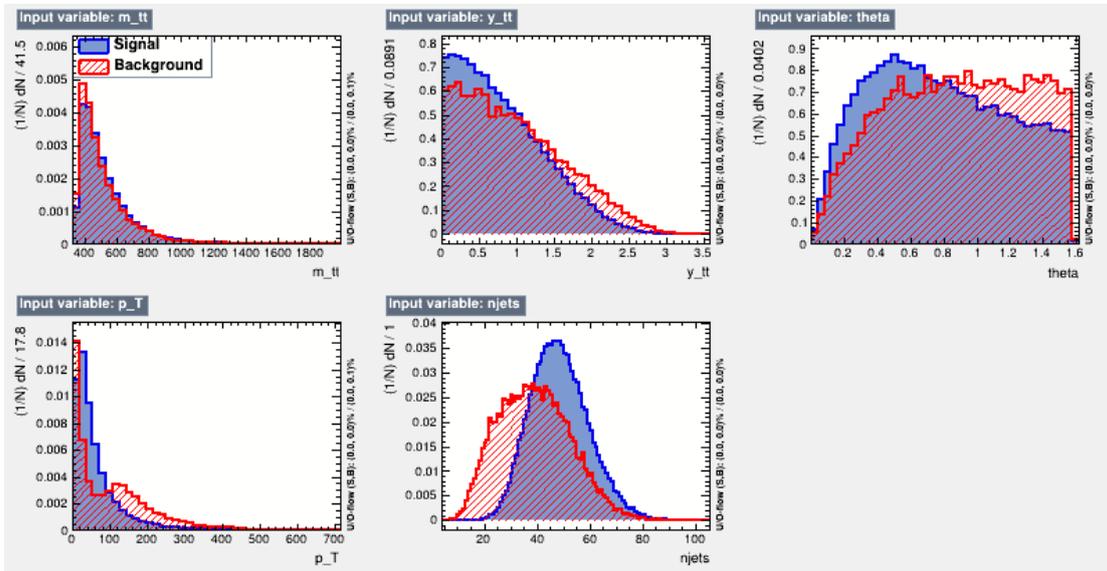

Figure 3.11: *Plots of the variables distributions distinguishing between signal and background events.*

As done previously, the evaluation of variables done by TMVA is useful to examine possible correlations between the variables as well as how they contribute to the discriminating process. Correlation matrices for the signal and the background are in fact presented in Fig. 3.12, and Table 3.4 along with 3.5 rank the observables by separation $\langle S^2 \rangle$ and importance I respectively. The variables in this process display some level of correlation.

The performance of the classifiers is then evaluated thanks to the CODs as well as the ROC curves, shown below in Fig. 3.13 and 3.14.
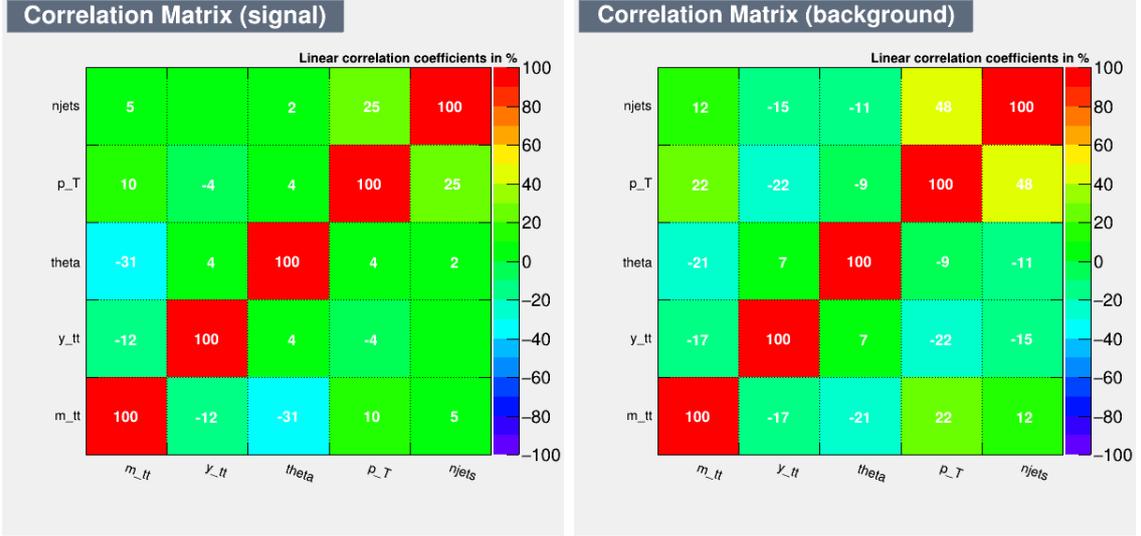
Figure 3.12: *Correlation matrices for signal and background events at NLO.*

Variables	$\langle S^2 \rangle$
N_j	1.590×10^{-1}
p_T	1.285×10^{-1}
$|y_{t\bar{t}}|$	1.985×10^{-2}
θ	1.472×10^{-2}
$m_{t\bar{t}}$	5.321×10^{-3}

Table 3.4: *Variables ranked according to their separation $\langle S^2 \rangle$.*

Fisher Linear Discriminant	Boosted Decision Trees	Multilayer Perceptron
N_j	N_j	p_T
p_T	p_T	N_j
$|y_{t\bar{t}}|$	θ	$m_{t\bar{t}}$
θ	$y_{t\bar{t}}$	$y_{t\bar{t}}$
$m_{t\bar{t}}$	$m_{t\bar{t}}$	θ

Table 3.5: *Variables ranked according to their importance I for each of the algorithms used.*
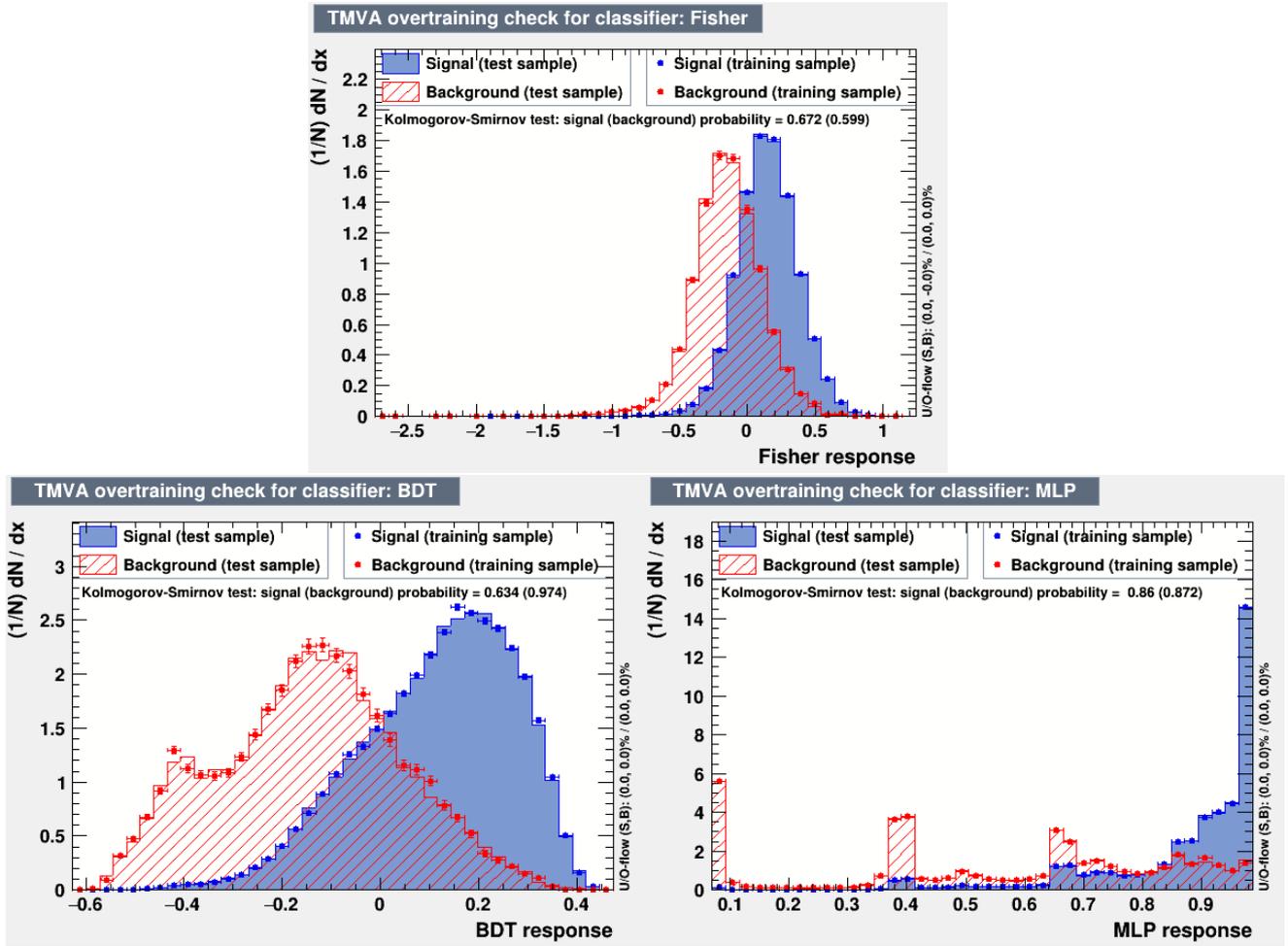
Figure 3.13: *Classifier Output Distributions for the three algorithms used with NLO events: FLD, BDT and MLP.*

Algorithm	Integral
Boosted Decision Trees	0.853
Multilayer Perceptron	0.850
Fisher Linear Discriminant	0.820

Table 3.6: *Algorithms ranked according to their performance at NLO, thanks to the ROC integral (right).*

As one can see, also from the integral of each ROC curve (shown in Table 3.6), the performance is significantly better. To highlight this we plotted the just obtained ROC curves with those resulting from a classification process using only kinematic variables. The comparison can be seen in Fig. 3.14, and the difference is evident.
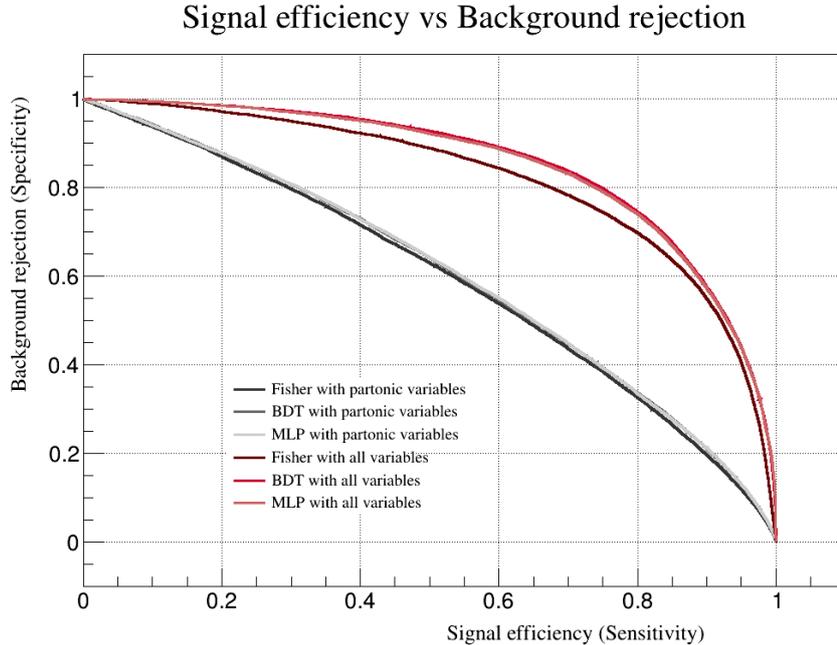
Figure 3.14: *Plots of two different sets of ROC curves at NLO. In grey one can see the performance of the different alogrithms using only kinematic variables, the same ones selected at Parton level. In red we have the result of the discriminating process with the addition of new variables.*

From Table 3.6 we conclude that the best performing algorithm is the *Boosted Decision Trees* (BDT), and as an example of one application of this analysis we evaluated the increase of purity of the sample after the classification process. In the starting sample of all the events, 81% were of type gg. We evaluated that after placing a cut on the BDT response variable at $t_{cut} = 0$ the purity increases from 0.81 to 0.92, at the cost of having an efficiency that goes from 1 to 0.74.

3.2.4 Alternative approach

Another approach to the problem can be formulated examining the plot of the $t\bar{t}$ system transverse momentum in Fig. 3.9. As one can see the three distributions displayed differ significantly in shape, and a pair produced through gluon-quark fusion has a much higher p_T than those from gg and $q\bar{q}$ events. In order to increase the purity of the sample one can place a cut on a high value of p_T, thus discarding most of the gg events, and performing the classification process on the remaining ones. This has the advantage of excluding the part of the sample with variables from the background that behave similarly to those from the signal. After the classification we can then choose the best

performing algorithm to further select a portion of the sample.

The first cut, on the $t\bar{t}$ pair transverse momentum, was chosen at 120 GeV. Just from this we obtain a 0.88 purity, while reducing the efficiency to 0.82. Then we train the TMVA methods, and we evaluate that their performance is very close to what described above. We show here the ROC curves and their integrals, in Fig. 3.15 and Table 3.6.

Algorithm	Integral
Boosted Decision Trees	0.852
Multilayer Perceptron	0.850
Fisher Linear Discriminant	0.831

Table 3.7: *Algorithms ranked according to their performance at NLO after a cut on $p_T = 120$ GeV, thanks to the ROC integral (right).*
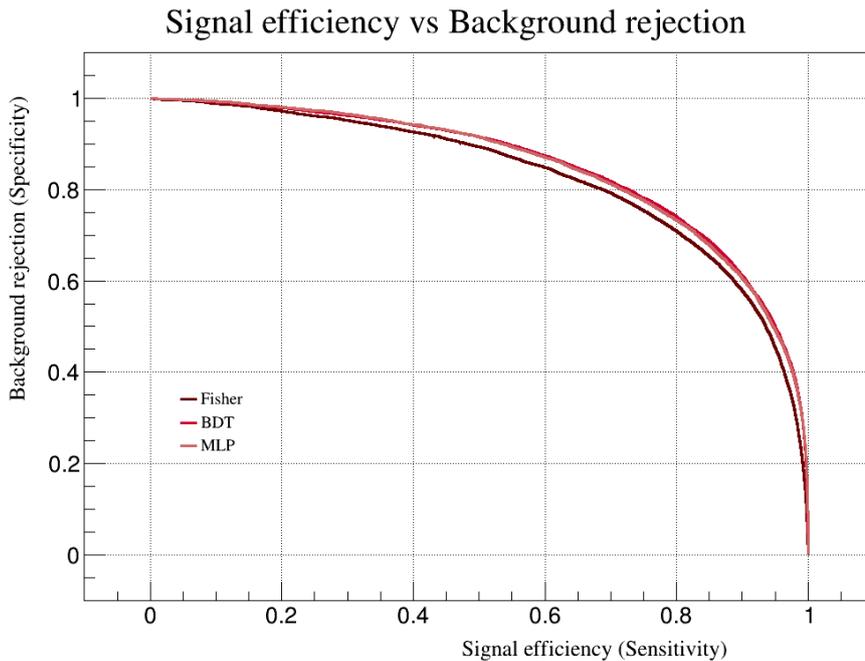

Figure 3.15: *Plots of the three ROC curves at NLO, evaluating the classification process after a cut on $p_T = 120$ GeV.*

After this process, the same cut for the BDT response variable at $t_{cut} = 0$ further increases the purity, going from 0.88 to 0.93. However the downside is the loss in efficiency, that results to be at 0.63.

Conclusions

This work presents the results of a multivariate approach to problem of discriminating the three $t\bar{t}$ pair production channels:

$$g\,g \;\to\; t\,\bar{t} \tag{3.2.1}$$

$$q\,\bar{q} \;\to\; t\,\bar{t} \tag{3.2.2}$$

$$g\,q \;\to\; t\,\bar{t}\,g\,(q)\;. \tag{3.2.3}$$

In order to do so, events were generated through Monte Carlo methods, at Leading Order only at Parton level as well as at Next to Leading Order both at Parton and at Particle level, and different observables were selected to help the classification process.

The multivariate analysis was carried out using the *Toolkit for Multivariate Analysis*, present in the ROOT package, selecting three of the available methods: the *Fisher Linear Discriminant*, the *Boosted Decision Trees* and the *Multilayer Perceptron*. Their performance was then evaluated in order to select the best one.

At Parton level three kinematic variables were selected: the invariant mass of the system, $m_{t\bar{t}}$, the absolute value of the rapidity of the $t\bar{t}$ pair, $|y_{t\bar{t}}|$ and the production angle θ. These alone lead to a poor classification, and we conclude an analysis at Parton level is not enough to obtain substantial results.

At NLO, in addition to the observables previously chosen, two more were evaluated: the transverse momentum of the $t\bar{t}$ system and the number of jets in the final state. With this addition the performance of the classifiers improves significantly, and the best performing algorithm turns out to be the Boosted Decision Trees, with a ROC integral of 0.853. After a cut on the BDT response variable at $t_{cut} = 0$ the purity of the sample increases from 0.81 to 0.92, at the cost of an efficiency of 0.74.

The thesis here presented constitutes just a first approach to the problem, hoping to improve the results and to provide a tool for discrimination on a real data sample, maybe expanding the work with a Detector level analysis.

Appendix A

Event generation

Hard events featuring a top pair production and up to 2 extra QCD partons are generated with Madgraph5_aMC@NLO v3.3.2 [13] at NLO QCD accuracy $\mathcal{O}(\alpha_s^3)$ within the Standard Model, in the 5 flavor scheme.

The hadronic center of mass energy is set to $\sqrt{s} = 13.0$ TeV, corresponding to the second run of the LHC. Electroweak corrections, known to be at the percent level in inclusive observables such as the total cross section [14], are neglected. The merging of samples with different jet multiplicites is handled by aMC@NLO [15] according to the FxFx prescription [16], with a jet p_T cut at the generator level set to 50 GeV and a merging scale of 100 GeV.

We use the NNPDF3.1 parton distribution [17] at NLO accuracy with $\alpha_s(m_Z) = 0.118$, available in the LHAPDF library [18] with ID 303400. The top mass is set to $m_t = 172$ GeV, and other SM input parameters are set to their recent experimental average [1]. The renormalization and the factorization scales are set on an event-by-event basis to one half of the total transverse mass of the hard interaction.

The electroweak decay of tops is handled by the MadSpin tool [19], that preserves both spin correlation and finite width effects to tree level accuracy. To facilitate event reconstruction, we decay the top pair in a different-flavor dilepton final state, and take CKM matrix to be the identity, so all events contain a $b\,\bar{b}$ pair.

Colored partons are showered using Pythia v8.244 [20]. The shower is only used to model color recombination, so the showering stops when a color neutral state is reached, and all electroweak effects are neglected. During the shower, the W^+tb and $W^-\bar{t}\bar{b}$ vertices are identified, the parton-level top momenta are defined as $p_W + p_b$, as appearing in the vertex.

Particles emerging from the shower are clustered into anti-kt jets [21] with $\Delta R = 0.4$.

Bibliography

[1] Particle Data Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022 (2022). http://pdg.lbl.gov. DOI: 10.1093/ptep/ptaa104.

[2] Claudio Severi et al. "Quantum tops at the LHC: from entanglement to Bell inequalities". In: *The European Physical Journal C* 82.4 (Apr. 2022). DOI: 10.1140/epjc/s10052-022-10245-9.

[3] S. P. Alamdari. "First Measurement of $\sigma(gg \to t\bar{t})/\sigma(p\bar{p} \to t\bar{t})$". PhD thesis. University of Toronto, 2008. DOI: 10.2172/929118.

[4] CDF Collaboration. "First measurement of the fraction of top-quark pair production through gluon-gluon fusion". In: *Physical Review D* 78.11 (Dec. 2008). DOI: 10.1103/PhysRevD.78.111101.

[5] Sylvie Braibant, Giorgio Giacomelli, and Maurizio Spurio. *Particelle e Interazioni Fondamentali: il mondo delle particelle*. Springer Milan, 2012.

[6] CDF Collaboration. "Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab". In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2626–2631. DOI: 10.1103/physrevlett.74.2626.

[7] DØ Collaboration. "Observation of the Top Quark". In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2632–2637. DOI: 10.1103/physrevlett.74.2632.

[8] J. H. Kuehn. *Theory of Top Quark Production and Decay*. 1997. DOI: 10.48550/ARXIV.HEP-PH/9707321.

[9] CMS Collaboration. "Measurement of the ratio $B(t \to Wb)/B(t \to Wq)$ in pp collisions at $\sqrt{(s)} = 8$ TeV". In: *Physics Letters B* 736 (Sept. 2014), pp. 33–57. DOI: 10.1016/j.physletb.2014.06.076.

[10] G. Mahlon and S. Parke. "Spin Correlation Effects in Top Quark Pair Production at the LHC". In: *Physical Review D* 81 (2010). http://arxiv.org/abs/1001.3422. DOI: 10.1103/PhysRevD.81.074024.

[11] M. Wolter. "Multivariate Analysis Methods in Physics". In: *Physics of Particles and Nuclei* 38.2 (2007), pp. 255–268. DOI: 10.1134/S1063779607020050.

[12] K. Albertsson et al. "TMVA - Toolkit for Multivariate Data Analysis". In: (2020). DOI: 10.48550/arXiv.physics/0703039.

[13] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). DOI: 10.1007/jhep07(2014) 079.

[14] Werner Bernreuther, Michael Fücker, and Zong-Guo Si. "Weak interaction corrections to hadronic top quark pair production". In: *Physical Review D* 74.11 (Dec. 2006). DOI: 10.1103/physrevd.74.113005.

[15] *FxFx merging with MadGraph5_aMC@NLO*. URL: http://amcatnlo.web.cern. ch/amcatnlo/FxFx_merging.htm.

[16] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). DOI: 10.1007/jhep12(2012) 061.

[17] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (Oct. 2017). DOI: 10.1140/epjc/s10052-017-5199-5.

[18] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *The European Physical Journal C* 75.3 (Mar. 2015). DOI: 10.1140/epjc/s10052-015-3318-8.

[19] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *Journal of High Energy Physics* 2013.3 (Mar. 2013). DOI: 10.1007/jhep03(2013)015.

[20] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01. 024.

[21] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063.

Ringraziamenti

Questo lavoro non sarebbe stato possibile senza l'aiuto dei miei relatori: ringrazio Maximiliano Sioli per la sua infinita pazienza e disponibilità in questi mesi di lavoro, e per avermi coinvolto in questo progetto che ho apprezzato molto. Ringrazio anche Rita FIoresi, per l'aiuto che mi ha fornito quando mi sono dovuto approcciare per la prima volta allo studio delle reti neurali e del machine learning, ma anche alla matematica e all'algebra lineare nel primo anno. Menzione d'onore va a Claudio Severi per aver avuto l'idea che ha fatto nascere questa tesi, e per il suo supporto per quanto riguardasse i dettagli più tecnici del lavoro qua presentato. Mi sono molto divertito in questo primo viaggio nel mondo della ricerca, e spero vivamente di poter continuare su questa strada.

Devo anche sicuramente ringraziare tutti i miei amici, sia i compagni del Club delle Algebre Eccezionali di Lie che i membri della Società Italiana di Geologia, così come tutti i colleghi del corso e i coinquilini passati. Ho apprezzato la compagnia di ognuno di voi in questi anni, che fosse per presentarsi alle 9 in punto su Teams per ripassare o per le escursioni clandestine sul tetto, e mi mancherà avervi accanto mentre proseguiamo i nostri studi. Spero di essere riuscito almeno in parte a ricambiare l'affetto che mi avete trasmesso e l'aiuto che mi avete fornito.

Ringrazio anche le persone che hanno reso possibile la mia permanenza in questa città e i miei studi, i miei genitori, le mie sorelle e la mia famiglia. Grazie della fiducia che mi avete dato e del sostegno prezioso.