CHRISTIAN MICHELSEN
NIELS BOHR INSTITUTE
UNIVERSITY OF COPENHAGEN

A PHYSICIST'S APPROACH TO MACHINE LEARNING
–
UNDERSTANDING THE BASIC BRICKS

SUPERVISOR:
TROELS PETERSEN
NIELS BOHR INSTITUTE
UNIVERSITY OF COPENHAGEN

First printing, December 2019

Abstract

Here will be a decent abstract at some point$^{\text{TM}}$.

Contents

Foreword

Part I

The first part of this thesis deals with the introductory theory of machine learning and its predictive power in estimating Danish housing prices.

This subproject was done in collaboration with Boligsiden without whom it would not have been possible. During this project, common python data science tools from the SciPy ecosystem[83] such as NumPy, Matplotlib, Pandas, Scikit-Learn, Scipy has been used extensively and should thus also be mentioned.

Part II

The second part of this thesis deals with particle physics and the discriminatory power of machine learning for quark-gluon identification and subsequent analysis.

4. *Particle Physics and LEP*

> *"Not only is the Universe stranger than we think, it is stranger than we can think."*
>
> — Werner Heisenberg

The aim of this chapter is to introduce the reader to the level of particle physics required for understanding the following chapter, in particular introducing the Standard Model in section 4.1, the theory behind quark hadronization in section 4.2, and the ALEPH detector at LEP in section 4.3. The goal is not to make a deep and thorough introduction to the field as this is not needed for the following analysis along with the fact that the author is no particle physicist himself.

4.1 *The Standard Model*

The *Standard Model* (SM) [41, 70, 85] of particle physics is the currently best known description of the elementary particles and thus describes the fundamental building blocks of our Universe. An overview of the particles explained by the Standard Model is shown in the typical tabular form seen in Figure 4.1. In general, particles comes in two categories: *bosons* and *fermions*.

The fermions, the left part of the figure, are particles with half-integer spin that obey Fermi-Dirac statistics and are further subdivided into *quarks* (upper left in figure) and *leptons* (lower left). The quarks interact with all of the four known forces[1], including the strong force. In contrary the leptons do not interact with the strong force. Quarks are never observed freely but are always combined into *hadrons* due to *color confinement* which is further explained in section 4.2. An example of this are protons which consists of two up-quarks and a down-quark. Leptons exist as either the charged leptons[2] or as neutral leptons, the so-called neutrinos[3]. The fermions come in three generations with increasing mass.

The bosons, the right part of the figure, are the force-carrying particles (with integer spin and which obey Bose-Einstein statistics) where the gluon g mediates the strong nuclear force (color charge), the photon γ mediates the electromagnetic force (charge), and the two $W^{\pm}$ and the Z bosons the weak nuclear force (weak isospin). The Higgs boson H, experimentally discovered in 2012 [33, 34],

[1] Gravity, electromagnetism, and the strong and weak force.

[2] The electron e, the muon μ, and the tau τ.

[3] The electron neutrine ν_e, the muon neutrino ν_μ, and the tau neutrino ν_τ.

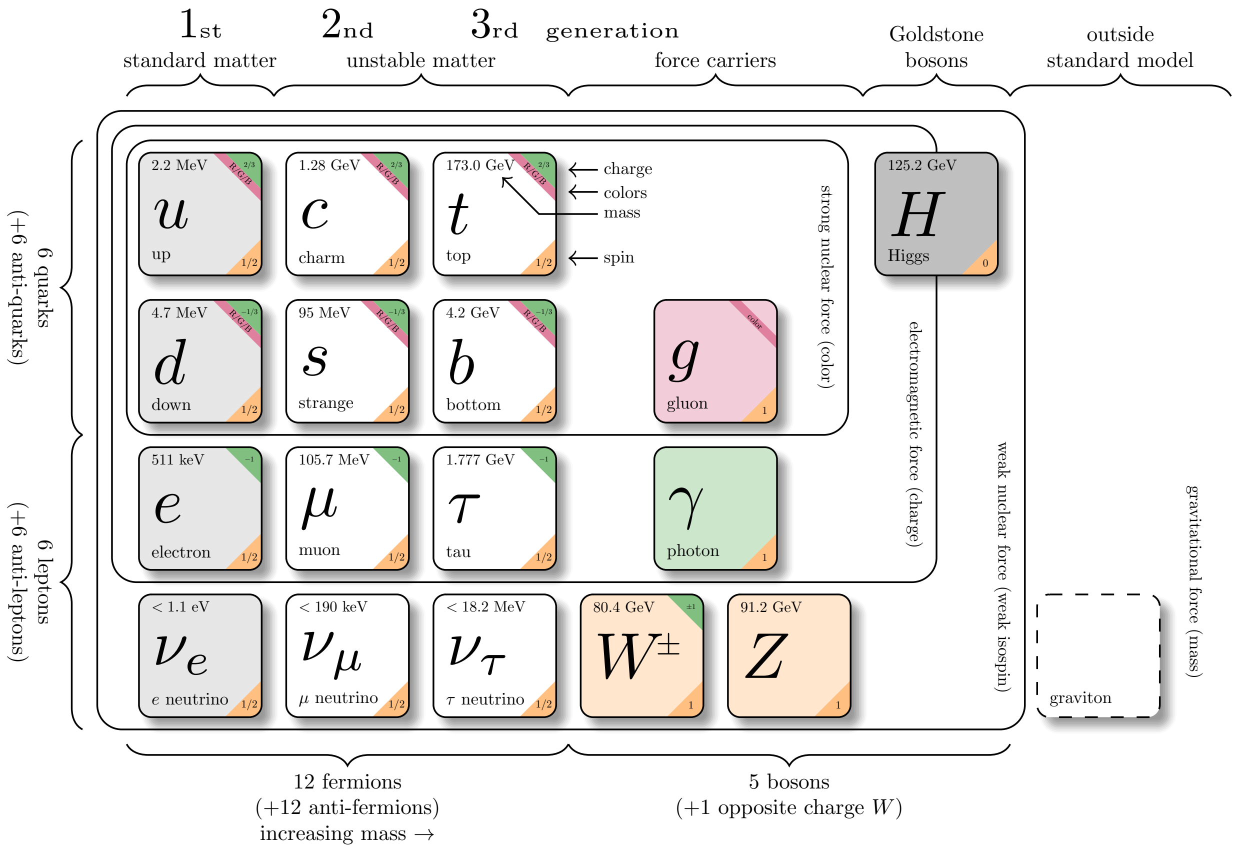

Figure 4.1: The Standard Model. Inspired by Purcell [65] using the template by Burgard [30] with manually updated masses according to Particle Data Group et al. [61].

does not mediate any forces but interacts with all massive particles and explains why particles have mass.

All particles have antiparticles which are particles with opposite charge but the same mass. Some particles are their own antiparticles[4], such as the Z. At the Large Electron Positron collider (LEP), see section 4.3, electrons e^- and their antiparticles positrons e^+ were collided at an energy of around 91 GeV. This particular energy was chosen since this is at the resonance peak of the Z. Its mass distribution follows a Cauchy distribution (also known as Breit-Wigner) with mean[5] $m_Z = (91.1876 \pm 0.0021)$ GeV and a full width of $\Gamma_Z = (2.4952 \pm 0.0023)$ GeV: LEP was as such a Z-factory. The Z, however, is only very short-lived with a half-life of $1/\Gamma_Z \sim 2.6 \times 10^{-25}$ s. The decay mode for this unstable Z particle is primarily to hadrons ((69.91 ± 0.06) %) where the ratio (R) for b-quarks is $R_b = (Z \to b\bar{b}) = (15.12 \pm 0.05)$ % and $R_g = (Z \to ggg) < (1.10 \pm 0.05)$ % for gluons [61]. The fact that the Z is neutral and its own anti-particle means that it generally decays to a particle–anti-particle pair (due to charge-conservation). Antiparticles are written with a bar on top, e.g. the $\bar{b}$-quark is the antiparticle of the b-quark.

[4] The photon, the Z, and the Higgs.

[5] Calculated in natural units where $c = \hbar = 1$ which will also be used throughout this thesis.

4.2 Quark Hadronization

The electron-positron e^+e^- annihilations at LEP are complicated events that require advanced high-energy particle physics theory to be properly understood. Most of the aspects of the process is well-described by now, however, especially the hadronization process is still an area of active research. To better get an overview of the different stages of the e^+e^- annihilations, see the Feynman diagram in Figure 4.2.

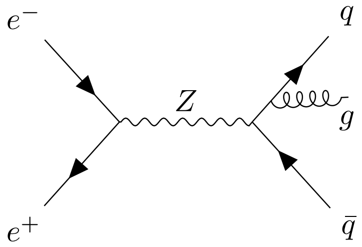

Figure 4.2: Feynman diagram showing the $e^+e^- \rightarrow Z^0$ production at LEP. The Z has several decay modes where the $Z \rightarrow q\bar{q}g$ is shown here.

Reading from left to right, the electron and the positron annihilates to a Z. This interaction is well-described by quantum electrodynamics (QED), a theory that has been around for more than 60 years by now. As mentioned in the previous section, the Z has several decays modes, yet most of these are background processes of no interest in this project and the focus for now will be the decay mode $Z \rightarrow q\bar{q}$ (Z to quark–anti-quark) as seen in the Feynman diagram. The particles produced by the Z-decay are called primary *partons*. Since this process involves quarks, and thus color charge, QED is no longer an adequate theory: quantum chromodynamics (QCD) is needed [16]. The $q\bar{q}$ pairs in this example acts as (color) dipoles from which a gluon can radiate. It can be shown with QCD that the gluon can only be radiated inside the cone that the $q\bar{q}$ pairs spans [24]. As mentioned in the introduction, quarks cannot exist freely (due to *confinement*) and we therefore cannot observe the individual partons in a $q\bar{q}g$ event produced in the Feynman diagram. Confinement is basically the QCD principle saying that quarks are always confined or bound inside hadrons. The initial partons (carrying color charge) are converted to (color-neutral) hadrons by non-perturbative QCD processes in what is called *hadronization*, and these hadrons can be measured.

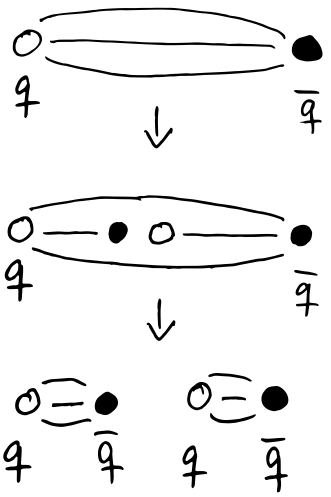

Figure 4.3: Illustration of the quarks splitting as explained by the Lund string model. For large charge separation the (color) field lines seem to be compressed to a tube-like region, where the strong interactions are mediated by the massless gluons (that couple to the color charge of quarks). When the two quarks are separated enough, the potential energy is released by the production of a new $q\bar{q}$ pair.

The hadronization process is not yet fully modelled and currently two competing models for predicting the hadronization pattern exists: the Lund string model and the cluster model. In this project only the former of the models will be used. The Lund string model [15] is the theoretical framework underlying the widely used Monte Carlo event generator PYTHIA [73]. The string model is based on the observation that (color) field lines between quarks seem to compress into a tube-like region mediated by gluons, see the top part of Figure 4.3. The field can be described by a linearly rising potential $V(r) = \kappa r$ at large distances[6], where r is the distance and κ is the strength of the potential [29]. This field is similar to the (constant) force of a string: $V(r) = \kappa r \Rightarrow F(r) = -\kappa$ where κ is the to be regarded as the spring tension. As quarks move apart, the potential energy stored in the "string" increases until it is large enough to "snap" and convert its potential energy into mass. This mass energy is released with the production of a new $q\bar{q}$ pair as this energetically favorable, see the rest of Figure 4.3.

An example of the hadronization process, or the transition from initial partons to final hadrons is sketched in Figure 4.4. Here the

[6] At small distances a Coulumb term has to be included, however, this term is assumed to be negligible by the Lund string model.

production of two kaons K^- and K^+, and two pions π^- and π^0 are shown. Since particles are created by "splits" in the "string", and the fact that there is energy-momentum conservation, they all have to share the total energy stored in the string. This is described by the fragmentation function:

$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(-\frac{bm^2}{z}\right), \tag{4.1}$$

where $0 \leq z \leq 1$ is the remaining momentum that the new hadron takes, a and b are constants, and m is the mass[7] [24]. When the system runs out of available momentum, it will stop producing new hadrons and the fragmentation function thus explains the distribution of final state particles. The Lund string model can be extended from only $q\bar{q}$ events to $q\bar{q}g$ events where it predicts cones spanning the angular regions qg and $\bar{q}g$ should receive enhanced particle production compared to the $q\bar{q}$ region. This prediction by the Lund string model is also measured in e^+e^- collisions [29].

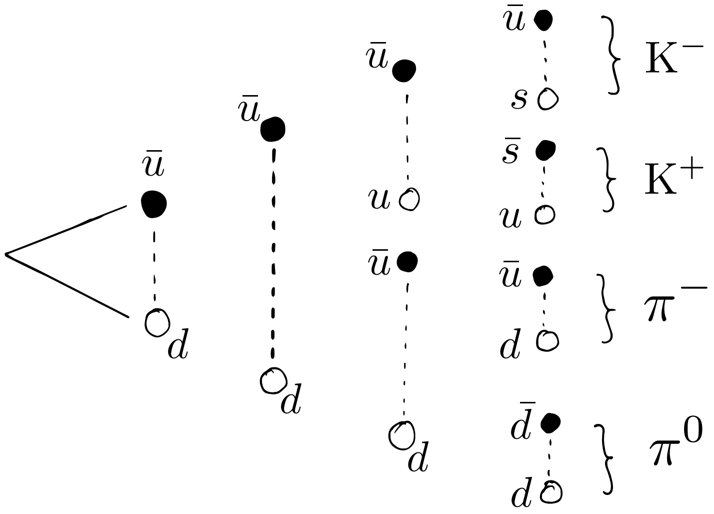

Figure 4.4: Illustration of the hadronization process by which $\bar{u}$- and d-quarks decay into four different mesons. The theoretical strings are shown as dashed lines and particles as circles, where filled circles are antiparticles.

The initial partons produced as Z decay therefore decay to final state hadrons[8] which create a whole "shower" in the direction of the initial parton: this is called a *parton shower* and it is this parton shower observed as particles, a *jet*, that is measured in the detector. The reverse computation from tracks measured in the detector is done with the use of *jet clustering* algorithms. The detector and the clustering algorithms are described in the following section.

4.3 *The ALEPH Detector and LEP*

The Large Electron Positron collider (LEP) was a particle collider at CERN in Switzerland operating from 1989 to 2000. It collided counter-rotating bunches of electrons and positrons in a giant ring with a circumference of more than 26 km. The first phase, LEP1, ran from 1989 to 1995 at the Z resonance 91 GeV and the second phase, LEP2, continued afterwards closer to 200 GeV for W^+W^-

[7] Where $m \rightarrow m_\perp$ for particles with transverse momentum.

[8] To either mesons which consist of two quarks (color–anti-color) or baryons (r-g-b) which consist of three quarks.

pair production [16], however, it is only the data collected at the energy around $\sqrt{s} = 91.3\,\text{GeV}$ called the *Z peak data* that is used throughout the rest of this project. There were four independent detectors at the LEP experiment, one of them ALEPH[9].

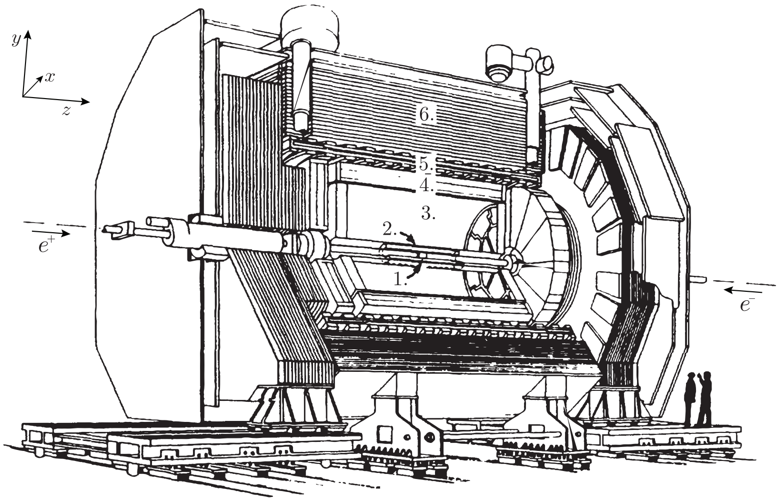

Figure 4.5: The ALEPH detector at LEP. 1) Vertex detector (VDET). 2) Drift chamber (ITC). 3) Time projection chamber (TPC). 4) Electromagnetic calorimeter (ECAL). 5) Superconducting magnet coil. 6) Hadron calorimeter (HCAL). Adapted from Buskulic et al. [31].

The *apparatus for LEP physics* (ALEPH) was a particle detector at LEP with a wide coverage, almost 4π, consisting of cylindrical subdetectors, see Figure 4.5, with the coordinate system shown in the upper left corner[10]. The polar angle θ is illustrated in Figure 4.6 together with the transverse (longitudinal) momentum $p_\perp$ (p_L) and the azimuthal angle ϕ in Figure 4.7. The ALEPH detector was designed to measure the energy deposited in calorimeters by charged and neutral particles, measure the momenta of charged particles, measure the distance of travel of short-lived particles, and to identify the three lepton flavors (electron, muon, tau) [31]. As can be seen in Figure 4.5, ALEPH consisted of five subdetectors (the vertex detector (VDET), the drift chamber (ITC), and the time projection chamber (TPC)) and two calorimeters (the electromagnetic (ECAL) and the hadronic calorimeters (HCAL)).

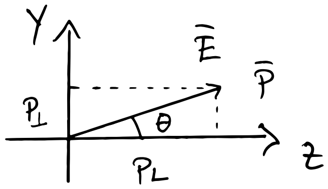

Figure 4.6: The polar angle θ defined in the zy coordinate system

The three innermost detectors allow for precise tracking of the charged particles produced in the parton shower and the two outer calorimeters of precise energy measurements for both charged and neutral particles going through the detector.

A hadronic event from a parton shower may leave a score of charged tracks resulting in hundreds of hits in the detectors (VDET, ITC, and TPC) which are fitted[11] with Kalman filters [50] to obtain global track fits, of which bad charged tracks are discarded for further analysis. The tracks are helical due to the presence of a 1.5 T magnetic field which curves the charged particles according to their transverse momentum, $p_\perp$.

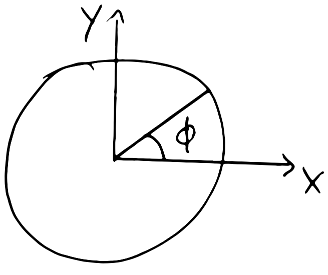

Figure 4.7: The azimuthal angle ϕ defined in the xy coordinate system.

The energy resolution σ of the calorimeters, or the *calorimeter performance*, is expected to increase with $\sqrt{E}$. In fact, it was found at ALEPH that the energy dependence of the resolution follows the

[9] Together with DELPHI, L3, and OPAL.

[10] The z-axis pointing along the beam direction, the y-axis pointing upwards, and the x-axis pointing towards the center of LEP.

[11] the process of fitting tracks is called *track reconstruction* in high energy particle physics.

parametrization [31]:

$$\sigma(E) = \Big((0.59 \pm 0.03) \cdot \sqrt{E/\text{GeV}} + (0.6 \pm 0.3)\Big)\,\text{GeV}. \tag{4.2}$$

Even though $\sigma(E)$ increases with E, the relative resolutions improves with higher energies. Since one never measures Nature directly, the results one obtains in a measurement are thus products of both model and experimental uncertainties folded together. To unfold the measurements to obtain experiment-independent results, the uncertainties are important to understand. Of course there are dozens of other uncertainties in an advanced experiment like ALEPH, however, the energy dependence is the primary focus in this project.

4.4 *Jet clustering*

Since the initial partons created as decay products from the Z are unstable themselves, what is measured in the detector is a whole shower of hadrons seen as charged tracks in the detectors and energy deposits in the calorimeters. However, say that the Z decayed to a $b\bar{b}$ event. In this case the two b's would be back-to-back and the final hadrons would be observed approximately in the same direction as the b's were created. The interest of the experiment is not to measure the final hadrons, but rather to infer information about the initial quarks and gluons. This is done via the reverse-engineering process called *jet clustering*. Over the years many clustering algorithms have been developed, however, most of these are younger than LEP. In the ALEPH experiment the JADE algorithm was used [20]. JADE is a sequential recombination algorithm where final state particles are initially described as individual so-called pseudo-jets which are then recursively merged to larger jets according to their inter-jet distance d_{ij}^2. The distance measure for JADE is:

$$d_{ij}^2 = \frac{2E_iE_j(1-\cos\theta_{ij})}{E_{\text{vis}}^2}, \tag{4.3}$$

where E_{vis} is the visible energy[12] and θ_{ij} is the angle between jet i and j. The JADE algorithm computes d_{ij}^2 for all combinations of jets and merges the two jets with the lowest d_{ij}^2, continuing like that recursively until $\min(d_{ij}^2) > d_{\text{cut}}^2$ for some predefined value of d_{cut}^2. In the dataset at hand, only the final jets were available and not the jet constituents, unfortunately.

4.5 *The variables*

The overall goal of the project is to be able to discriminate quarks and gluons using only vertex variables. The reason for the last condition is that the goal is to better understand the shape distributions of gluons in which there is still significant differences between Monte Carlo (MC) simulations and Data. Therefore only vertex

[12] The total sum of energies in the event.

variables will be used to avoid any biases introduced by using shape-related variables to detect differences in shape-distributions. The vertex variables are a subset of all variables which include the three variables `projet`, `bqvjet`, and `ptlrel`. These three particular variables have each shown discriminatory power in separating b-quarks from light quarks and gluons.

`projet`: Probability of significant lifetime. For each track in the jet an impact parameter δ is computed. This parameter is the minimum distance between the estimated Z decay point and the track itself and its sign depends on whether or not the point of closest approach is in front of or behind the Z decay point (relative to the momentum). From δ the significance $\mathcal{S}$ – which is δ/σ_δ – is computed and is thus a measure of the certainty of a measured track being from primary vertex. High values of $\mathcal{S}$ is typically an indicator of b jets, since long-lived particles typically decay in front of the Z relative to the jet direction, while *uds*-jets generally have small significance and might as well have negative values of $\mathcal{S}$. An illustration of the difference in significance between *uds*-jets and b-jets can be seen in Figure 4.8. From $\mathcal{S}$ the track probability $\mathcal{P}_{\text{track}}$ of a track originating at the decay point of the Z can be computed, which can further be aggregated across all tracks within a jet to form the jet probability $\mathcal{P}_{\text{jet}}$ which `projet` is a function of [37]. Whether or not $\mathcal{P}_{\text{jet}}$ is strictly a probability can be discussed but it is related to the probability of all tracks within a jet to originate from long-lived particles, which itself is a good indicator of being a b- (or c-) jet. This variable further has the advantage of being independent of any vertex algorithm.

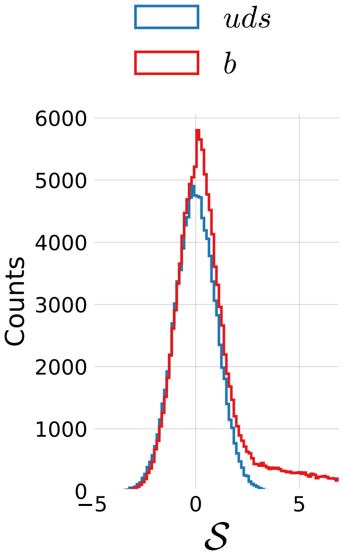

Figure 4.8: Distribution showing the difference in significance $\mathcal{S}$ between *uds*-jets and b-jets. Based on own, simulated data to illustrate this difference.

`bqvjet`: b-quark vertex of jet. For any jet with well measured[13] charged tracks, a fit with a (hypothetical) secondary vertex is performed. The difference in χ^2 between the null hypothesis that all good tracks originate from the same primary vertex and the alternative hypothesis that a secondary vertex exists in addition to the primary one is calculated. For the long-lived massive b and c quarks this typically results in large differences in χ^2 compared to *uds*- and gluon jets which have much lower $\Delta\chi^2$-values [16]. The `bqvjet` is related to the $\Delta\chi^2$-value from the secondary vertex algorithm. This value is dependent of the vertex algorithm, but still explores other areas of phase space than `projet`, however, they are still very correlated. The linear correlations[14] ρ_{q_i} between `projet` and `bqvjet` for q_i jets are $\rho_b = 0.80, \rho_c = 0.65, \rho_{uds} = 0.23, \rho_g = 0.29$.

`ptlrel`: Relative Lepton Momentum. If any leptons (in the case of $e^\pm$ or $\mu^\pm$) are measured in the jet by the detector, this is a good sign of the jet originating from a b-quark as

[13] Meaning that there are at least four TPC hits and the fit has a reduced χ^2 of less than four [16].

[14] Based on MC truth.

$\sim 11\,\%(e) + \sim 11\,\%(\mu)$ decay semi-leptonically [13][15]. The high mass of the b quark leads to high $p_\perp$ for the leptons relative to the jet axis which is exactly measured by `ptlrel`.

The fact that the heavy b-quarks have much longer lifetimes than the lighter *uds*-quarks stems from their much lower coupling magnitudes written as the CKM matrix $\mathbf{V}$ [61]:

$$\mathbf{V} = \begin{array}{c} \\ u \\ c \\ t \end{array} \begin{array}{c} \begin{array}{ccc} d & s & b \end{array} \\ \begin{pmatrix} 0.97446 & 0.22452 & 0.00365 \\ 0.22438 & 0.97359 & 0.04214 \\ 0.00896 & 0.04133 & 0.99911 \end{pmatrix} \end{array}. \tag{4.4}$$

The matrix element $|V_{ij}|^2$ is proportional to the transition-probability of quark i transitioning to quark j. From the CKM matrix it can be seen that u and d quarks couples strongly together, likewise with c-s and b-t quark pairs. When a Z decays into a b-quark, this quark couples strongly with the top quark, however, due to the high mass of the top quark compared to the b-quark, the b-quark cannot decay into a t-quark but must (almost always) decay to a c, however, still with low probability, $V_{bc} \ll 1$. This, together with the fact that $V_{bu} \ll V_{bc}$ explains the long life-time of b quarks, $\tau_b \sim 1.3 \times 10^{-12}\,\mathrm{s}$ [68]. This is also why the three variables above are very common variables for b-tagging algorithms. That c-quarks also have relative long life-times, $\tau_c \sim 1.1 \times 10^{-12}\,\mathrm{s}$ [68], are not due to the CKM elements, as for b-quarks, but rather due to the c-decay being governed by the weak force through virtual W^* bosons, a force that is much weaker than the strong force (hence the name). The low phase space in a c-quark decay makes the c-quark longer-lived. This also happens for b-quarks which further explains why c-quarks share many similarities with b-quarks but also resembles resembles light-quarks (which are very long-lived.).

The rest of the non-vertex variables are:

`ejet`: The energy of the jet E_{jet}

`costheta`: The cosine of the θ angle defined in Figure 4.6.

`phijet`: The angle ϕ of defined in Figure 4.7: ϕ.

`sphjet`: The sphericity tensor $\mathbf{S}$ is defined as:

$$S^{(\alpha\beta)} = \frac{\sum_{i=1}^{N} p_i^{(\alpha)} p_i^{(\beta)}}{\sum_{i=1}^{N} |p_i|^2} \quad \alpha, \beta \in \{x, y, z\}, \tag{4.5}$$

and the sphericity is determined as $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ where $\lambda_1 \geq \lambda_2 \geq \lambda_3,\ \lambda_1 + \lambda_2 + \lambda_3 = 1$ are the three eigenvalues of the sphericity tensor. The sphericity $0 \leq S \leq 1$ is a measure of the angular distribution of the tracks and clusters in a jet. When $S = 0$ the jets form a perfect sphere, compared to $S = 1$ for a perfect line. The `sphjet` variable is the sphericity of the jet when calculated in its boosted rest frame, also known as *boosted sphericity*.

[15] $\mathcal{B}(B \to l\nu X) = (10.5 \pm 0.3)\,\%$, $l = \{e, \mu\}$ [13].

`pt2jet` : The sum of the square of transverse momentum w.r.t. the jet axis: $\sum_i p^2_{\perp,i}$.

`muljet` : The rescaled multiplicity of the jet.

For further details about the variables, see Armstrong [16].

The variables explained above are all used in the following analysis where the machine learning model is trained on only the vertex variables to probe differences in the shape-variables. The goal of this is to better understand the gluon hadronization process to minimize differences in MC simulations and ultimately get a better understanding of the rules governed by Nature.

B. Quarks vs. Gluons Appendix

List of Figures

List of Tables

Bibliography

[1] Advanced Topics in Machine Learning (ATML). URL https://kurser.ku.dk/course/ndak15014u.

[2] Allstate Claims Severity - Fair Loss. URL https://kaggle.com/c/allstate-claims-severity.

[3] Dmlc/xgboost. URL https://github.com/dmlc/xgboost.

[4] HEP meets ML award | The Higgs Machine Learning Challenge. URL https://higgsml.lal.in2p3.fr/prizes-and-award/award/.

[5] The Large Electron-Positron Collider | CERN. URL https://home.cern/science/accelerators/large-electron-positron-collider.

[6] Microsoft/LightGBM. URL https://github.com/microsoft/LightGBM/blob/b3975555d7023fd05f8e56326905fe7b185109de/src/treelearner/serial_tree_learner.cpp#L282.

[7] Scikit-hep/uproot. URL https://github.com/scikit-hep/uproot.

[8] Datashader: Revealing the Structure of Genuinely Big Data. URL https://github.com/holoviz/datashader.

[9] O. . Www.OIS.dk - Din genvej til ejendomsdata. URL https://www.ois.dk/.

[10] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems. URL http://tensorflow.org/.

[11] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin. *Learning From Data*. AMLBook. ISBN 978-1-60049-006-4.

[12] D. Albanese, S. Riccadonna, C. Donati, and P. Franceschi. A practical tool for maximal information coefficient analysis. 7. ISSN 2047-217X. doi: 10.1093/gigascience/giy032. URL https://doi.org/10.1093/gigascience/giy032.

[13] H. Albrecht, H. Ehrlichmann, T. Hamacher, R. P. Hofmann, T. Kirchhoff, A. Nau, S. Nowak, H. Schröder, H. D. Schulz, M. Walter, R. Wurth, C. Hast, H. Kolanoski, A. Kosche,

A. Lange, A. Lindner, R. Mankel, M. Schieber, T. Siegmund, B. Spaan, H. Thurn, D. Töpfer, D. Wegener, M. Bittner, P. Eckstein, M. Paulini, K. Reim, H. Wegener, R. Eckmann, R. Mundt, T. Oest, R. Reiner, W. Schmidt-Parzefall, W. Funk, J. Stiewe, S. Werner, K. Ehret, W. Hofmann, A. Hüpper, S. Khan, K. T. Knöpfle, M. Seeger, J. Spengler, D. I. Britton, C. E. K. Charlesworth, K. W. Edwards, E. R. F. Hyatt, H. Kapitza, P. Krieger, D. B. MacFarlane, P. M. Patel, J. D. Prentice, P. R. B. Saull, K. Tzamariudaki, R. G. Van de Water, T. S. Yoon, D. Reßing, M. Schmidtler, M. Schneider, K. R. Schubert, K. Strahl, R. Waldi, S. Weseler, G. Kernel, P. Križnič, T. Podobnik, T. Živko, V. Balagura, I. Belyaev, S. Chechelnitsky, M. Danilov, A. Droutskoy, Y. Gershtein, A. Golutvin, G. Kostina, D. Litvintsev, V. Lubimov, P. Pakhlov, F. Ratnikov, S. Semenov, A. Snizhko, V. Soloshenko, I. Tichomirov, and Y. Zaitsev. A model-independent determination of the inclusive semileptonic decay fraction of B mesons. 318(2):397–404. ISSN 0370-2693. doi: 10.1016/0370-2693(93)90146-9. URL http://www.sciencedirect.com/science/article/pii/0370269393901469.

[14] E. Anderson. The Species Problem in Iris. 23(3):457–509. ISSN 00266493. doi: 10.2307/2394164. URL www.jstor.org/stable/2394164.

[15] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. 97(2):31–145. ISSN 0370-1573. doi: 10.1016/0370-1573(83)90080-7. URL http://www.sciencedirect.com/science/article/pii/0370157383900807.

[16] S. R. Armstrong. A Search for the standard model Higgs boson in four jet final states at center-of-mass energies near 183-GeV with the ALEPH detector at LEP. URL http://wwwlib.umi.com/dissertations/fullcit?p9910371.

[17] M. Awad and R. Khanna. Support Vector Regression. In M. Awad and R. Khanna, editors, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, pages 67–80. Apress. ISBN 978-1-4302-5990-9. doi: 10.1007/978-1-4302-5990-9_4. URL https://doi.org/10.1007/978-1-4302-5990-9_4.

[18] R. J. Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series)*. WileyBlackwell, reprint edition. ISBN 0-471-92295-1. URL http://www.amazon.co.uk/Statistics-Statistical-Physical-Sciences-Manchester/dp/0471922951%3FSubscriptionId%3D13CT5CVB80YFWJEPWS02%26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0471922951.

[19] J. T. Barron. A General and Adaptive Robust Loss Function. URL http://arxiv.org/abs/1701.03077.

[20] W. Bartel et al. Experimental study of jets in electron-positron annihilation. 101(1):129–134. ISSN 0370-2693. doi: 10.1016/0370-2693(81)90505-0. URL http://www.sciencedirect.com/science/article/pii/0370269381905050.

[21] E. Becht, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell. Evaluation of UMAP as an alternative to t-SNE for single-cell data. page 298430, . doi: 10.1101/298430. URL https://www.biorxiv.org/content/10.1101/298430v1.

[22] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell. Dimensionality reduction for visualizing single-cell data using UMAP. 37 (1):38–44, . ISSN 1546-1696. doi: 10.1038/nbt.4314. URL https://www.nature.com/articles/nbt.4314.

[23] J. Bergstra and Y. Bengio. Random Search for Hyper-parameter Optimization. 13:281–305. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2188385.2188395.

[24] C. Bierlich. Rope hadronization, geometry and particle production in pp and pA Collisions. URL https://lup.lub.lu.se/search/ws/files/18474576/thesis.pdf.

[25] Bolighed. Bolighed - usikkerhed i data-vurderingen. URL https://bolighed.dk/om-bolighed/spoergsmaal-og-svar/#boligvaerdi.

[26] L. Breiman. Random Forests. 45(1):5–32. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

[27] E. Brochu, V. M. Cora, and N. de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. URL http://arxiv.org/abs/1012.2599.

[28] R. Brun and F. Rademakers. ROOT — An object oriented data analysis framework. 389(1):81–86. ISSN 0168-9002. doi: 10.1016/S0168-9002(97)00048-X. URL http://www.sciencedirect.com/science/article/pii/S016890029700048X.

[29] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, H. Hoeth, F. Krauss, L. Lonnblad, E. Nurse, P. Richardson, S. Schumann, M. H. Seymour, T. Sjostrand, P. Skands, and B. Webber. General-purpose event generators for LHC physics. 504(5):145–233. ISSN 03701573. doi: 10.1016/j.physrep.2011.03.005. URL http://arxiv.org/abs/1101.2599.

[30] C. Burgard. Standard model of physics | TikZ example. URL http://www.texample.net/tikz/examples/model-physics/.

[31] D. Buskulic et al. An investigation of Bdo and Bso oscillation. 322(4):441–458. ISSN 0370-2693. doi: 10.1016/0370-2693(94)91177-0. URL http://www.sciencedirect.com/science/article/pii/0370269394911770.

[32] T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. pages 785–794. doi: 10.1145/2939672.2939785. URL http://arxiv.org/abs/1603.02754.

[33] T. A. Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. 716(1):1–29, . ISSN 03702693. doi: 10.1016/j.physletb.2012.08.020. URL http://arxiv.org/abs/1207.7214.

[34] T. C. Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. 716(1):30–61, . ISSN 03702693. doi: 10.1016/j.physletb.2012.08.021. URL http://arxiv.org/abs/1207.7235.

[35] A. Diaz-Papkovich, L. Anderson-Trocmé, C. Ben-Eghan, and S. Gravel. UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. 15(11). ISSN 1553-7390. doi: 10.1371/journal.pgen.1008432. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6853336/.

[36] S. D. DST. Price Index (EJ14) - Statistics Denmark. URL https://www.dst.dk/en/Statistik/emner/priser-og-forbrug/ejendomme.

[37] D. et al. Buskulic. A precise measurement of hadrons. 313(3): 535–548. ISSN 0370-2693. doi: 10.1016/0370-2693(93)90028-G. URL http://www.sciencedirect.com/science/article/pii/037026939390028G.

[38] F. Faye. Frederik Faye / deepcalo. URL https://gitlab.com/ffaye/deepcalo.

[39] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. 7(2):179–188. ISSN 2050-1439. doi: 10.1111/j.1469-1809.1936.tb02137.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x.

[40] Y. Freund and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In P. Vitányi, editor, *Computational Learning Theory*, pages 23–37. Springer Berlin Heidelberg. ISBN 978-3-540-49195-8. AdaBoost.

[41] S. L. Glashow. Partial-symmetries of weak interactions. 22(4): 579–588. ISSN 0029-5582. doi: 10.1016/0029-5582(61)90469-2.

URL http://www.sciencedirect.com/science/article/pii/0029558261904692.

[42] N. Guttenberg, N. Virgo, O. Witkowski, H. Aoki, and R. Kanai. Permutation-equivariant neural networks applied to dynamics prediction. URL http://arxiv.org/abs/1612.04530.

[43] A. E. Harvey and S. Peters. Estimation Procedures for Structural Time Series Models. doi: 10.1002/for.3980090203.

[44] T. Hastie and R. Tibshirani. Generalized Additive Models: Some Applications. 82(398):371–386. ISSN 01621459. doi: 10.2307/2289439. URL www.jstor.org/stable/2289439.

[45] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer-Verlag, 2 edition. ISBN 978-0-387-84857-0. URL //www.springer.com/la/book/9780387848570.

[46] K. Hornik. Approximation capabilities of multilayer feedforward networks. 4(2):251–257. ISSN 0893-6080. doi: 10.1016/0893-6080(91)90009-T. URL http://www.sciencedirect.com/science/article/pii/089360809190009T.

[47] P. Huber and E. Ronchetti. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley. ISBN 978-1-118-21033-8. URL https://books.google.dk/books?id=j1OhquR_j88C.

[48] S. Hviid, Juul. Working Paper: A regional model of the Danish housing market. URL http://www.nationalbanken.dk/en/publications/Pages/2017/11/Working-Paper-A-regional-model-of-the-Danish-housing-market.aspx.

[49] F. James and M. Roos. Minuit – a system for function minimization and analysis of the parameter errors and correlations. 10:343–367. doi: 10.1016/0010-4655(75)90039-9.

[50] R. E. Kalman. A new approach to linear filtering and prediction problems. 82:35–45.

[51] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3146–3154. Curran Associates, Inc. URL http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

[52] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. URL http://arxiv.org/abs/1412.6980.

[53] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. 49(4): 764–766. ISSN 0022-1031. doi: 10.1016/j.jesp.2013.03.013. URL http://www.sciencedirect.com/science/article/pii/S0022103113000668.

[54] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 4768–4777. Curran Associates Inc. ISBN 978-1-5108-6096-4. URL http://dl.acm.org/citation.cfm?id=3295222.3295230.

[55] S. M. Lundberg, G. G. Erion, and S.-I. Lee. Consistent Individualized Feature Attribution for Tree Ensembles - SHAP. . URL http://arxiv.org/abs/1802.03888.

[56] S. M. Lundberg, G. G. Erion, and S.-I. Lee. Consistent Individualized Feature Attribution for Tree Ensembles. . URL http://arxiv.org/abs/1802.03888.

[57] A. L. Maas. Rectifier nonlinearities improve neural network acoustic models.

[58] L. McInnes and J. Healy. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. URL http://arxiv.org/abs/1802.03426.

[59] T. C. Mills. *Time Series Techniques for Economists / Terence c. Mills*. Cambridge University Press Cambridge [England] ; New York. ISBN 0-521-34339-9 0-521-40574-2. URL http://www.loc.gov/catdir/toc/cam031/89007187.html.

[60] I. Mulalic, H. Rasmussen, J. Rouwendal, and H. H. Woltmann. The Financial Crisis and Diverging House Prices: Evidence from the Copenhagen Metropolitan Area. ISSN 1556-5068. doi: 10.2139/ssrn.3041272. URL https://www.ssrn.com/abstract=3041272.

[61] Particle Data Group et al. Review of Particle Physics. 98 (3):030001. doi: 10.1103/PhysRevD.98.030001. URL https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[62] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. 12:2825–2830.

[63] E. Polley and M. van der Laan. Super Learner In Prediction. URL https://biostats.bepress.com/ucbbiostat/paper266.

[64] J. Proriol, J. Jousset, C. Guicheney, A. Falvard, P. Henrard, D. Pallin, P. Perret, and B. Brandl. TAGGING B QUARK EVENTS IN ALEPH WITH NEURAL NETWORKS (comparison of different methods : Neural Networks and Discriminant Analysis). page 27.

[65] A. Purcell. Go on a particle quest at the first CERN webfest. URL https://cds.cern.ch/record/1473657.

[66] S. Ravanbakhsh, J. Schneider, and B. Poczos. Deep Learning with Sets and Point Clouds. URL http://arxiv.org/abs/1611.04500.

[67] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. Detecting Novel Associations in Large Data Sets. 334(6062):1518–1524. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1205438. URL http://science.sciencemag.org/content/334/6062/1518.

[68] J. W. Rohlf. *Modern Physics from A to Z*. John Wiley and Sons. ISBN 978-0-471-57270-1.

[69] P. J. Rousseeuw and C. Croux. Alternatives to the Median Absolute Deviation. 88(424):1273–1283. ISSN 0162-1459. doi: 10.1080/01621459.1993.10476408. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476408.

[70] A. Salam. Weak and electromagnetic interactions. In *Selected Papers of Abdus Salam*, volume Volume 5 of *World Scientific Series in 20th Century Physics*, pages 244–254. WORLD SCIENTIFIC. ISBN 978-981-02-1662-7. doi: 10.1142/9789812795915_0034. URL https://www.worldscientific.com/doi/abs/10.1142/9789812795915_0034.

[71] L. Scodellaro. B tagging in ATLAS and CMS. URL http://arxiv.org/abs/1709.01290.

[72] L. Shapley. A value for n-person games. In *The Shapley Value*, volume 28 of *Annals of Math Studies*, pages 307–317. doi: 10.1017/CBO9780511528446.003.

[73] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An Introduction to PYTHIA 8.2. 191:159–177. ISSN 00104655. doi: 10.1016/j.cpc.2015.01.024. URL http://arxiv.org/abs/1410.3012.

[74] S. J. Taylor and B. Letham. Forecasting at scale. doi: 10.7287/peerj.preprints.3190v2. URL https://peerj.com/preprints/3190.

[75] i. team. Iminuit – A python interface to minuit. URL https://github.com/scikit-hep/iminuit.

[76] R. Tibshirani. Regression Shrinkage and Selection via the Lasso. 58(1):267–288. ISSN 0035-9246. URL www.jstor.org/stable/2346178.

[77] A. Tikhonov. *On the Stability of Inverse Problems*, volume vol. 39 of *Doklady Akademii Nauk SSSR*.

[78] S. Toghi Eshghi, A. Au-Yeung, C. Takahashi, C. R. Bolen, M. N. Nyachienga, S. P. Lear, C. Green, W. R. Mathews, and W. E. O'Gorman. Quantitative Comparison of Conventional and t-SNE-guided Gating Analyses. 10. ISSN 1664-3224. doi: 10.3389/fimmu.2019.01194. URL https://www.frontiersin.org/articles/10.3389/fimmu.2019.01194/full.

[79] d. L. M. J. van, E. C. Polley, and A. E. Hubbard. Super Learner. 6(1). ISSN 1544-6115. doi: 10.2202/1544-6115.1309. URL https://www.degruyter.com/view/j/sagmb.2007.6.issue-1/sagmb.2007.6.1.1309/sagmb.2007.6.1.1309.xml.

[80] L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. 9:2579–2605. ISSN ISSN 1533-7928. URL http://www.jmlr.org/papers/v9/vandermaaten08a.html.

[81] F. van Veen. The Neural Network Zoo. URL http://www.asimovinstitute.org/neural-network-zoo/.

[82] V. Vapnik. Principles of Risk Minimization for Learning Theory. In *Proceedings of the 4th International Conference on Neural Information Processing Systems*, NIPS'91, pages 831–838. Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-222-9. URL http://dl.acm.org/citation.cfm?id=2986916.2987018.

[83] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors. SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. page arXiv:1907.10121. URL https://ui.adsabs.harvard.edu/abs/2019arXiv190710121V/abstract.

[84] I. Wallach and R. Lilien. The protein–small-molecule database, a non-redundant structural resource for the analysis of protein-ligand binding. 25(5):615–620. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp035. URL https://academic.oup.com/bioinformatics/article/25/5/615/183421.

[85] S. Weinberg. A Model of Leptons. 19(21):1264–1266. doi: 10.1103/PhysRevLett.19.1264. URL https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[86] H. Wickham. Tidy data. 59(10):1–23. ISSN 1548-7660. doi: 10.18637/jss.v059.i10. URL https://www.jstatsoft.org/v059/i10.

[87] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola. Deep Sets. URL http://arxiv.org/abs/1703.06114.

Index