Classification of Physics Events Using Machine Learning

Capstone Proposal for Udacity's Machine Learning Engineer Nanodegree

Gabriel Santucci[1]

[1]Nucleon Decay and Neutrino Group, Stony Brook University Physics Department

May 2017

Abstract

Statistical analysis in high energy physics experiments are most of the time trying to classify events into two classes, usually defined as signal and background classes. This proposal is particularly interested in studying applications of machine learning classifiers to optimize the separation of events coming from proton decay and atmospheric neutrino interactions in the Super-Kamiokande experiment. SK has searched for the mode $p \to \bar{\nu}K^+$ in the prompt-gamma channel. The official results are used as a benchmark model to be improved in this work using machine learning methods.

Proposal

1 Domain Background

There is a type of theories in particle physics called Grand Unified Theories (GUTs). The main idea behind all variations of GUTs is to have a unified description of 3 fundamental interactions of nature: the strong, weak and electromagnetic forces. There are many types of GUT models [1] each of them with different predictions, but one thing they all have in common is nucleon[1] decay.

As far as we know, the proton is a stable[2]particle and so is the neutron inside a nucleus (although free neutrons can decay). Since the typical energy of GUTs is far beyond the reach of any particle physics experiment, nucleon decay is the best bet for searching for evidence of GUTs.

One particular class of GUTs is the so called SUSY GUTs, when supersymmetry is present in the theory. In some models, the preferred channel for proton decay is $p \to \bar{\nu}_\mu K^+$, where the proton decays to two other particles: an anti-neutrino and a charged Kaon.

[1]Nucleon is the generic name for protons and neutrons, particles that live inside the nucleus of an atom.
[2]Being stable means that the particle will not spontaneously decay into other particles.

Detecting this kind of event is very challenging, but some experiments were built in the 80's and 90's to search for this kind of event. One of the main difficulties to detect this type of signal is due to neutrinos that are created in Earth's atmosphere and then interact inside the detector, leaving a signature very similar to the one left by the decay of a proton.

In physics the interesting class that is being studied is called 'Signal' and all other possible interactions that produce similar data in the detector are collectively called 'Background'[3]. These can be mapped into {1,0} classes for machine learning algorithms to perform binary classification.

I propose to study improvements that machine learning classifiers can make in the identification of these two classes: proton decay signal and background interactions coming from atmospheric neutrino events in the Super-Kamiokande experiment.

2 Problem Statement

To date, the biggest experiment looking for proton decay is the Super-Kamiokande (SK) experiment [2]. SK is a big water tank containing 50 kilo-tones of ultra pure water located 1 km deep underground on a mine in the Japanese alps. When a charged particle travels at extremely high speeds inside the tank, it produces light, this is known as the Cherenkov effect [3]. In the walls of the tank there are photo-multiplier tubes (PMTs), which essentially are light detectors [4]. Using these PMTs we know when a particle passed through the detector leaving a trace and producing light. We then have a data set with the time and charge of each PMT that received a hit. A reconstruction algorithm is then applied on this data so that physics quantities can be studied, like momentum, energy, direction and position of each particle that participated in the event.

As described in section 1, the proton decay mode that we are interested here consists of an anti-neutrino and a charged Kaon. The anti-neutrino has no electric charge, so it does not leave a trace in the detector and the Kaon also can not be seen, even though it has charge. That is because it does not have enough energy to be traveling at very high speeds, so it does not produce any Cherenkov light. But since the Kaon also decays, the hope is to see its decay products. Most of the time, it decays to a neutrino (which can not be seen) and an anti-muon[4], which is charged and has enough energy to be detected. Since this is a 2-body decay (the initial particle decays into 2 particles) and we know the masses of the final state particles, we also know the exact energy and momentum of the muon we are looking for (using energy and momentum conservation).

The problem is that after all this, the only visible particle that we can detect is the mono-energetic muon[5]. This would make the search impossible, due to the amount of

[3]The physics and ML names will be used interchangeably.

[4]This is the only Kaon decay mode that will be used in the analysis here, but there is another decay mode that can be seen by SK.

[5]This means that the energy of the muon is always the same. Also, since the SK can not detect the sign of the particle's charge (+ or -), we will not make a distinction between muons and anti-muons.

background events coming from atmospheric neutrino interactions. The solution is to look only for protons that decay inside the Oxygen nucleus of the water molecules, but not for the ones coming from the Hydrogen atoms. The difference is that when a proton decay inside the Oxygen, the remaining nucleus is left in an excited state and can emit a low energy photon (also called gamma) from nuclear de-excitation [5]. If we look for this photon[6] in coincidence with a mono-energetic muon, we then have a very particular signature in the detector that can be used to differentiate signal and background events. Figure 1 shows a diagram of our final state particles.

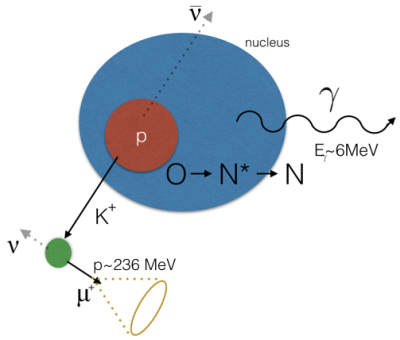

Figure 1: A proton inside an Oxygen nucleus decays to $\bar{\nu} + K^+$ and the remaining nucleus emits a 6 MeV photon. Later the Kaon decays into a ν and a mono-energetic μ^+ of about 236 MeV.

Even though the situation is better now, it is still a challenge to identify if a signal comes from a proton that decayed or a neutrino that entered the detector and interacted with some nucleus. There are billions[7] of neutrinos entering the detector every second. The chance of a neutrino interacting with a nucleus inside the tank is extremely small, but since the flux of incoming neutrinos is so high, we are bound to see some neutrino events[8] every hour or so.

Some of these events leave a trace in the detector that is very similar to our signal, so distinguishing between signal and background events is challenging. SK has searched for this decay mode before and the strategy used was to apply simple cuts in reconstructed variables such as momentum and number of particles present in the event [7]. By looking at

[6]The photon also does not have charge, but it produces electron-positron pairs as it travels in the water and these particles produce the Cherenkov light.

[7]Neutrinos come from many different sources, the main ones are the atmosphere and the Sun. Most of them cross our bodies, the Earth and all else and leave intact, since the chance of a neutrino interacting with something is incredibly small. Only from the Sun, 65 billion neutrinos cross a person's thumb every second.

[8]The detection of neutrinos in SK was awarded the Nobel prize in physics in 2015 for proving that neutrinos have mass [6].

many distributions of these variables (features) and applying hard boundary cuts on them, a selection criteria was used to determine if an event is coming from signal or background.

The rejection of background events is already very powerful in the standard analysis, however it would be better if it could be further improved. Due to the fact that proton decay was never observed, it is crucial to reduce the number of False Positives as much as possible so that if an event is detected, we are confident that it comes from proton decay and not background. Even if no event is seen in the data, a lower limit for the proton lifetime can be calculated using a Poisson distribution. Increasing this limit is important to further restrict GUT models based on their predictions for the proton lifetime.

The goal of this study is to improve the selection criteria for this proton decay mode using Machine Learning techniques. This could enhance the signal-background separation leading to a better limit on the proton lifetime or ultimately the discovery of proton decay. This would be a huge discovery that will certainly change the current view of particle physics and deeper our knowledge of elementary particles and the fundamental interactions of nature.

3 Datasets and Inputs

I will use 2 sets of data in this project, one for each class to be classified. Both sets are similarly generated using many steps of simulation. The first step is to generate the events using Monte Carlo event generators. The generator used in SK for neutrino interactions is called NEUT [8] and the proton decay generator is a custom code of the collaboration that simply generates a vector of kinematic variables for each particle present in the decay.

Once the events are generated, it is necessary to do the simulation of the detector. This is where we simulate how the particles created in the first step will interact as they travel through the water inside the SK tank. This code is also a custom simulator called skdetsim, which is based on a general package called Geant3 [9]. Finally, once all the physics simulation is done, the events need to be reconstructed. SK has 2 main algorithms to perform event reconstruction, APFit and fiTQun. In this study we will use fiTQun, which uses a maximum likelihood method based on different hypotheses for an event to reconstruct the kinematics of all particles present in the event.

After all these steps, we have a data set containing many features for every particle present in the event and a label for proton decay (class 1) simulation or neutrino interaction events (class 0). A pre-selection criteria will be applied in the data before we start the analysis. These cuts are necessary to guarantee that fiTQun will perform well, namely, I will only use events that are fully contained inside the fiducial volume (FCFV) of the tank[9]. Also, we need 1-ring events, which means that at most only one particle is detected at a time. This is because our signal is the mono-energetic muon and the low energy

[9]This means that the event started and ended inside the detector and 2 m away from the tank walls, without any activity happening outside.

photon. The photon energy is so small that it will not produce enough light to be seen as a Cherenkov ring in the detector, while the muon will be clearly seen. Therefore, we can greatly reduce the number of background by only looking at events that contain only one particle.

Once this pre-selection is applied we can finally start using ML to optimize our event classification. The different features I plan to use in the analysis are:

- $\ln\left(\frac{L_\mu}{L_e}\right)$: This is the log of the ratio of likelihoods for the electron and muon hypotheses. FiTQun will try to fit the event using the hypothesis that an electron generated the ring and after using a muon, for each fit we have the likelihood of how well the fit explains the data[10]. One can then compare the likelihood of different hypotheses to see which one is more likely.

- $\ln\left(\frac{L_\mu}{L_\pi}\right)$: Similarly, one can compare the likelihood of the ring being created by a muon and a charged pion.

- $\ln\left(\frac{L_\mu}{L_{\mu\gamma}}\right)$: And finally, one can compare the likelihood of the event having a single muon present or a low energy photon being also present.

- P_μ: This is the reconstructed momentum[11] of the muon.

- P_γ: The reconstructed momentum of the photon.

- ΔT: The time difference between the muon ring appearing and the photon, in the hypothesis where both are present in the event. Ideally, for signal events this would have a non-zero time difference between them, since the Kaon travels some time before decaying. For neutrino events, the muon is produced instantaneously, so the time difference should be close to zero.

- ΔX: The distance between the muon and the Michel electron[12] vertices.

The proton decay sample generated contains 100,000 events, out of which approximately 30,000 are events where the Kaon decayed to a muon and the low energy photon was present. After applying the pre-selection cuts of only FCFV 1-ring events, the number goes down to about 22,000 events. The atmospheric neutrino event sample was generated using 500 years of Monte Carlo simulation. This means that we simulated a number of events equivalent to SK running for 500 years without stopping. Neutrino interactions are so rare, that a huge amount of events is necessary to be able to properly estimate the number of events in a low-background search like ours. The total number of neutrino

[10] Similar to a goodness of fit.

[11] In high energy physics the unit of energy and momentum used is the electronVolt (eV), such that 1 MeV is 1 million eV.

[12] Michel electron is the electron that comes after the muon decays [10].

events is on the order of 2.5 million. After applying pre-selection cuts, these number is still very big, so extra reduction is necessary.

Another layer of cuts is then applied to reduce the background further. The selected cuts are loose, such that there is still a high number of signal events while the number of background events is drastically reduced. With this extra layer, signal and background samples have similar numbers of events. The choice of these cuts are also chosen based on a physics prior, we know that the final analysis must have more strict cuts than these. So it will not bias our results if we only use these neutrino background events in the analysis. It is guaranteed that the events that are left out here will never be mistaken as signal events. The definition of pre-selection[13] cuts is:

- $\ln\left(\frac{L_\mu}{L_e}\right) > 0$: Events should look more like muons than electrons.
- 210 MeV $< P_\mu <$ 270 MeV: The muon momentum is known to be about 236 MeV, due to detector limitations, we can not measure momentum with infinite resolution. Therefore a window around the expected momentum is allowed.
- $P_\gamma < 2$ MeV: The momentum of the photon should be at least 2 MeV. This will get rid of events where no photon is found.
- $\Delta T > 0$ ns: The time difference between the muon and the photon should be at least 0 nanoseconds. We know that the muon will come after the photon in our signal, so we get rid of a lot of background that has negative ΔT values.

After applying this filter we are finally ready to begin being analyzed. We still have to prepare the data further before feeding it to some ML algorithm, like standardizing each feature, etc. But this will be done in the final project submission. Our final set of data contains 18652 signal events and 24435 neutrino events.

4 Solution Statement

As described in section 3, the amount of background events is reduced by a factor of 100. The remaining events are can potentially be mistaken as a proton decay signal which would increase the false positive rate. But the amount of remaining background is still high and needs to be reduced event further[14].

The goal here is to use machine learning algorithms to learn an optimized way of classifying events as signal or background. The way the standard analysis is done is by using a decision tree, since it simply applies boundary cuts on different features of the

[13] More details of the pre-selection will not be provided here since they are not the main purpose of this study.

[14] Sections 5 and 6 describe why it is so important to have such a small number of false positives in this type of search.

data. A natural choice of algorithm for the new analysis could be Boosted Decision Trees to continue in the same direction. Since systematic uncertainties have to be taken into account for a full analysis, it is preferable to use simple boundary cuts compared with more complex methods, such as neural nets.

On the other hand, this approach is completely new, so there is no reason to be attached to the standard analysis and new algorithms can be tested. The plan is to test different methods like for example Random Forest and Support Vector Machines. Both these algorithms can find a new boundary that separates both classes in feature space.

Applying a dimensionality reduction algorithm is another common practice in ML applications, but it will not be used here. There are two main reasons for that, one is that our feature space is not that big compared to image recognition for instance, where the number of features is thousands of times bigger than ours. Also, we want to keep track of the features that are being used in the feature space itself, since they have a one-to-one mapping with physical quantities. If an algorithm like Principal Component Analysis is applied, the new space is a linear combination of the original features which complicates the interpretation of final results and the treatment of systematic errors. Systematic uncertainties are beyond the purpose of this study, but it is necessary to keep it in mind when choosing an algorithm for the analysis.

Here I just want to find an algorithm with simple interpretation that performs well given some metric that will be discussed in Section 6. The goal is to perform better than the standard analysis and there are two ways this can be achieved. Either we improve the efficiency[15] of keeping signal events while having the same amount of background events. Or we keep a similar efficiency while reducing the number of background. Of course, achieving both higher efficiency and reduced number of background is the best possible scenario. But if only one of them is achieved the effort can be considered a success.

5 Benchmark Model

As discussed in Section 2, SK has already performed a search for the mode we are studying here. This particular mode is sometimes called prompt-gamma method, since it relies on the coincidence measurement of the muon and the photon that is promptly emitted by the nucleus. The details of the analysis, as well as the final results can be found in [7], but figure 2 shows a table that summarizes the main results of that search.

The criteria A1-8 are the different cuts applied in that analysis, the background rate is explained in more detail in Section 6.2, eq. 4, but basically it refers to the expected number of false positives if the detector had 1 million tones of volume and took data for 1 year. The expected background is the number of false positives in the analyzed data after each cut A$-i$ is applied, and the Data column corresponds to the actual number observed.

[15] The definition of efficiency that will be used in this work is given by eq. 3 in Section 6.

	SK-I/III/IV				SK-II			
Criterion	Bkg. Rate	Exp. Bkg.	Data	Signal Eff.	Bkg. Rate	Exp. Bkg	Data	Signal Eff.
A-1	35240.8	7432.3	7497	0.575	34910.6	1717.6	1712	0.566
A-2	24865.7	5244.2	5240	0.520	22239.7	1094.2	1051	0.473
A-3	2496.6	526.5	531	0.494	2161.0	106.3	91	0.440
A-4	2443.7	515.4	520	0.485	2067.8	101.7	87	0.420
A-5	2400.3	506.2	514	0.479	2030.0	99.9	82	0.414
A-6	2302.7	485.6	488	0.436	1931.5	95.0	78	0.368
A-7	1.34	0.28	0	0.084	5.84	0.29	0	0.063
A-8	1.11	0.24	0	0.084	2.75	0.14	0	0.062

Figure 2: Summary of the results obtained in the standard analysis done in SK. Definition of all quantities is given in the text.

Finally, the Signal Efficiency is defined in eq. 3 in Section 6.1 and corresponds to recall, i.e., the probability that an event is classified as signal given that it is signal.

The table contains values for all SK eras, from I to IV. SK-II is different due to an accident that reduced the number of PMTs, so the photo-coverage is different. We are using simulated data for SK-IV only, while the table contains the averaged results of SK I, III and IV. SK-IV performance if better than I and III due to better electronics and other factors, but since the performance is not super different. We can use the numbers in the first column as our benchmark model.

The efficiency in the last column is reported to be

$$\epsilon_B = 0.084 = 8.4\%, \tag{1}$$

where B stands for Benchmark. This is one of the main numbers we wanna compare to, so using the definition of efficiency given in 3 and the necessary scale factor discussed in Section 6.1, we can make a comparison between all classifiers used in this study and the benchmark model.

Also, since the amount of data analyzed will be different, we have to compare the number of false positives in our analysis with the Mton $\cdot$ year rate provided in the table. Section 6.2 explain the necessary transformations we need to do with FP to compare with the benchmark rate. Eq. 4 gives the rescale for our analysis and the benchmark result we can define as:

$$R_B = 1.1 \text{ event per Mega-tone per year.} \tag{2}$$

Section 6 discusses different metrics that will be used in this study. They will be useful when comparing different classifiers among themselves, but to compare our final results with the benchmark model provided by the standard analysis of SK, I will use the efficiency and rate defined here in eqs. 1 and 2.

6 Evaluation Metrics

In this work, two numbers are crucial to judge a classifier performance: efficiency and expected background rate.

6.1 Efficiency

Let us define efficiency as:

$$\epsilon = r \text{ (recall)} = \frac{TP}{TP + FN}, \tag{3}$$

where TP is the number of true positives[16] and FN is the number of false negatives[17]. So efficiency is the same as recall and it measures how well the classifier is able to select events of the signal class. Ideally this will be as high as possible while keeping the number of false positives (FP[18]) as low as possible.

To be able to compare the efficiency of this study with the standard efficiency presented in Section 5 it is necessary to rescale the total number of events in class 1: FP+FN. That is because ϵ_B uses a slightly different definition of efficiency, so we have to multiply by the total number of class 1 events in our sample and divide by the total number of class 1 events before the filtering process described in section 3. The total number of class 1 events in our sample is 18652 and the total number of events used by the standard analysis is 82687. Therefore, $18652/82687 = 0.23$ is the scale factor we need to apply in our efficiency when comparing with the benchmark efficiency, $\epsilon_B = 0.084$.

6.2 Expected Background

The expected number of background events is related to the number of FP in the analysis. As described in Section 3, an atmospheric neutrino sample corresponding 500 years of data was used. But the typical way of presenting a background expectation value is given in Mton $\cdot$ year. This is the amount of background events expected to be detected if the detector took data for 1 year and its size was 1 million tones. By doing that, it is easy to compare different experiment results, since not all experiments have the same size or run for the same amount of time.

To achieve that, we first need to divide our number of false positives by 500, to have an yearly rate and then scale our volume. SK has 50 kilo-tones of water, but only 22.5 kilo-tones are used to take data (this is the fiducial volume as described in Section 3). Thus, we have:

[16] The number of events correctly classified as class 1.
[17] The number of events incorrectly classified as class 0.
[18] The number of events incorrectly classified as class 1.

$$FP \to \frac{FP}{500} \cdot \frac{1}{0.0225} = 0.089 \cdot FP, \tag{4}$$

where the 0.0225 corresponds to the fiducial volume of SK in Mton. Another useful way of reporting background is by given the expected number of background events that correspond to the time the experiment actually took data. This is called the live time of the experiment and it is defined by the number of days that data was taken. SK has started taking data since 1996, but throughout the years, many different things changed in the experiment (all the SK eras are described in [2]) so the number of years that will be used in this analysis is 2519.89 days = 6.90 years. Since we have the number of FP in 500 years of data, we simply have to scale that down to 6.90 years to obtain the expected number of FP in the live time, given that the volume of the detector is the same. Thus,

$$FP \to FP \cdot \frac{6.90}{500} = FP \cdot 0.0138, \tag{5}$$

where the scale factor is simply the ratio of both live times.

6.3 Possible Metrics

6.3.1 The F1 score

With these definitions of efficiency and background rates, a metric can be formed to be used as a score for each classifier that we test. Given that we need to have the best efficiency possible while reducing the number of FP as much as possible, one popular choice would be the F1 score, the weighted average of recall and precision given by:

$$\mathrm{F1} = 2\frac{pr}{p+r}, \tag{6}$$

where r is the recall defined as efficiency in eq. 3 and p is the precision defined as:

$$p = \frac{TP}{TP+FP}, \tag{7}$$

thus, p incorporates the FP quantity that we want to minimize. For a low number of false positives, $p \to 1$.

In case the classifier is performing extremely well, recall and precision approach 1 as the number of FN and FP go to zero, making F1 $\to 1$. This metric can certainly serve as a comparison between two different classifiers, but it can raise a problem when looking at the number of FP. Since p and r are treated equally in F1, we might achieve high scores without reducing the number of FP enough (comparing it to the benchmark rate for example).

6.3.2 The 5-sigma standard

As it was seen in Section 5, the expected background rate has to be as low as possible to claim a discovery or to increase the proton lifetime. Thus, it might be better to use another metric that incorporates the need of making FP low. A popular choice in high energy physics is S/B or $S/\sqrt{B}$, where S represents signal and B the expected background. We can use the definition of signal efficiency (recall) and the expected number of background events in the simulated live time as the measurements of S and B. As our efficiency increases, the capacity of detecting signal (S) increases and as the expected number of background events (given by eq. 5) decreases, our confidence that a selected event is indeed a proton decay event increases.

This definition of $\frac{S}{\sqrt{B}}$ is one way an experiment can report sensitivity results for new discoveries [11]. In the high energy physics community, the standard threshold is that $\frac{S}{\sqrt{B}} > 5$ to be able to claim the discovery of a new particle or phenomena. What this means is that the background expectation has to fluctuate at least 5 standard deviations[19] to be able to explain your observation. Given that a 5σ fluctuation is extremely unlikely[20], it indicates the sign of new physics.

This way if we observe 1 event that is classified as signal and we have an efficiency of 0.9 for example, we have $S = 0.9$. If we expect 0.2 background events in the amount of data being analyzed (live time), we have $\sqrt{B} = 0.44$ and then

$$\frac{S}{\sqrt{B}} = \frac{0.9}{0.44} \approx 2,$$

this simple example shows the importance of how low the number of FP has to be. Even if we have a 90% efficiency and a low background rate of only 0.2 events, it is still not unlikely to obseve one event in the data. Given that there is a reasonable chance that 0.2 can fluctuate to 0.9. For a 90% efficient classifier, we need to expect less than 0.03 events in the data to have a 5σ sensitivity.

6.3.3 Conclusion

I plan to study at least the two metrics presented here in this proposal and see which one is better to achieve the wanted results.

[19] Assuming a Poisson distribution for the number of background events, then the fluctuation of n events is given by $\sigma = \sqrt{n}$

[20] One has to calculate the p-value associated with the measurement to quantify how unlikely the observation is. Assuming a gaussian distribution for σ gives 1 chance in 3.5 million. A lot more strict than the 1 in 20, used by other fields with p-value of 5%.

7 Project Design

I plan to used the data generated accordingly to the process described in Section 3. I will also add new data sets, generated in the same manner to use as test/validation sets. The data needs to be explored and visualized to make sure all the chosen features can be used in the final analysis and if their behavior is performing as expected. Figure 3 shows an example of data visualization, we can see the reconstructed muon momentum in the proton decay and the neutrino background samples. The blue curve contains all events in the sample, including events that do not have a prompt-gamma or events where the Kaon has a different decay mode[21]. The green curve is the signal sample, as expected we can see the peak of the distribution at $\approx$ 240 MeV and the red curve shows the background distribution. Also as expected, the background sample shows a continuous spectrum of momentum with no peak present. This illustrates the type of exploration necessary to understand the data and optimize the selection analysis based on its behavior.

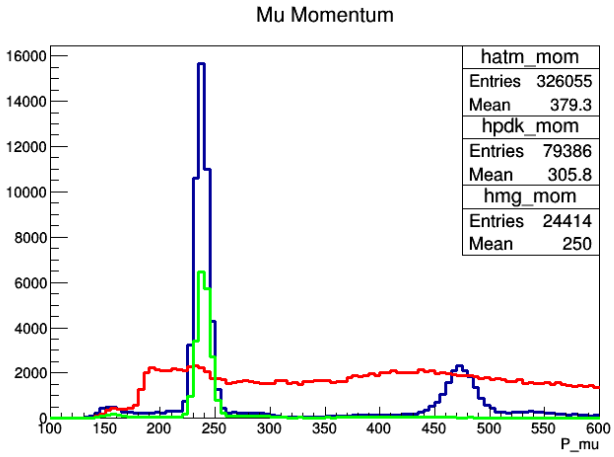

Figure 3: Reconstructed muon momentum for all events in the proton decay sample (Blue), only the events that satisfy the signal requirement (Green) and the neutrino background sample (Red).

After data exploration, I will pre-process the data to optimize the way classifiers work. For example, I can standardize all features so that all of them are dimensionless and have a similar magnitude. This way the classifier does not get tricked into thinking one variable is more important, for example the muon momentum, since it is a hundred times higher than the photon momentum.

Once data is ready to be fed to the algorithms, I will test the performance of different classifiers like BTDs and SVMs using the metrics discussed in Section 6. The basic pipeline

[21] That is the reason for the $\approx$ 475 MeV bump.

will be choosing a cross-validation scheme like train-test split or stratified k-fold and perform a grid search on hyper parameters of the algorithms. Once some set of parameter is found, we can compare different algorithms and the benchmark model presented in Section 5. Due to the fact that these comparisons will possibly use different metrics, I might revisit the parameter set to compare only with the benchmark, in case the number of FP needs to be reduced.

References

[1] W. de Boer, **Grand Unified Theories and Supersymmetry in Particle Physics and Cosmology**. arXiv:9402266

[2] Super-Kamiokande collaboration, **The Super-Kamiokande detector**, Nuclear Instruments and Methods in Physics Research A 501 (2003) 418-462. SK detector

[3] For more see: Wiki Cherenkov Radiation

[4] For more see: Wiki Photomultipliers

[5] H. Ejiri, **Nuclear deexcitations of nucleon holes associated with nucleon decays in nuclei**, Phys. Rev. C vol. 48, 3. Nuclear deexcitations of Oxygen

[6] For more see: 2015 Nobel

[7] Super-Kamiokande collaboration, **Search for Proton Decay via $p \to \nu K^+$ using 260 kiloton·year data of Super-Kamiokande**. arXiv:1408.1195

[8] Y. Hayato, **A neutrino interaction simulation program library NEUT**. NEUT

[9] Geant can be downloaded from: Geant

[10] For more: Wiki Michel Electron

[11] A simple but complete explanation with the example of the Higgs Boson discovery: Higgs discovery