PULSE SHAPE DISCRIMINATION

1. *Scientific background:*

Neutron and Gamma interaction mechanisms:

Gamma rays or X rays interact with matter by three primary processes- Compton Scattering, Photoelectric effect and Pair Production. Fig 2.1 illustrates their relative importance by plotting the total mass attenuation coefficient as a function of photon energy for Al and Pb. The mass attenuation coefficient is the measure of probability of the gamma ray interacting with the material.

Diagram

Description automatically generated

Fig 2.1: Photon mass attenuation coefficients for Al and Pb as a function of photon energy. Dashed lines show individiual contributions from each of the primary processes.

If a photon beam is well collimated, all the three interaction processes cause the beam to be attenuated as it passes through matter. It has been observed that the intensity decreases exponentially with the length of path traversed through the matter.

Neutrons interact with matter via nuclear reactions which particularly depend on the energy and the nuclei with which the neutrons collide. Neutrons may be scattered in which case the energy is transferred to the recoiling (charged) nuclei. Alternatively, they could be absorbed in a variety of different processes. Fusion is the main contributor to absorption cross section. While traversing a slab of material, neutron beams, like gamma beams are also attenuated due to absorption and scattering and the intensity shows an exponential decay wit the path traversed. Fig 2.2 shows a comparative plot depicting the neutron and gamma attenuation in NE213 as a function of incident energy.

Diagram

Description automatically generated

Fig 2.2 Neutron Gamma attenuation as a function of incident energy in NE213.

PSD:

Diagnostics for the energetic particles released in a thermonuclear fusion are based on scintillators, that exhibit scintillation, the property of luminescence, emission of radiation when the material is excited by ionizing radiation. A schematic diagram of a detector system is shown in Fig. 1. Some of the light released in the phosphor is collected in the [photomultiplier](https://www.sciencedirect.com/topics/engineering/photomultiplier) tube, which consists of a set of electrodes with photosensitive surfaces. When a light photon strikes the surface, an electron is emitted by the photoelectric effect, it is accelerated to the next surface where it dislodges more electrons, and so on, and a multiplication of current is achieved. An amplifier then increases the electrical signal to a level convenient for counting or recording.

Diagram

Description automatically generated

### Fig 1. Scintillation Counters

The major components of the scintillator light decay in times of the order of a few nanoseconds. This means that in principle, organic scintillators can operate at very high counting rates. However, there is a weaker, longer lived component of the radiation from many scintillators that corresponds to delayed fluorescence. Consequently, the total light output can often be represented by the sum of the two exponential decays referred to as the fast and slow components of the scintillation. The slow component has a characteristic decay time in the range of a few hundred nanoseconds. The fraction of the total light observed in this weaker, slower component is a function of the type of particle inducing the radiation. Heavier particles have higher specific ionization and produce more delayed fluorescence light. Figure 3 illustrates the time dependence of scintillator pulses in stilbene, a solid organic crystal scintillator, when the crystal is excited by different types of radiation.

Diagram

Description automatically generated

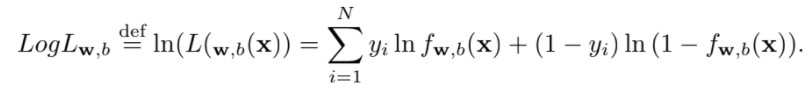
Fig 3: The time dependence (in nanosecond) of scintillation pulses in stilbene when excited by different incident particle types.

In organic scintillators the fraction of light emitted during delayed fluorescence depends on the exciting particle, therefore it is possible to differentiate between neutron and photon interactions through pulse shape discrimination (PSD).

1. *Overview of the classification tools:*

# Logistic Regression:

# Logistic regression is not a regression, but a classification learning algorithm. Logistic regression is named for the function used at the core of the method, the logistic function. The [logistic function](https://en.wikipedia.org/wiki/Logistic_function), also called the sigmoid function is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits. It is defined as y= 1 / (1 + exp(-x)). Let us define our model as: f(x) = 1/ (1 + exp-(w.x+b)). By looking at the graph of the standard logistic function, we can see how well it fits our classification purpose: if we optimize the values of w and b appropriately, we could interpret the output of f(x) as the probability being positive or negative. For example, if it is higher than or equal to the threshold 0.5 we would say that the class of x is positive; otherwise, it’s negative. In practice, the choice of the threshold could be different depending on the problem. For instance, assume that we have a labeled example (xi, yi) in our training data. Assume also that we have found (guessed) some specific values w and b of our parameters. If we now apply our model f to xi we will get some value 0 <p< 1 as output. If yi is the positive class, the likelihood of yi being the positive class, according to our model, is given by p. Similarly, if yi is the negative class, the likelihood of it being the negative class is given by 1 - p. To find the best values of w and b for our model we use the optimization criterion called maximum log-likelihood. The log-likelihood is defined like follows:



Because ln is a strictly increasing function, maximizing this function is the same as maximizing its argument, and the solution to this new optimization problem is the same as the solution

to the original problem. A typical numerical optimization procedure used in such cases is gradient descent as there is no closed form solution to the above optimization.

Gaussian Mixture Models:

Gaussian mixture models are a probabilistic model for representing [normally distributed](https://brilliant.org/wiki/multivariate-normal-distribution/) subpopulations within an overall population. [Mixture models](https://brilliant.org/wiki/mixture-model/) in general don't require knowing which subpopulation a data point belongs to, allowing the model to learn the subpopulations automatically. Since subpopulation assignment is not known, this constitutes a form of [unsupervised learning](https://brilliant.org/wiki/unsupervised-learning/). A Gaussian mixture model is parameterized by two types of values, the mixture **component weights**and the component **means(***μk*​**)** and **variances/covariances(***σk***)**. Training of the model involves [expectation maximization](https://brilliant.org/wiki/expectation-maximization-algorithm/), a technique most commonly used to estimate the mixture model's parameters. Expectation maximization for mixture models consists of two steps. The first step, known as the **expectation** step or **E** step, consists of calculating the expectation of the component assignments *Ck*​for each data point *xi*​ ∈ *X* given the model parameters *ϕk*​, *μk*​, and *σk*​. The second step is known as the **maximization** step or **M** step, which consists of maximizing the expectations calculated in the E step with respect to the model parameters. This step consists of updating the values *ϕk*​, *μk*​, and *σk.* The entire iterative process repeats until the algorithm converges, giving a maximum likelihood estimate.Once the EM algorithm has run to completion, the fitted model can be used to perform various forms of inference. The two most common forms of inference done on GMMs are [density estimation](https://brilliant.org/wiki/density-estimation/) and [clustering](https://brilliant.org/wiki/clustering/).

Chart, line chart

Description automatically generated

Support Vector Machines:

Support Vector Machine (SVM) is a supervised [machine learning algorithm](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle) which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line). Maximizing the distances between nearest data point (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as **Margin.** The SVM kernel is a function that takes low dimensional input space and transforms it to a higher dimensional space i.e. it converts not separable problem to separable problem. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then finds out the process to separate the data based on the labels or outputs we have defined.

Chart, scatter chart

Description automatically generated

1. *Results*

Train-test split (50/50):

Logistic Regression accuracy- 99.9897%

GMM purity score- 74.3466%

SVM score- 99.9897%

Train-test split (70/30):

Logistic Regression accuracy- 98.9885%

GMM purity score- 74.3466%

SVM score- 96.8343%

Train-test split (80/20):

Logistic Regression accuracy- 99.9914%

GMM purity score- 74.3466%

SVM score- 97.1097%

1. *Conclusion*

Three machine learning tools have been applied to the task of pulse shape discrimination – Logistic Regression, Gaussian Mixture Model, and Support Vector machine. Logistic regression out performed the task as compared to others. Besides SVM also proved to be a good approach. Moreover, both these techniques output the probability of each pulse to be a neutron or a gamma ray, allowing detailed analysis of the final classification. This added value can become particularly useful in the case of neutron spectrometry, in which it is crucial to exclude the gamma rays to reach reliable conclusions about the plasma properties. GMM could not give a satisfactory performance probably because the clusters (Total area/Tail area) for neutrons and gamma rays are situated very close to each other.

With regard to future developments, from a machine learning perspective significant improvements could be obtained by implementing alternative metrics like neural networks, KNN, decision trees and so on.