**Gaussian Discriminant Analysis**

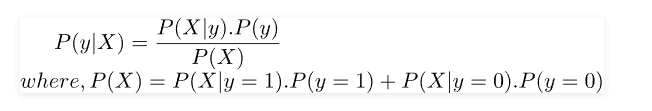
There are two types of Supervised Learning algorithms are used in Machine Learning for classification.

1. Discriminative Learning Algorithms
2. Generative Learning Algorithms

Logistic Regression, Perceptron, and other Discriminative Learning Algorithms are examples of discriminative learning algorithms. These algorithms attempt to determine a boundary between classes in the learning process. A Discriminative Learning Algorithm might be used to solve a classification problem that will determine if a patient has malaria. The boundary is then checked to see if the new example falls on the boundary, **P (y|X)**, i.e., given a feature set X, what is its probability of belonging to the class "y".

Generative Learning Algorithms, on the other hand, take a different approach. They try to capture each class distribution separately rather than finding a boundary between classes. A Generative Learning Algorithm, as mentioned, will examine the distribution of infected and healthy patients separately. It will then attempt to learn each distribution's features individually. When a new example is presented, it will be compared to both distributions, and the class that it most closely resembles will be assigned, **P (X|y)** for a given **P(y)** here, P(y) is known as a class prior.

These Bayes Theory predictions are used to predict generative learning algorithms:



By analyzing only, the numbers of **P (X|y)** as well as **P(y)** in the specific class, we can determine P(y), i.e., considering the characteristics of a sample, how likely is it that it belongs to class "y".

Gaussian Discriminant Analysis (GDA) is a probabilistic generative model used for classification tasks. It assumes that the data from each class follows a Gaussian (normal) distribution, and it leverages Bayes' theorem to calculate the posterior probability of a data point belonging to a particular class given its features. In this detailed explanation, we'll cover the underlying principles, assumptions, equations involved in GDA, as well as how the model learns from the data and makes predictions.

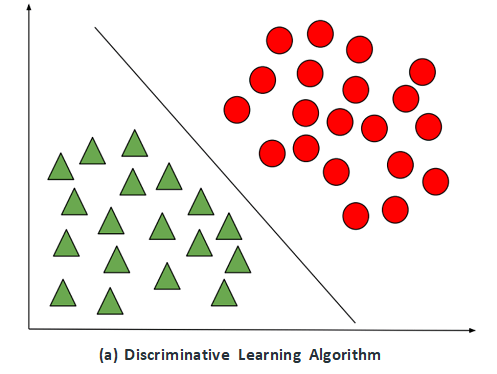
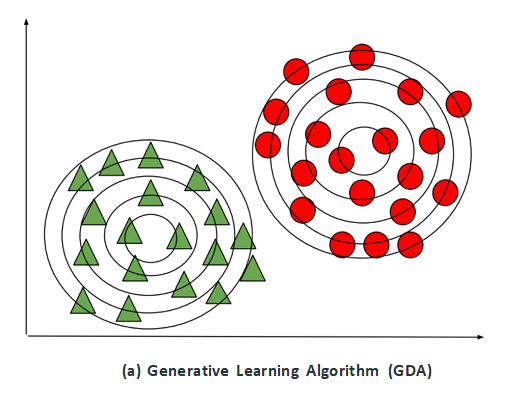
Gaussian Discriminant Analysis is a Generative Learning Algorithm that aims to determine the distribution of every class. It attempts to create the Gaussian distribution to each category of data in a separate way. The likelihood of an outcome in the case using an algorithm known as the Generative learning algorithm is very high if it is close to the centre of the contour, which corresponds to its class. It diminishes when we move away from the middle of the contour. Below are images that illustrate the differences between Discriminative as well as Generative Learning Algorithms.

1. Underlying Principles:

* GDA is based on the assumption that the data from each class is normally distributed. This means that the features of each class follow a Gaussian distribution with specific mean and variance.
* The model aims to learn the parameters (mean and variance) of the Gaussian distribution for each class from the training data.
* GDA calculates the posterior probability of an instance belonging to each class given its features using Bayes' theorem and then assigns the instance to the class with the highest probability.

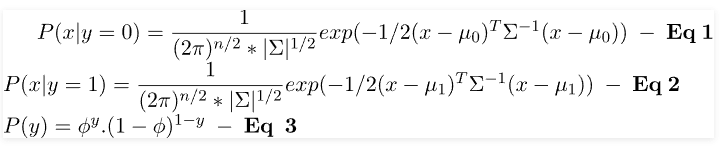
2. Assumptions:

* GDA assumes that the features in each class follow a multivariate Gaussian distribution.
* The classes have the same covariance matrix, meaning the features' variances and covariances are shared across all classes.
* The data points are assumed to be generated independently.



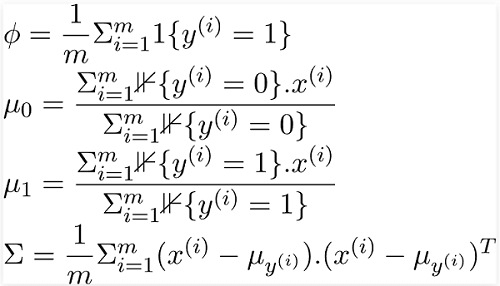
Let's take a look at the case of a classification binary problem in which all datasets have **IID** (Independently and identically distributed). To determine **P (X|y)**, we can use Multivariate Gaussian Distribution to calculate a probability density equation for every particular class. In order to determine P(y) or the class prior for each class, we can make use of the Bernoulli distribution since all sample data used in binary classification could be 0 or 1.

So the probability distribution, as well as a class prior to a sample, could be determined using the general model of Gaussian and **Bernoulli distributions:**



To understand the probability distributions in terms of the above parameters, we can formulate the likelihood formula, which is the product of the probability distribution as well as the class before every data sample (Taking the probability distribution as a product is reasonable since all samples of data are considered IID).

Gaussian Discriminant Analysis

In accordance with the principle of Likelihood estimation, we need to select the parameters so as to increase the probability function, as shown in Equation 4. Instead of maximizing the Likelihood Function, we can boost the Log-Likelihood Function, a strict growing function.Gaussian Discriminant Analysis

In the above equations, "**1{condition}**" is the indicator function that returns 1 if this condition holds; otherwise returns zero. For instance, 1{y = 1} returns 1 only if the class of the data sample is 1. Otherwise, it returns 0 in the same way, and similarly, in the event of 1{y = 0}, it will return 1 only if the class of the sample is 0. Otherwise, it returns 0.

## Model Learning Process:

* Parameter Estimation: The learning process in GDA involves estimating the parameters (mean and covariance matrix) of the Gaussian distribution for each class from the training data.
* Mean Estimation: For each class, the mean vector 'μ' is calculated as the average of all feature vectors of instances belonging to that class.
* Covariance Matrix Estimation: To estimate the shared covariance matrix 'Σ', the covariance of the features is calculated across all classes. This involves computing the covariance matrix for each class and then taking their weighted average, where the weights are proportional to the number of instances in each class.

The parameters derived can be used in equations 1, 2, and 3, to discover the probability distribution and class before the entire data samples. The values calculated can be further multiplied in order to determine the Likelihood function, as shown in Equation 4. As previously mentioned, it is the probability function, i.e., P (X|y). P(y) is integrated into the Bayes formula to calculate P (y|X) (i.e., determine the type 'y' of a data sample for the specified characteristics 'X').

Thus, Gaussian Discriminant Analysis works extremely well with a limited volume of data (say several thousand examples) and may be more robust than Logistic Regression if our fundamental assumptions regarding data distribution are correct.

In conclusion, Gaussian Discriminant Analysis (GDA) is a powerful and intuitive probabilistic generative model used for classification tasks. By assuming that the data from each class follows a multivariate Gaussian distribution, GDA captures the underlying probability distribution of the features within each class, allowing for effective modeling of complex data patterns and the incorporation of uncertainty in the classification process.

Throughout this detailed explanation, we have explored the fundamental principles, key assumptions, and essential equations involved in GDA. The model's core principle revolves around estimating the parameters of the Gaussian distribution, specifically the mean vector and the covariance matrix, for each class from the training data. These estimates form the foundation for calculating the likelihood of a data point belonging to a particular class, which is crucial for posterior probability calculation using Bayes' theorem.

One of the strengths of GDA lies in its ability to handle multiple features and account for the correlations between them through the covariance matrix. Moreover, the assumption of shared covariance across classes allows the model to capture the underlying data structure more efficiently, especially when the number of features is large and they exhibit interdependencies.

The model learning process in GDA entails the estimation of the parameters for each class through mean estimation and covariance matrix estimation. This step ensures that the Gaussian distributions accurately represent the statistical properties of the data within each class, making the model more robust and better suited for classification tasks.

In the prediction phase, GDA employs the Maximum A Posteriori (MAP) decision rule to classify new data points. By calculating the posterior probabilities for each class using Bayes' theorem, the model can effectively evaluate the likelihood of a data point belonging to each class given its features. The class with the highest posterior probability is then selected as the predicted class for the instance.

Gaussian Discriminant Analysis's elegant probabilistic approach not only provides a reliable classification mechanism but also inherently captures uncertainty in the model. This ability to quantify uncertainty is valuable in real-world scenarios where making informed decisions is crucial, especially in cases where misclassification may have significant consequences.

However, it is important to consider some limitations of GDA. The assumption of Gaussian distribution may not always hold in real-world data, and if the data does not align well with the Gaussian model, the performance of GDA may be compromised. Additionally, the shared covariance assumption may not be suitable for datasets with vastly different covariance structures across classes.

In summary, Gaussian Discriminant Analysis is a versatile and interpretable classification model that leverages the Gaussian distribution to probabilistically analyze and classify data. Its ability to handle multiple features, account for feature correlations, and incorporate uncertainty makes it a valuable tool in various fields, including pattern recognition, image processing, and medical diagnosis. However, as with any modeling approach, it is crucial to assess the underlying data characteristics and the appropriateness of GDA's assumptions before applying the model to ensure accurate and reliable classification results.