Go Back to Machine Learning

Course Content

LDA and QDA

Before understanding Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), let us understand the concepts of **prior and posterior probabilities** with the help of the **Bayes theorem**.

Prior and Posterior probabilities

Let us take an example to understand these probabilities: Suppose there are two types of people in a society, namely good and bad people, and let's assume that the probability of class k (good or bad) is given as follows:

$$P(Y = k) = \pi_k$$

Here, $P(y = k)$ is called the **prior probability**. If the count of good people is 48 out of a total of 80 people, then the prior probability of good people is 48/80 = 0.6 and that for the bad people is 32/80 = 0.4

Now, every person has some distinct features that help to classify them into these classes. If there is only one independent variable, the probability of a person to have some feature X given that person is belonging to class k is given in the form of the following **normal distribution:**

$$P(X \mid Y = k) = \gamma_k \, exp\left\{-\frac{(X - \mu_k)^2}{2\sigma_k^2}\right\}$$

Where, $P(X|Y = k)$ is the probability of X given the class of the person is k, γ_k is the normalizing factor, μ_k and σ_k are the mean and the standard deviation of the normal distribution.

If we have more than one independent variable, the above expression of probability is given as follows:

$$P(X|Y = k) = N(\mu_k, C_k) = \Upsilon_k \, exp\left\{-\frac{1}{2}(X - \mu_k)^T C_k^{-1}(X - \mu_k)\right\}$$

While building models with multiple variables, it is required to consider the **variation among them.** Standard deviation is not capable of seeing the variation in more than one variable. Due to this, another quantity namely the covariance matrix C_k is introduced in this equation.

The covariance matrix is the generalization of variance between two variables to the variance between multiple variables. The diagonal entries of the covariance matrix represent the variance of that particular variable.

The above equation is used to generate the probability of a person to have features X given the person is belonging to class k, i.e., good or bad. The expressions of μ_k and C_k are given as follows:

$$\mu_k = E[X|Y = k]$$

$$C_k = E\left[(X - \mu_k)(X - \mu_k)^T |Y = k\right]$$

Where, E represents **the expectation or the mean**.

Now, to make a prediction, we need to find the probability of a person belonging to class k given the person has features X, i.e., we need to find $P(Y=k|X)$ which is called the **posterior probability.**

Let's suppose that we already know the prior probability and the parameters of the normal probability distribution of each class. Then to find this, we need to use an important theorem called the **Bayes theorem**. The mathematical expression of the Bayes theorem is given as follows:

$$P\left(Y=k|X\right) = \frac{\pi_k P\left(X|Y=k\right)}{P\left(X\right)}$$

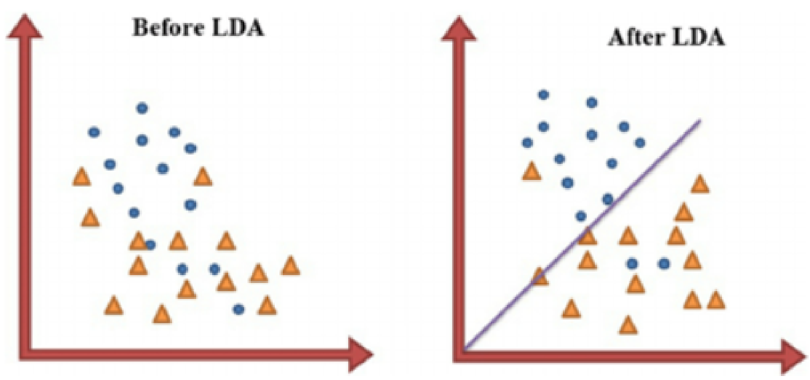
LDA and QDA

Let us now get into the concepts of LDA and QDA.

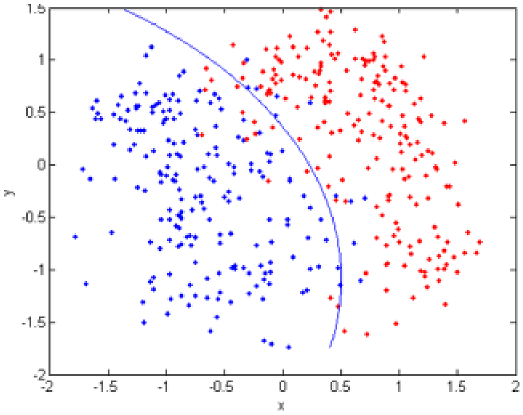
The concepts of prior and posterior probabilities are useful in LDA and QDA. The expression of probability in LDA and QDA is the logarithm of the posterior probability explained above. The predicted class is the one that maximizes the log of the posterior probability.

In the case of LDA, it works with the condition that all the covariance matrices are the same, while QDA contains the logarithmic expression for probability with different covariance matrices for each class.

LDA stands for Linear Discriminant Analysis. It is known so, because it creates a linear separation between the existing classes while solving a classification problem. It finds the line that best separates the data into two classes. It is preferred when there are a small number of observations.



In Quadratic Discriminant Analysis, the purpose is similar to LDA, but it finds a quadratic / nonlinear curve that best separates the data into two classes. The quadratic curve is more flexible than the linear one and can separate classes more efficiently in the case of a large dataset. Due to this, QDA performs better over large datasets.



Help