

Generative Models for Classification

In the case of [Logistic Regression](#) we directly model $\Pr(Y = k|X = x)$ using the [Sigmoid Function](#) (Logistic Function) its a simple conditional probability approach **Predicting Y given X** .

Considering the alternative Probability of the the predictors X given a class Y which is the logic behind **Bayes Theorem**, When the distribution of X is normal the model turn to be very similar to Logistic Regression.

Why its needed?

- If There is a substantial separation between the two classes the **parameter estimates** for logistic regression model will be unstable
- If the predictors X are normally distributed and the sample size n is small Using **Generative Models** will be more accurate than Logistic Regression

Suppose that we want to classify an observation into on of the K classes where $K \geq 2$, means the [Response](#) Y can take K possible classes.

- Let π_k be the **Prior Probability** that the a random [Observation](#) comes from the k th class
- Let $f_k(X) = \Pr(X|Y = k)$ is the **Density Function** of X
 - $f_k(x)$ will be large if there is **high probability** of that observation being in the k th class

$$\Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum^K \pi_i f_i(x)}$$

- This is just the [Bayes' theorem](#) Formula
- Let $p_k(x) = \Pr(Y = k|X = x)$ and will also be called the **Posterior** probability
- So our goal is to know the probability of the [Observation](#) being in the k th class Given the class k aka [Response](#) Y

Linear Discriminant Analysis Vs Logistic Regression

First before diving into **LDA** ,Lets establish a clear difference between What [Logistic Regression](#) Does and What **LDA Does** , cause at the end they both Classify a new observation into a class.

What Logistic Regression Does

- It tries to directly model the **posterior probability**

$$P(Y = k|X = x) = \frac{e^{\vec{x}\vec{\beta}}}{1 + e^{\vec{x}\vec{\beta}}}$$

- You just want to separate the classes,Its **Discriminative** just making distinctions
- We don't care how the data X looks like inside each class, We just wanna class each observation
- We also don't care about how the data was generated

What LDA Does

- Its a **generative**, it models how the data X looks within each class
- With the assumption of the data X taking a normal distribution form

$$P(X|Y = k) = \mathcal{N}(\mu_k, \Sigma)$$

- Then use [Bayes' theorem](#) to compute

$$P(Y = k|X = x) = \frac{P(X|Y = k)P(Y = k)}{P(X)}$$

- Then classify x by choosing the class with the highest **posterior** $P(Y = k|X = x)$
- Simply its saying *This new point looks most like the kind of data that class k generates*