

4.8.3

Suppose we have K classes and if an observation belongs to k -th class, then x comes from a one dimensional normal distribution.

$$x \sim N(\mu_k, \sigma_k^2)$$

where,

x : The feature we are observing which is assumed to be normally distributed within each class.

K : The number of classes

μ_k : The mean of the one-dimensional normal distribution of class k

σ_k^2 : The variance of the one dimensional normal distribution of class k

The objective is, by applying Bayes Theorem, we get

$$P(Y=k | x=a) \propto P(x=a | Y=k) * P(Y=k)$$

$$P(x=a | Y=k) = \frac{1}{\sigma_k * \sqrt{2\pi}} * \exp\left[-\frac{(a - \mu_k)^2}{2\sigma_k^2}\right]$$

our goal is to maximize $P(Y=k | x=a)$

So, we have to maximize

$$P(Y=k) \propto \frac{1}{\sigma_k * \sqrt{2\pi}} * \exp\left[-\frac{(a - \mu_k)^2}{2\sigma_k^2}\right]$$

\therefore The above equation is a quadratic in a and μ_k so it can't be a linear one.

\therefore Bayes classifier is not linear but quadratic