LDA Mean And Variance Estimates

Maximum Likelihood Estimation for the parameters of **Linear Discriminant Analysis**, As stated in Generative Models for Classification that **LDA** assumes that X is normally distributed and the variance σ^2 is shared across all classes K ,While each class k have a mean μ_k.
Using the **MLE** to estimates these parameters

Estimate of μ_k

$$L(\mu_k) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu_k)/2\sigma^2}$$

- Taking the log likelihood $l(\mu_k) = \log(L(\mu_k))$

$$l(\mu_k) = \sum^{k} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu_k)^2}{2\sigma^2}$$

$$l(\mu_k) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum^{k}(x-\mu_k)^2$$

$$l(\mu_k) = \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum^{k}(x-\mu_k)^2$$

- Now we differentiate with the respect to μ_k and solve for 0

$$\frac{d}{d\mu_k} l(\mu_k) = \frac{1}{2\sigma^2}\sum^{k}(x-\mu_k) = 0$$

$$\frac{d}{d\mu_k} l(\mu_k) = \sum^{k}(x-\mu_k) = \sum x - \sum \mu_k$$

Note : $\sum \mu_k = n_k\mu_k$

$$\sum^{k} x_i = n_k\mu_k$$

$$\hat{\mu_k} = \frac{1}{n_k}\sum x_i$$

Estimate of σ^2

Unlike the mean μ_k where you only focus on the mean of the class k , The variance σ^2 is shared across all the datasets, Where

- dataset (x_i, y_i)
- Each data point x_i belongs to class $y_i \in 1\dots K$

$$L = \prod^{n} P(x_i|y_i) = \prod_{k=1}^{K}\prod_{i:y_i=k} P(x_i|y_i=k) = \prod_{k=1}^{K}\prod_{i:y_i=k}\mathcal{N}(x_i|\mu_k\sigma^2)$$

$$L(\sigma^2) = \prod^{K}\prod_{i:y_i=k}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu_k)^2/2\sigma^2}$$

- Taking the **log-likelihood** $l(\sigma^2) = \log(L(\sigma^2))$

$$l(\sigma^2) = \sum^{K}\sum_{i:y_i}\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\mu_k)^2$$

- Now we differentiate with the respect to σ^2

$$l(\sigma^2) = \frac{n}{2}\log(2\pi\sigma^2)\sum^{K} - \frac{1}{2\sigma^2}\sum_{i:y_i=k}(x_i-\mu_k)^2$$

$$\frac{d}{d\sigma^2} l(\sigma^2) = -\frac{n}{2} . \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum^{K} \sum_{i:y_i=k} (x_i - \mu_k)^2 = 0$$

- Now multiply By $(\sigma^2)^2$

$$-n\sigma^2 + \sum^{K} \sum_{i:y_i=k} (x_i - \mu_k)^2 = 0$$

$$\sigma^2 = \frac{1}{n} \sum^{K} \sum_{i:y_i=k} (x_i - \mu_k)^2$$

- With Bassel's correction

$$\sigma^2 = \frac{1}{n-K} \sum^{K} \sum_{i:y_i=k} (x_i - \hat{\mu}_k)^2$$

- To make it **unbiased** using $n - K$
- Its the **weighted average** of sample variance for each class