Linear Discriminant Analysis (LDA)

Oliver Zhao

Linear Discriminant Analysis (LDA) attempts to model the distribution of the features $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ based on the classes $\mathbf{Y} = \{\mathbf{y_1}, \ldots, \mathbf{y_n}\}$ as $P(\mathbf{X} = x|\mathbf{Y} = k)$. Then Bayes' theorem is applied to obtain the distribution estimates for $P(\mathbf{Y} = k|\mathbf{X} = x)$.

Let π_k represent the probability that a randomly chosen sample comes from the kth class and let $f_k(x) = P(\mathbf{X} = x|\mathbf{Y} = k)$ be the probability density function of $\mathbf{X}$ for a sample from the kth class. Bayes' theorem states that

$$P(\mathbf{Y} = k|\mathbf{X} = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)}. \tag{1}$$

For conciseness, we shall use the abbreviation

$$p_k(X) = P(\mathbf{Y} = k|\mathbf{X}). \tag{2}$$

According to the equation above, in order to obtain $p_k(X)$, we can just plug in our calculated estimates of π_k and $f_k(X)$. π_k is easy enough to estimate just from sampling $\mathbf{Y}$, but finding $f_k(X)$ is not trivial. To simplify the approximation of $f_k(x)$, we assume that $f_k(x)$ is Gaussian, having the following form

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right), \tag{3}$$

Where μ_k and σ_k^2 is the mean and variance for the kth class. Additionally, we assume that $\sigma_1^2 = \sigma_2^2 = \sigma_K^2$. Note that if this assumption is not made in a two-class problem, the classification method becomes Quadratic Discriminant Analysis (QDA). Through these given assumptions, we can redefine $p_k(x)$ as

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma_k} exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right)}. \tag{4}$$

The Bayes classifier assigns a data sample $\mathbf{X} = x$ to the class that has the highest value of $p_k(x)$. To make the computations simple, notice that by taking the log on both sides, we can instead maximize the value of $\delta_k(x)$ where

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + log(\pi_k). \tag{5}$$

In summary, LDA functions by approximating the Bayes classifier by using estimates for π_k, μ_k, and σ^2, where

$$
\begin{aligned}
\hat{\pi}_k &= n_k/n, \\
\hat{\mu}_k &= \frac{1}{n_k} \sum_{i:y_i=k} x_i, \\
\hat{\sigma}^2 &= \frac{1}{n-K} \sum_{k=1}^{K} \sum_{y_i=k} (x_i - \hat{\mu_k})^2.
\end{aligned}
\tag{6}
$$