Diagonal Linear Discriminant Analysis

Saffet Gökçen Şen

May 18, 2020

Assume that a classification is to be made. A generative discrete model is used for this task. Then:

$$p\left(y=c|\mathbf{x},\boldsymbol{\theta}\right)=\frac{p\left(\mathbf{x}|y=c,\boldsymbol{\theta}\right)p\left(y=c|\boldsymbol{\theta}\right)}{p\left(\mathbf{x}|\boldsymbol{\theta}\right)} \tag{1}$$

Assume that a multivariate Gaussian distribution is used for $p\left(\mathbf{x}|y=c,\boldsymbol{\theta}\right)$. If the covariance matrix for the multivariate Gaussian is common to all classes, then this analysis is called the linear discriminant analysis (LDA). In addition, it is assumed that the covariance matrix is diagonal. Hence, we are in the diagonal LDA. Let $p\left(\mathbf{x}|y=c,\boldsymbol{\theta}\right)$ be written more explicitly:

$$p\left(\mathbf{x}|y=c,\boldsymbol{\theta}\right)=\frac{1}{\left(2\pi\right)^{D/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_c\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_c\right)\right] \tag{2}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with the j^{th} diagonal being equal to σ_j^2. j is from 1 to D. σ_j^2 is the variance of the j^{th} feature.

$$\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_c\right)\right]_j=\frac{1}{\sigma_j}\left(x_j-\mu_{c_j}\right)\Rightarrow \tag{3}$$

$$\left(\mathbf{x}-\boldsymbol{\mu}_c\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_c\right)=\sum_{j=1}^{D}\frac{1}{\sigma_j^2}\left(x_j-\mu_{c_j}\right)^2 \tag{4}$$

μ_{c_j} is the mean of the j^{th} feature for the class c. σ_j is the variance of the j^{th} feature. μ_{c_j} is estimated as follows:

$$\hat{\mu}_{c_j}=\frac{1}{N_{c_j}}\sum_{i=1}^{N_{c_j}}x_{c_j}^{(i)} \tag{5}$$

The j^{th} feature has a variance σ_j^2. This variance can be estimated using the data from each class c. The unbiased estimate for the variance of the j^{th} feature with the class being equal to c is given by

$$\hat{\sigma}_{c_j}^2=\frac{1}{N_{c_j}-1}\sum_{i:y_i=c}\left(x_{ij}-\hat{\mu}_{c_j}\right)^2 \tag{6}$$

The pooled variance is calculated using the following weighted average:

$$\sigma_j^2 = \frac{\sum_{c=1}^{C}(N_{c_j} - 1)\hat{\sigma}_{c_j}^2}{\sum_{c=1}^{C}(N_{c_j} - 1)} = \frac{\sum_{c=1}^{C}(N_{c_j} - 1)\hat{\sigma}_{c_j}^2}{N - C} \tag{7}$$

References

Machine Learning A Probabilistic Perspective, Kevin P. Murphy