Introduction to Machine Learning

Classification: Discriminant Analysis

compstat-lmu.github.io/lecture_i2ml

LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA follows a generative approach

$$\pi_k(\mathbf{x}) = \mathbb{P}(y = k \mid \mathbf{x}) = \frac{\mathbb{P}(x|y = k)\mathbb{P}(y = k)}{\mathbb{P}(x)} = \frac{p(\mathbf{x}|y = k)\pi_k}{\sum\limits_{j=1}^{g} p(\mathbf{x}|y = j)\pi_j}$$
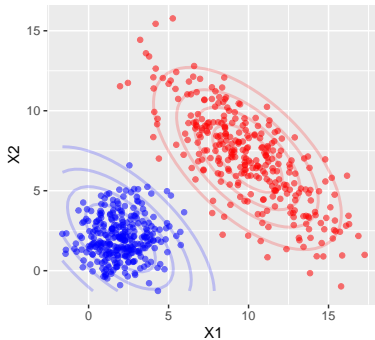
where we now have to pick a distributional form for $p(\mathbf{x}|y = k)$.

LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA assumes that each class density is modeled as a *multivariate Gaussian*:

$$p(\mathbf{x}|y = k) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu_k})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu_k})\right)$$

with equal covariance, i. e. $\Sigma_k = \Sigma \quad \forall k$.

LINEAR DISCRIMINANT ANALYSIS (LDA)

Parameters θ are estimated in a straight-forward manner by estimating

$$
\begin{aligned}
\hat{\pi}_k &= n_k/n, \text{ where } n_k \text{ is the number of class } k \text{ observations} \\
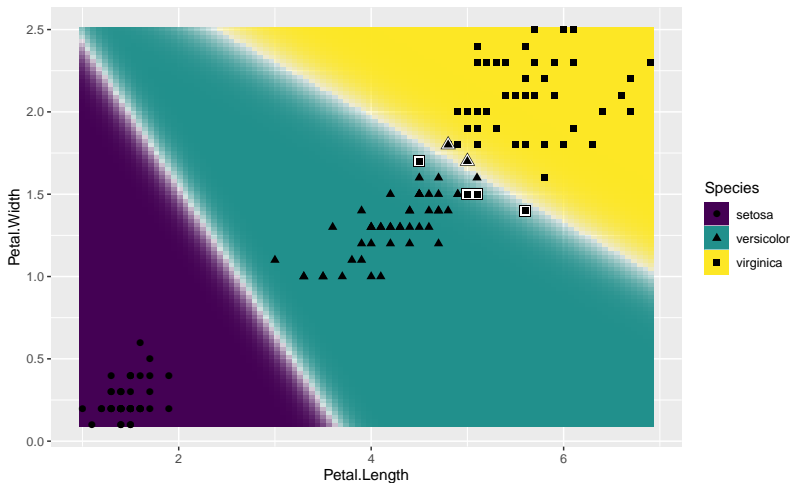\hat{\mu}_k &= \frac{1}{n_k} \sum_{i:y^{(i)}=k} \mathbf{x}^{(i)} \\
\hat{\Sigma} &= \frac{1}{n-g} \sum_{k=1}^{g} \sum_{i:y^{(i)}=k} (\mathbf{x}^{(i)} - \hat{\mu}_k)(\mathbf{x}^{(i)} - \hat{\mu}_k)^T
\end{aligned}
$$
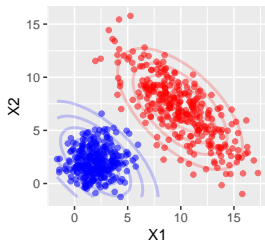
LDA AS LINEAR CLASSIFIER

Because of the equal covariance structure of all class-specific
Gaussian, the decision boundaries of LDA are linear.

LDA AS LINEAR CLASSIFIER

We can formally show that LDA is a linear classifier, by showing that the posterior probabilities can be written as linear scoring functions - up to any isotonic / rank-preserving transformation.

$$\pi_k(\mathbf{x}) = \frac{\pi_k \cdot p(\mathbf{x}|y = k)}{p(\mathbf{x})} = \frac{\pi_k \cdot p(\mathbf{x}|y = k)}{\sum\limits_{j=1}^{g} \pi_j \cdot p(\mathbf{x}|y = j)}$$

As the denominator is the same for all classes we only need to consider

$$\pi_k \cdot p(\mathbf{x}|y = k)$$

and show that this can be written as a linear function of $\mathbf{x}$.

LDA AS LINEAR CLASSIFIER

$$\pi_k \cdot p(\mathbf{x}|y = k)$$
$$\propto \quad \pi_k \exp\left(-\tfrac{1}{2}\mathbf{x}^T\Sigma^{-1}x - \tfrac{1}{2}\boldsymbol{\mu_k}^T\Sigma^{-1}\boldsymbol{\mu_k} + \mathbf{x}^T\Sigma^{-1}\boldsymbol{\mu_k}\right)$$
$$= \quad \exp\left(\log \pi_k - \tfrac{1}{2}\boldsymbol{\mu_k}^T\Sigma^{-1}\boldsymbol{\mu_k} + \mathbf{x}^T\Sigma^{-1}\boldsymbol{\mu_k}\right)\exp\left(-\tfrac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right)$$
$$= \quad \exp\left(\boldsymbol{\theta}_{0k} + \mathbf{x}^T\boldsymbol{\theta}_k\right)\exp\left(-\tfrac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right)$$
$$\propto \quad \exp\left(\boldsymbol{\theta}_{0k} + x^T\boldsymbol{\theta}_k\right)$$

by defining $\boldsymbol{\theta}_{0k} := \log \pi_k - \tfrac{1}{2}\boldsymbol{\mu_k}^T\Sigma^{-1}\boldsymbol{\mu_k}$ and $\boldsymbol{\theta}_k := \Sigma^{-1}\boldsymbol{\mu_k}$.

We have again left-out all constants which are the same for all classes k, so the normalizing constant of our Gaussians and $\exp\left(-\tfrac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right)$)

By finally taking the log, we can write our transformed scores as linear:

$$f_k(x) = \boldsymbol{\theta}_{0k} + x^T\boldsymbol{\theta}_k$$

QUADRATIC DISCRIMINANT ANALYSIS (QDA)

QDA is a direct generalization of LDA, where the class densities are now Gaussians with unequal covariances Σ_k.

$$p(\mathbf{x}|y = k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu_k})^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu_k}) \right)$$

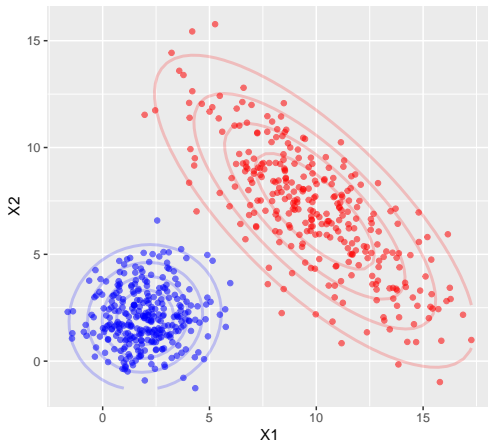Parameters are estimated in a straight-forward manner by:

$$\hat{\pi}_k = \frac{n_k}{n}, \text{ where } n_k \text{ is the number of class } k \text{ observations}$$

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i:y^{(i)}=k} \mathbf{x}^{(i)}$$

$$\hat{\Sigma}_k = \frac{1}{n_k - 1} \sum_{i:y^{(i)}=k} (\mathbf{x}^{(i)} - \hat{\mu}_k)(\mathbf{x}^{(i)} - \hat{\mu}_k)^T$$

QUADRATIC DISCRIMINANT ANALYSIS (QDA)

- Covariance matrices can differ over classes.
- Yields better data fit but also requires estimation of more parameters.

QUADRATIC DISCRIMINANT ANALYSIS (QDA)

$$
\begin{aligned}
\pi_k(\mathbf{x}) &\propto \pi_k \cdot p(\mathbf{x}|y=k) \\
&\propto \pi_k |\Sigma_k|^{-\frac{1}{2}} \exp(-\frac{1}{2}\mathbf{x}^T \Sigma_k^{-1} \mathbf{x} - \frac{1}{2}\boldsymbol{\mu_k}^T \Sigma_k^{-1} \boldsymbol{\mu_k} + \mathbf{x}^T \Sigma_k^{-1} \boldsymbol{\mu_k})
\end{aligned}
$$

Taking the log of the above, we can define a discriminant function that is quadratic in x.

$$
\log \pi_k - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}\boldsymbol{\mu_k}^T \Sigma_k^{-1} \boldsymbol{\mu_k} + \mathbf{x}^T \Sigma_k^{-1} \boldsymbol{\mu_k} - \frac{1}{2}\mathbf{x}^T \Sigma_k^{-1} \mathbf{x}
$$

QUADRATIC DISCRIMINANT ANALYSIS (QDA)