

1 Problem 1

1.1 Mathematical Analysis

Let's consider this generic formula, to classify new vector \mathbf{x} :

$$\arg \max_{c \in \mathcal{C}} \frac{p(y=c)p(\mathbf{x}|y=c)}{\sum_{c \in \mathcal{C}} p(y=c)p(\mathbf{x}|y=c)}.$$
 In our case the class conditional probabilities

are modelled as normal distributions so this means, using the notation from [1] that the above formula translates in:

$$\arg \max_{c \in \mathcal{C}} \frac{\pi_c |2\pi \Sigma_c|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\mathbf{x} - \mu_c)^T \Sigma_c^{-1}(\mathbf{x} - \mu_c)]}{\sum_{c \in \mathcal{C}} \pi_c |2\pi \Sigma_c|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\mathbf{x} - \mu_c)^T \Sigma_c^{-1}(\mathbf{x} - \mu_c)]}.$$
 Where π_c are the prior

probabilities, μ_c the means and Σ_c the covariance matrices for every class. Notice that performing the previous maximization is the same (by the property of the exponential) as minimizing the following formula: $(\mathbf{x} - \mu_c)^T \Sigma_c^{-1}(\mathbf{x} - \mu_c)$ which is called Mahalanobis distance and which is indeed what we implemented in our code because we wanted to be able to call a single function from both QDAlearn and LDAlearn functions, and hence we needed to be general. But to see why we get linear and quadratic boundaries let's keep analyzing the exponential formula. In LDA we make the assumption that: $\forall c \in \mathcal{C} \Sigma_c = \Sigma$ so we can rewrite the above formula as:

$$\arg \max_{c \in \mathcal{C}} \frac{\pi_c \exp[-\frac{1}{2}(\mathbf{x} - \mu_c)^T \Sigma^{-1}(\mathbf{x} - \mu_c)]}{\sum_{c \in \mathcal{C}} \pi_c \exp[-\frac{1}{2}(\mathbf{x} - \mu_c)^T \Sigma^{-1}(\mathbf{x} - \mu_c)]}.$$
 With some calculations we can

go further and see the equality with:

$$\arg \max_{c \in \mathcal{C}} \frac{\exp[\log(\pi_c) + \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c]}{\sum_{c \in \mathcal{C}} \exp[\log(\pi_c) + \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c]} = \quad (1)$$

$$\arg \max_{c \in \mathcal{C}} \frac{\exp[-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}] \cdot \exp[\log(\pi_c) + \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c]}{\exp[-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}] \cdot \sum_{c \in \mathcal{C}} \exp[\log(\pi_c) + \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c]}.$$
 (2)

So we can see that the only quadratic term in \mathbf{x} got cancelled out. Since we only care about the maximum class probability we can disregard about the denominator (which doesn't vary with \mathbf{c}) and consider only:

$$\arg \max_{c \in \mathcal{C}} \exp[\log(\pi_c) + \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c].$$
 Since \log is a monotonic function we can disregard the exponentiation and only care about what is inside it, getting: $\arg \max_{c \in \mathcal{C}} \log(\pi_c) + \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c.$

In our case the prior probabilities π_c for every class where all very similar and about 20% so we can disregard them, and consider only:

$$\arg \max_{c \in \mathcal{C}} \mu_c^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c.$$
 The previous formula is **linear** in \mathbf{x} and indeed

if we consider the boundary with any two classes $c, c' \in \mathcal{C}$ given by the equation: $p(y = c' | \mathbf{x}) = p(y = c | \mathbf{x})$ iff $\boldsymbol{\mu}_{c'}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_{c'}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{c'} = \boldsymbol{\mu}_c^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_c^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_c$ we see that this is the equation of a line. In the case of QDA we cannot perform the canceling from (1) to (2) and so we end up with a quadratic term in \mathbf{x} in the final equation and this is why we end up with quadratic boundaries.

1.2 Experimental Results

The accuracy achieved with LDA is 97% while with QDA is 94%: evidently the data is best separated by lines and not by curves. In the following we can see the graphic results of LDA and QDA:

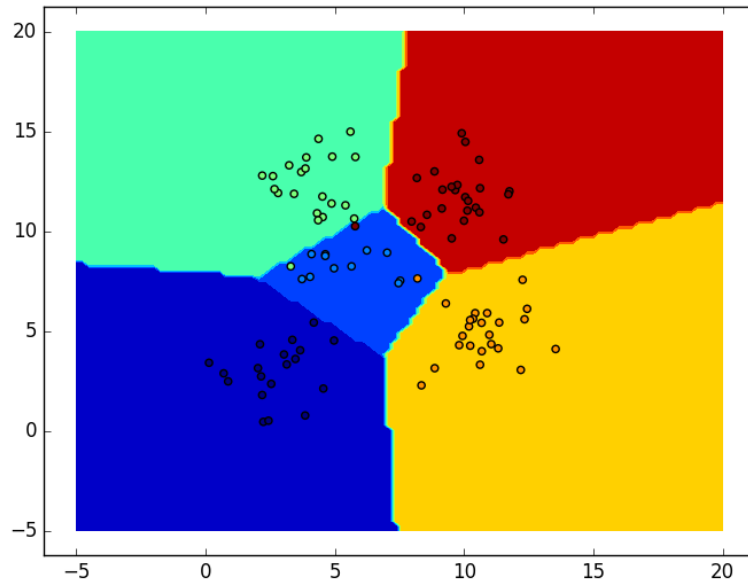


Figure 1: LDA

References

- [1] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

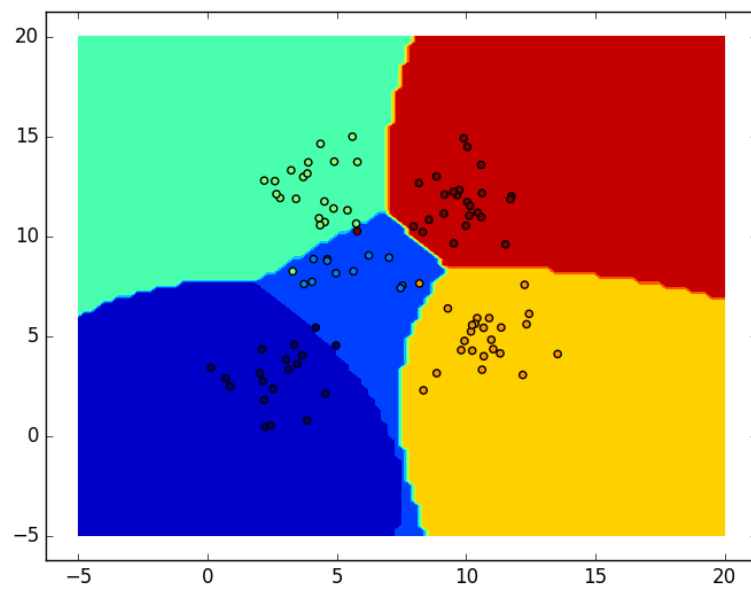


Figure 2: QDA