Predictive Analytics
Regression and Classification

Lecture 9 : Part 1

Sourish Das

Chennai Mathematical Institute

Aug-Nov, 2019

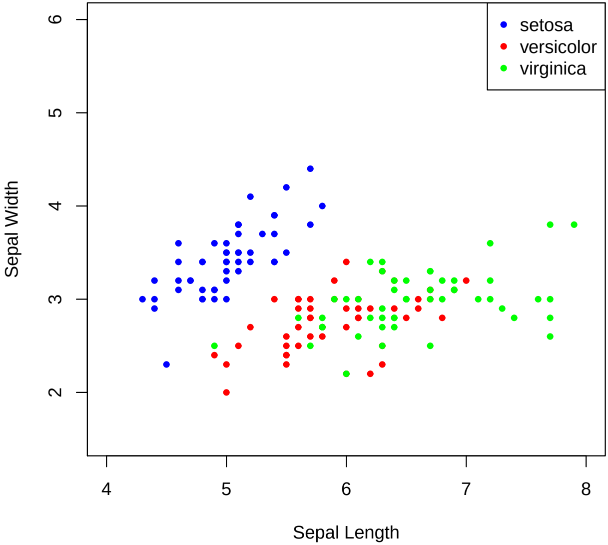
Classify the three species of the Iris Flower

Iris Versicolor **Iris Setosa** **Iris Virginica**

How the dataset looks like?

Sepal Length (X_1)	Sepal Width (X_2)	Species	Group/Label (k)
5.1	3.5	setosa	1
7.0	3.2	versicolor	2
6.7	3.3	verginica	3
⋮	⋮	⋮	⋮

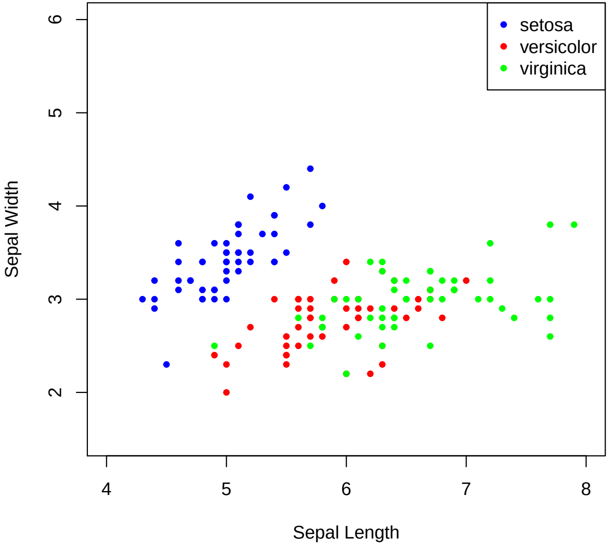
Classify the three species of the Iris Flower



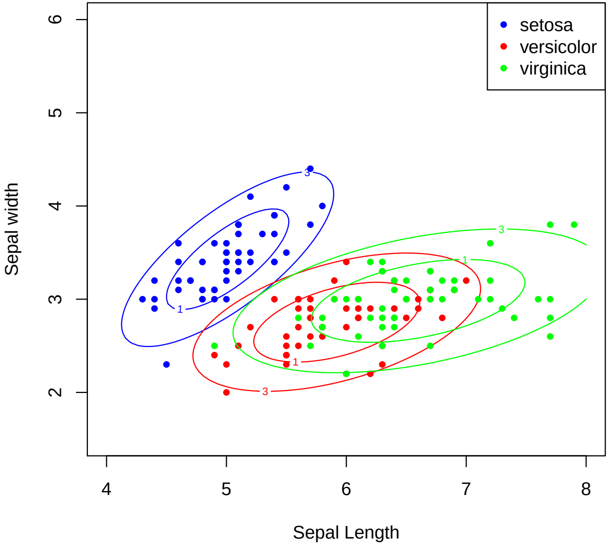
Classify the three species of the Iris Flower

- Suppose $\mathbf{X}_{k=1} = (X_1, X_2)$ is the vector of features of species setosa
- Similarly, $\mathbf{X}_{k=2} = (X_1, X_2)$ is the vector of features of species versicolor
- And, $\mathbf{X}_{k=3} = (X_1, X_2)$ is the vector of features of species virginica
- We can assume $\mathbf{X}_k = (X_1, X_2)$ follows joint probability distribution with pdf as $f_k(x)$
- Given a new test point $\mathbf{X} = (X_1, X_2)$, we want to classify the new flower into one of the three species.

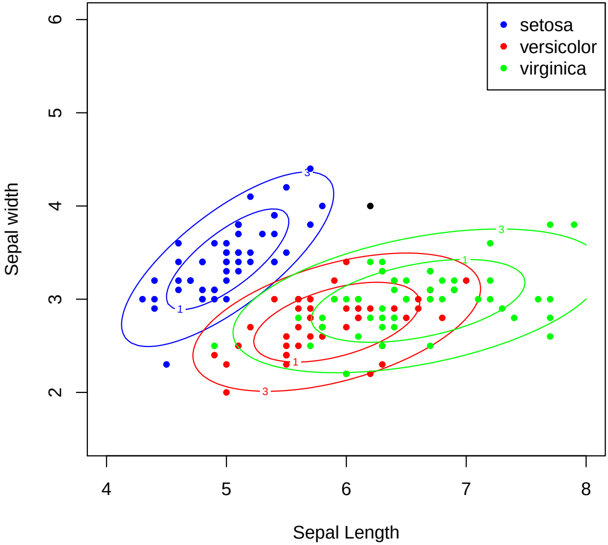
Classify the three species of the Iris Flower



Classify the three species of the Iris Flower



Classify the three species of the Iris Flower



Discriminant Analysis

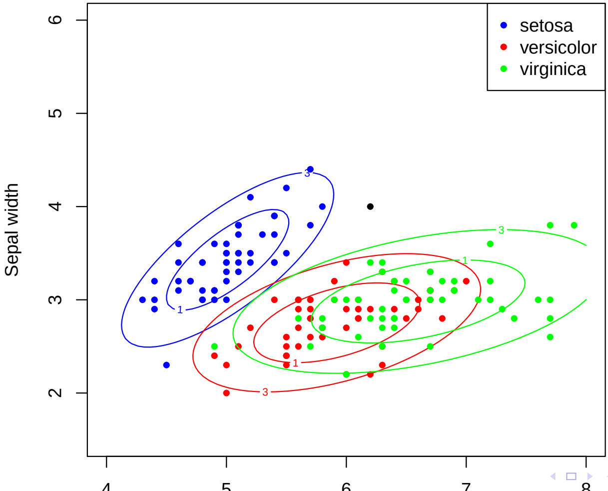
- Suppose $f_k(x)$ is the class-conditional density of X in class $G = k$
- π_k be the prior probability of class k, with $\sum_{k=1}^{K} \pi_k = 1$.
- Using Bayes Theorem:

$$\mathbb{P}(G = k | X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}$$

- In terms of ability to classify, having the $f_k(x)$ is almost equivalent to having the quantity $\mathbb{P}(G = k | X = x)$.

Classify the three species of the Iris Flower

setosa	versicolor	virginica
0.861	0.029	0.110



Discriminant Analysis

- Many techniques are there to model $f_k(x)$
- linear and quadratic discriminant analysis use Gaussian densities
- Finite mixture models (some what complicated)

$$f_k(x) = \sum_{i=1}^{I} p_i N(\mu_i, \Sigma_i)$$

- Nonparametric density estimation (very complicated)

$$f_k(x) = \sum_{i=1}^{\infty} p_i N(\mu_i, \Sigma_i) = \lim_{I \to \infty} \sum_{i=1}^{I} p_i N(\mu_i, \Sigma_i)$$

Linear Discriminant Analysis

- We model each class density as multivariate Gaussian

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

- Linear discriminant analysis (LDA) arises in the special case when we assume

$$\Sigma_k = \Sigma \quad \forall k$$

Linear Discriminant Analysis

- We want to compare two classes k and l,
- Let's look at the ratio

$$
\begin{aligned}
\log \frac{\mathbb{P}(G=k|X=x)}{\mathbb{P}(G=l|X=x)} &= \log \frac{f_k(x)}{f_l(x)} + \log \frac{\pi_k}{\pi_l} \\
&= \log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k + \mu_l)^T \Sigma^{-1} (\mu_k - \mu_l) \\
&\qquad + x^T \Sigma^{-1} (\mu_k - \mu_l)
\end{aligned}
$$

is an equation linear in x.

Linear Discriminant Analysis

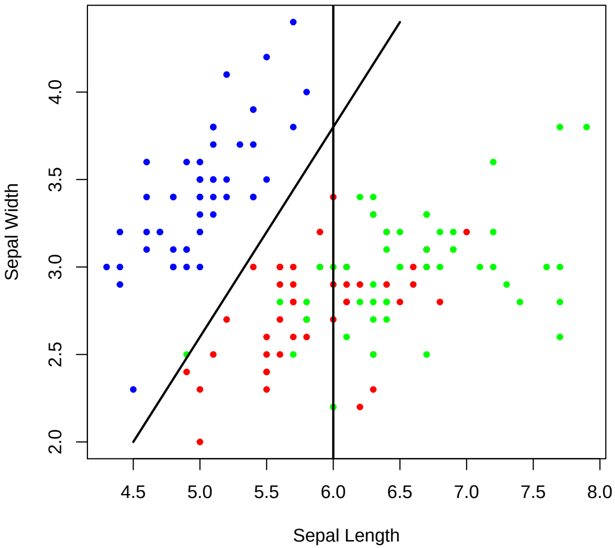
- $\Sigma_k = \Sigma \ \forall k$ cause the normalization factors to cancel, as well as the quadratic part in the exponents.
- The decision boundary between classes k and l is linear
- From above the linear discriminant functions

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

- Best decision rule:

$$G(x) = \mathrm{argmax}_k \, \delta_k(x)$$

Linear Discriminant Analysis



Linear Discriminant Analysis

- In practice we do not know the parameters of the Gaussian distributions
- Need to estimate using our training data
  - $\hat{\pi}_k = \frac{f_k}{n}$, where f_k is the number of class-k observations
  - $\hat{\mu}_k = \sum_{g_i=k} x_n / f_k$
  - $\hat{\Sigma} = \sum_{k=1}^{K} \sum_{g_i=k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T / (n - K)$
- These estimates are MLE

Two Classes LDA

- The LDA for two classes are very simple.
- The LDA rule classifies to class 2 if

$$x^T \hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) > c$$

where

$$c = \frac{1}{2}\hat{\mu}_2^T \hat{\Sigma}^{-1}\hat{\mu}_2 - \frac{1}{2}\hat{\mu}_1^T \hat{\Sigma}^{-1}\hat{\mu}_1 + \log(f_1/n) - \log(f_2/n)$$

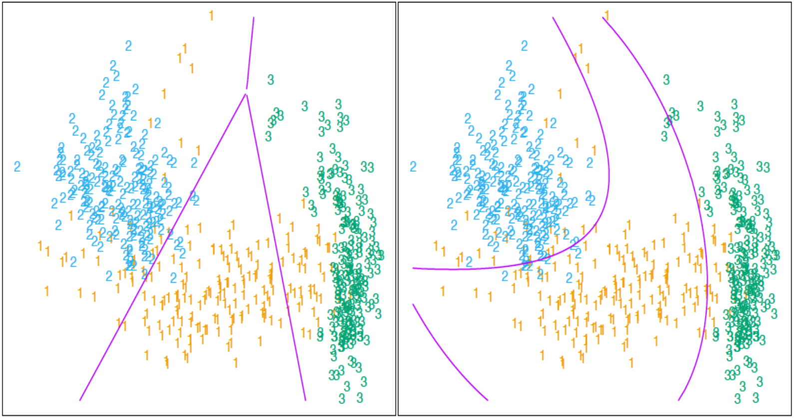
Quadratic Discriminant Analysis

- $\Sigma_k \neq \Sigma$ at least for one k
- Convinient cancellation will not work any more
- Then QDA function is

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log\pi_k.$$

- The decision boundary between each pair of classes k and l is described by quadratic equation $\{x : \delta_k(x) = \delta_l(x)\}$

LDA or QDA



source: "Introduction to Statistical Learning" by James, Witten, Hastie and Tibshirani
https://faculty.marshall.usc.edu/gareth-james/ISL/

Thank You

sourish@cmi.ac.in