QTM 347 Machine Learning

Lecture 5: LDA and QDA

Ruoxuan Xiong

Suggested reading: ISL Chapter 4

Lecture plan

- LDA and QDA

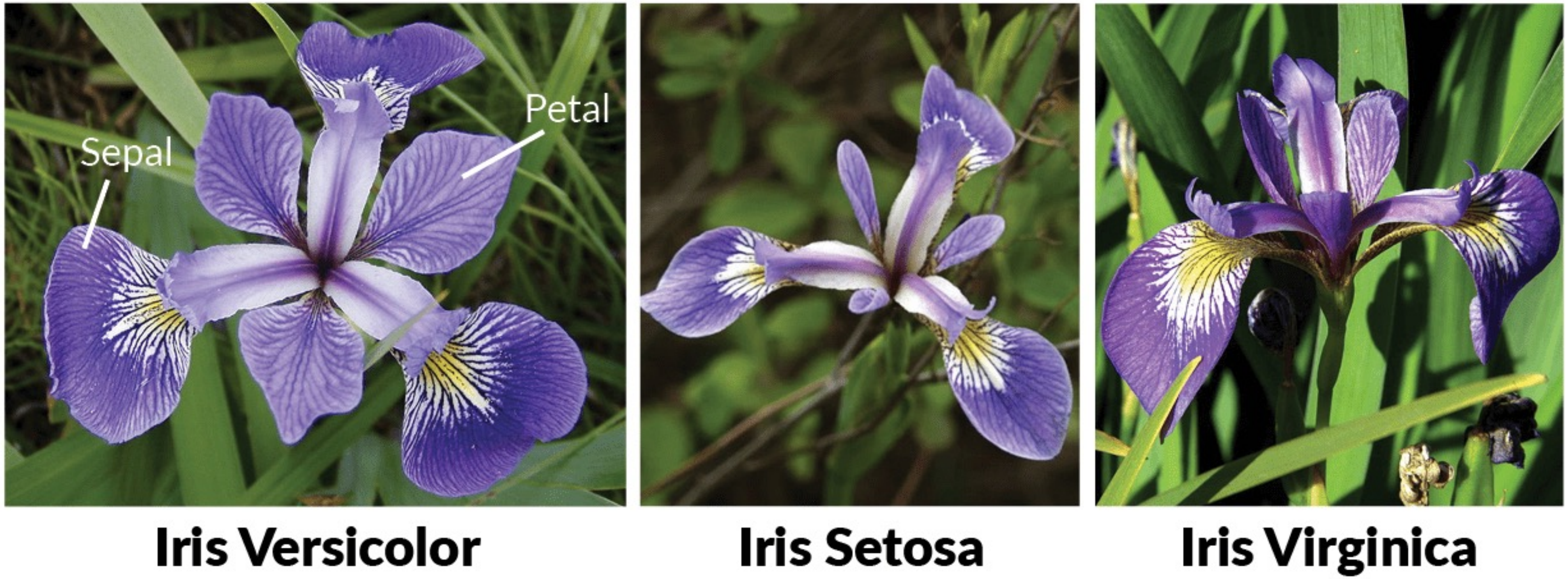
Generative vs discriminative methods

- Generative methods
  1. Model the joint probability $p(x, y)$
  2. Assume some distribution for conditional distribution of X given $Y = k$, $P(X = x|Y = k)$
  3. Bayes theorem is applied to obtain $P(Y = k|X = x)$ and classify
  - E.g., linear discriminant analysis (LDA), quadratic discriminant analysis (QDA)

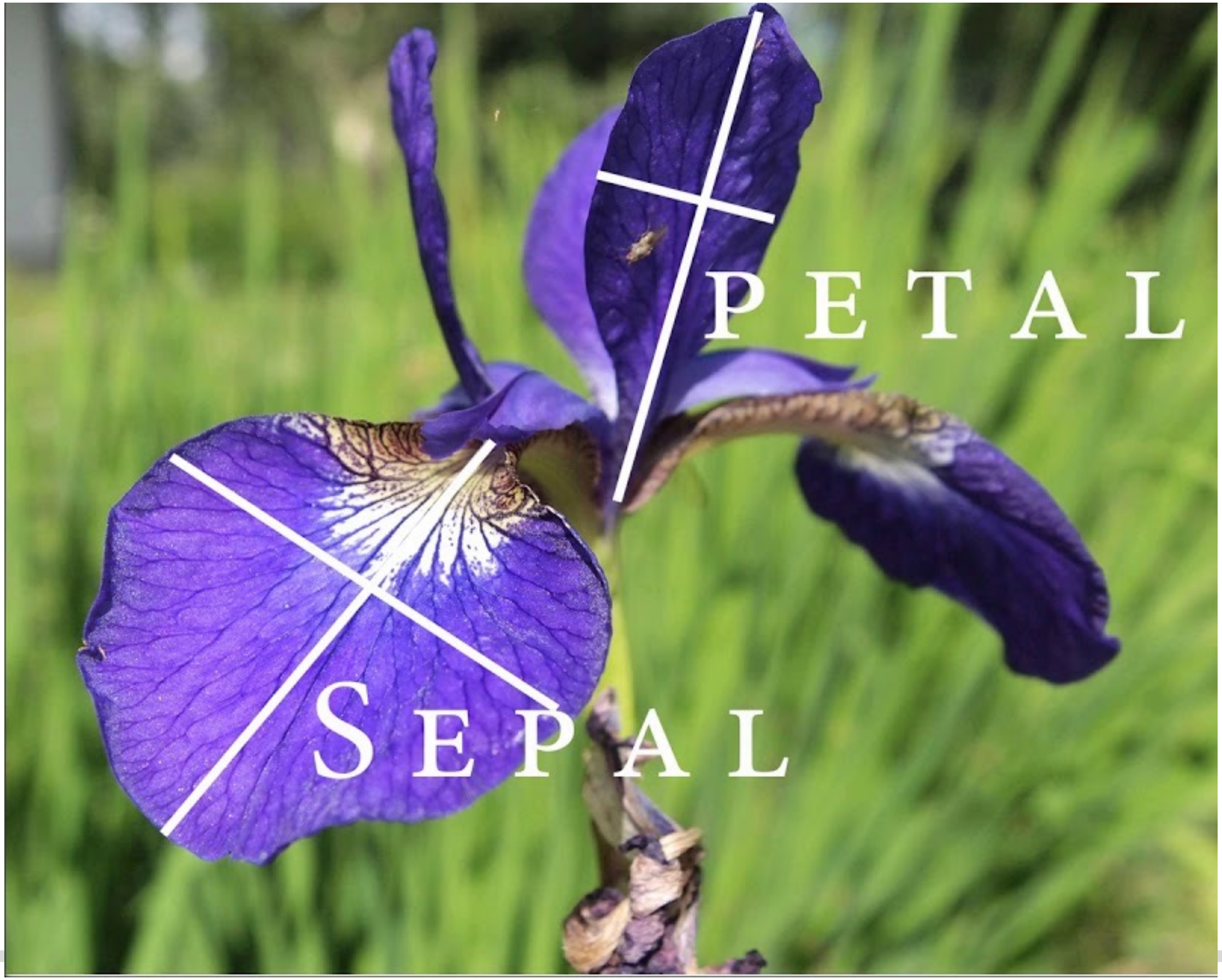
- Discriminative methods
  - Directly model $P(Y = k|X = x)$ and classify
  - E.g., logistic regression

Example: An iris data set

- Perhaps the best known database in the pattern recognition literature
- Predict class of iris plant
- There are three classes



Sepal and petal of iris

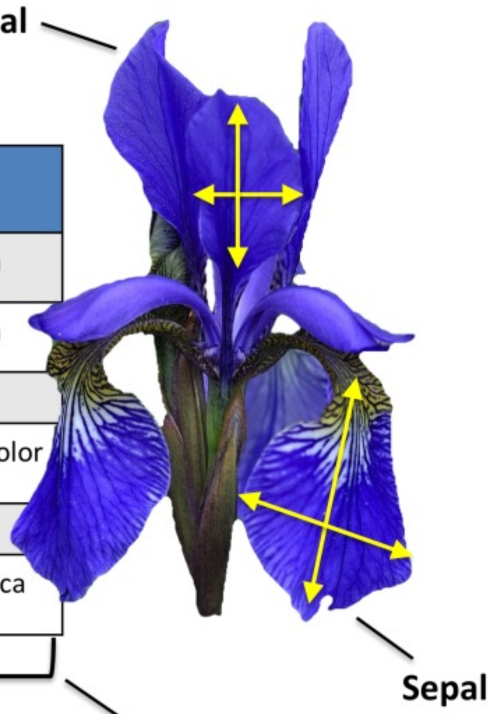


Example: An iris data set

- 50 samples from each of three class of *Iris (versicolor, setosa, virginica)*
- Four features: sepal length, sepal width, petal length, petal width

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica



Features
(attributes, measurements, dimensions)

Class labels
(targets)

Estimating $\pi_k = P(Y = k)$

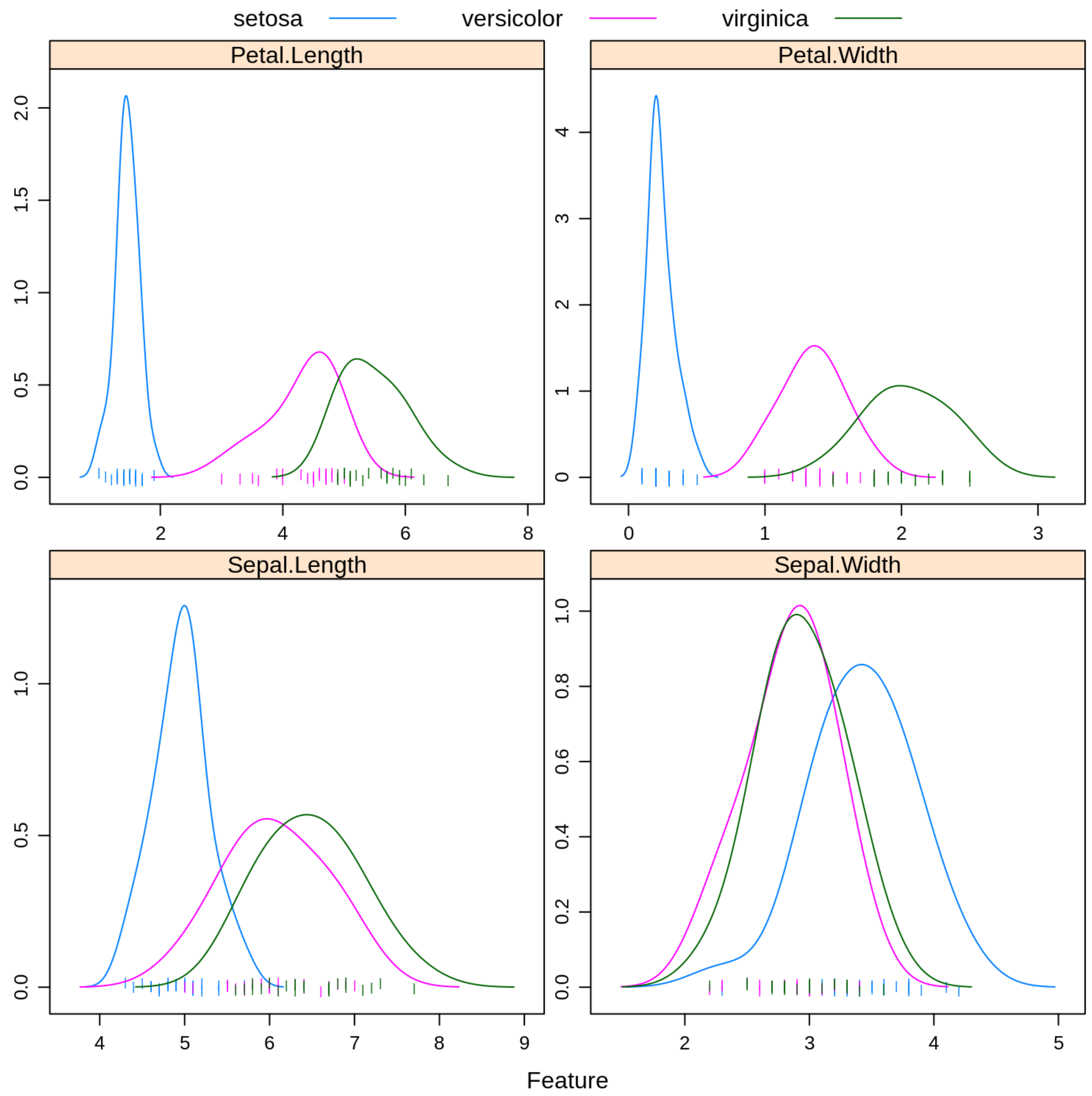
- The fraction of training samples of class k

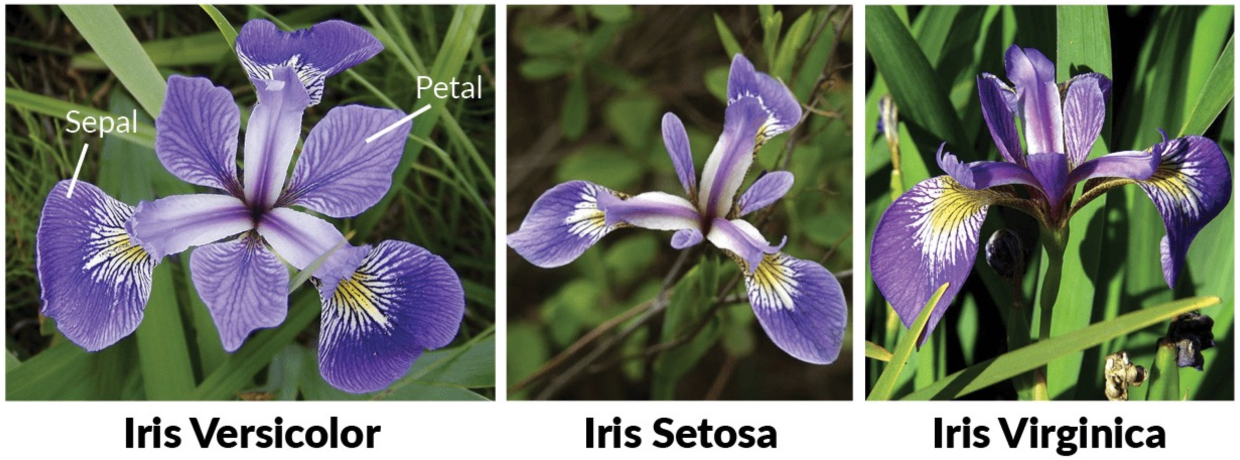
$$\hat{\pi}_k = \hat{P}(Y = k) = \frac{\#\{i: y_i = k\}}{n}$$

- Iris data: 50 samples from each of three class of *Iris (versicolor, setosa, virginica)*. Then

$$\hat{\pi}_{setosa} = \hat{\pi}_{versicolor} = \hat{\pi}_{vriginica} = \frac{50}{50 + 50 + 50} = \frac{1}{3}$$

Distribution of features





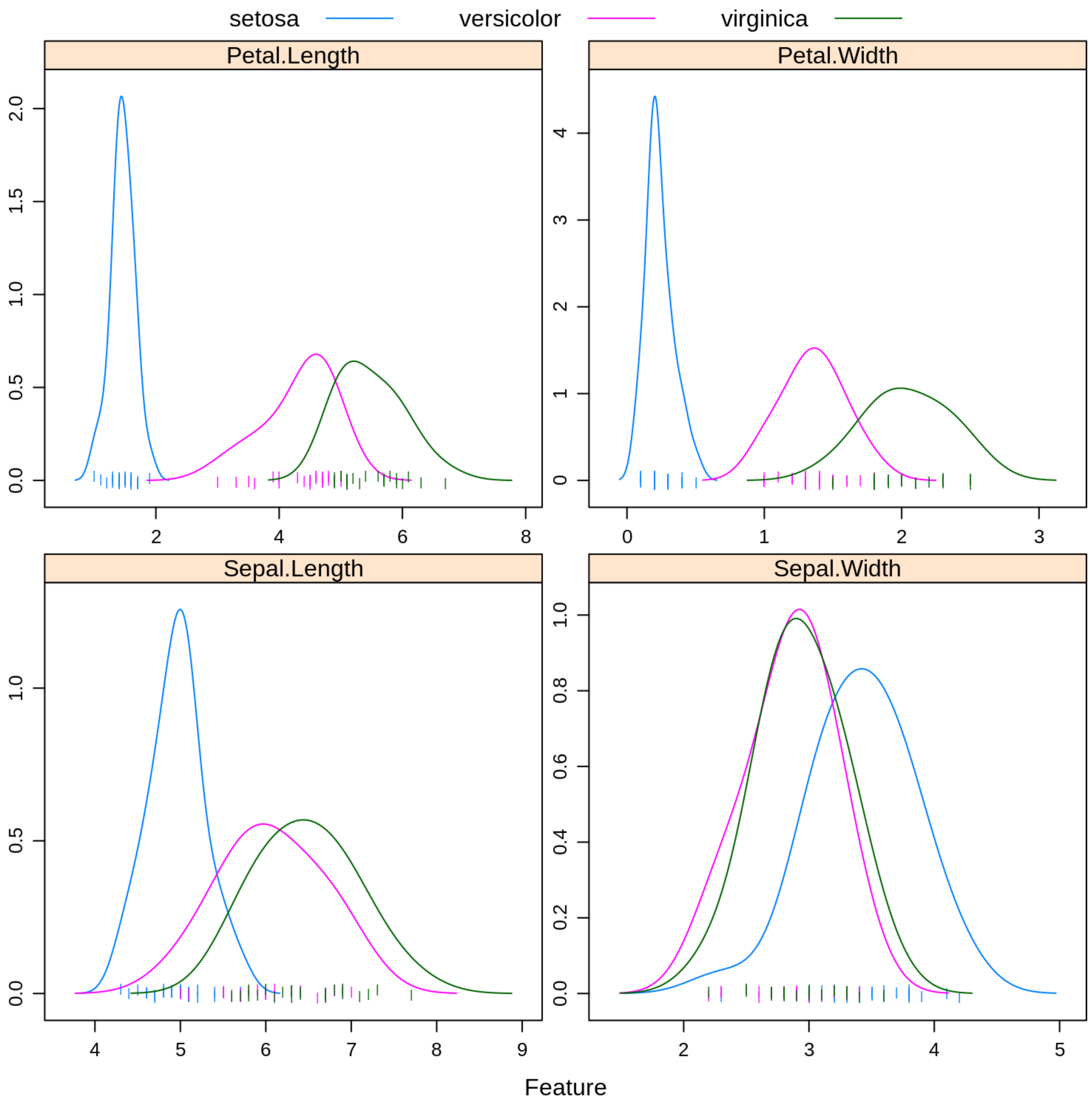
Linear discriminant analysis

- Model $P(X = x \mid Y = k)$

  - $X = \begin{bmatrix} sepal\ length \\ sepal\ width \\ petal\ length \\ petal\ width \end{bmatrix}$

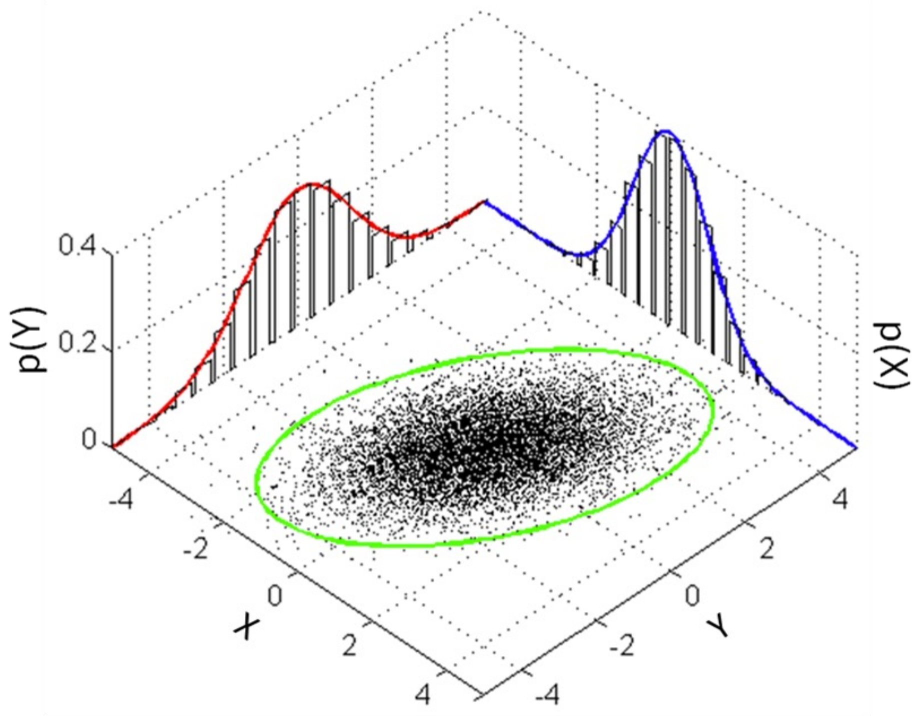
  - $Y \in \{versicolor, setosa, virginica\}$

by a *Multivariate Normal Distribution* $N(\mu_k, \Sigma)$ with mean μ_k, covariance matrix Σ



Multivariate normal distribution

- Illustration of a two-dimensional multivariate normal distribution
  - Centered at zero
  - Projection to every dimension (blue and red) is still a Gaussian

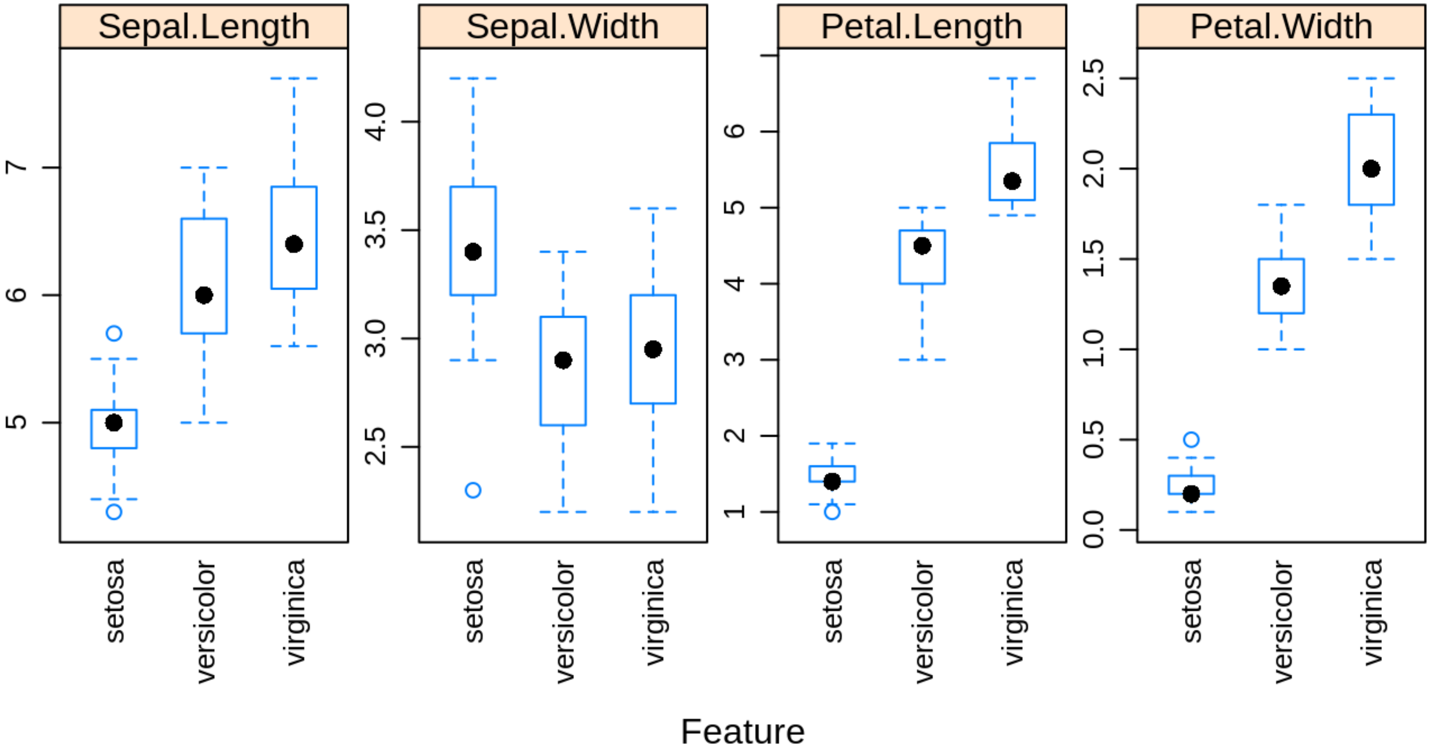


μ_k in Linear discriminant analysis

- $\mu_{setosa} = \begin{bmatrix} \overline{setosa\ sepal\ length} \\ \overline{setosa\ sepal\ width} \\ \overline{setosa\ petal\ length} \\ \overline{setosa\ petal\ width} \end{bmatrix}$

- Bar represents average value

- Black dots of setosa in the box plots

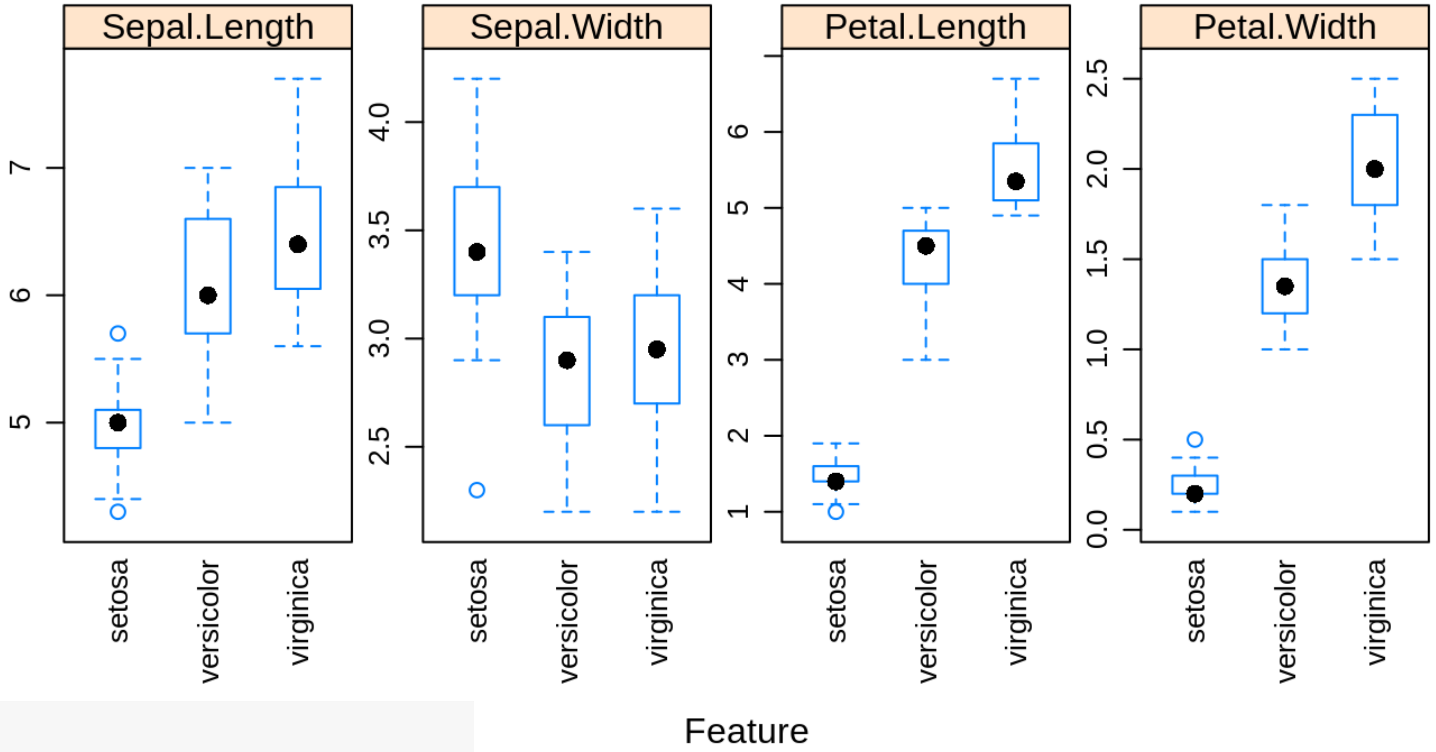


Estimating the center μ_k

- Estimate the center of each class μ_k:

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i: y_i = k} x_i$$

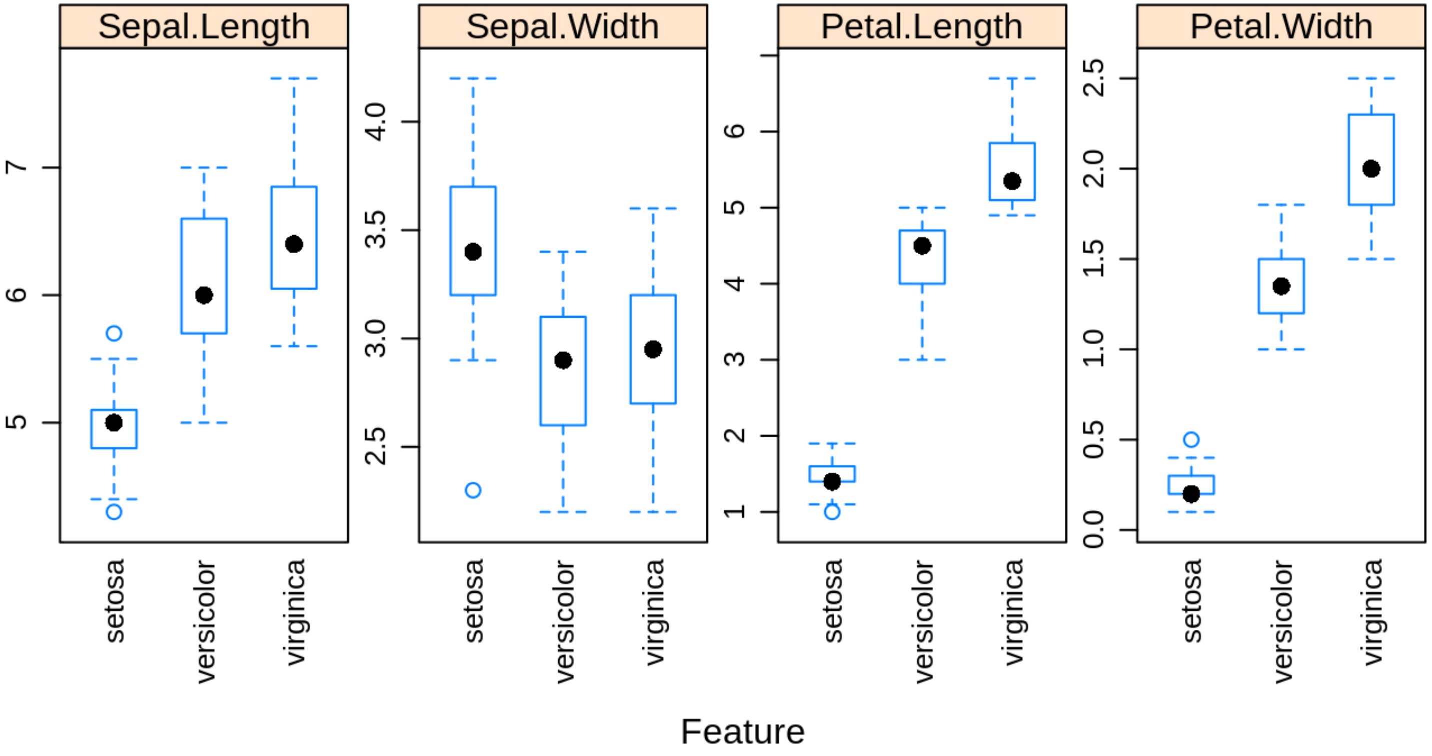
where $n_k = \#\{i: y_i = k\}$




```
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa         4.958621    3.420690     1.458621    0.237931
## versicolor     6.063636    2.845455     4.318182    1.354545
## virginica      6.479167    2.937500     5.479167    2.045833
```

Σ in Linear discriminant analysis

- Σ is the same for $versicolor, setosa, virginica$
  - Diagonal entries equal to variance of each feature for all classes
    - Proportional to the width of the box plots
  - Off-diagonal entries equal to covariance between two features for all classes
- What if Σ should be different for different class?



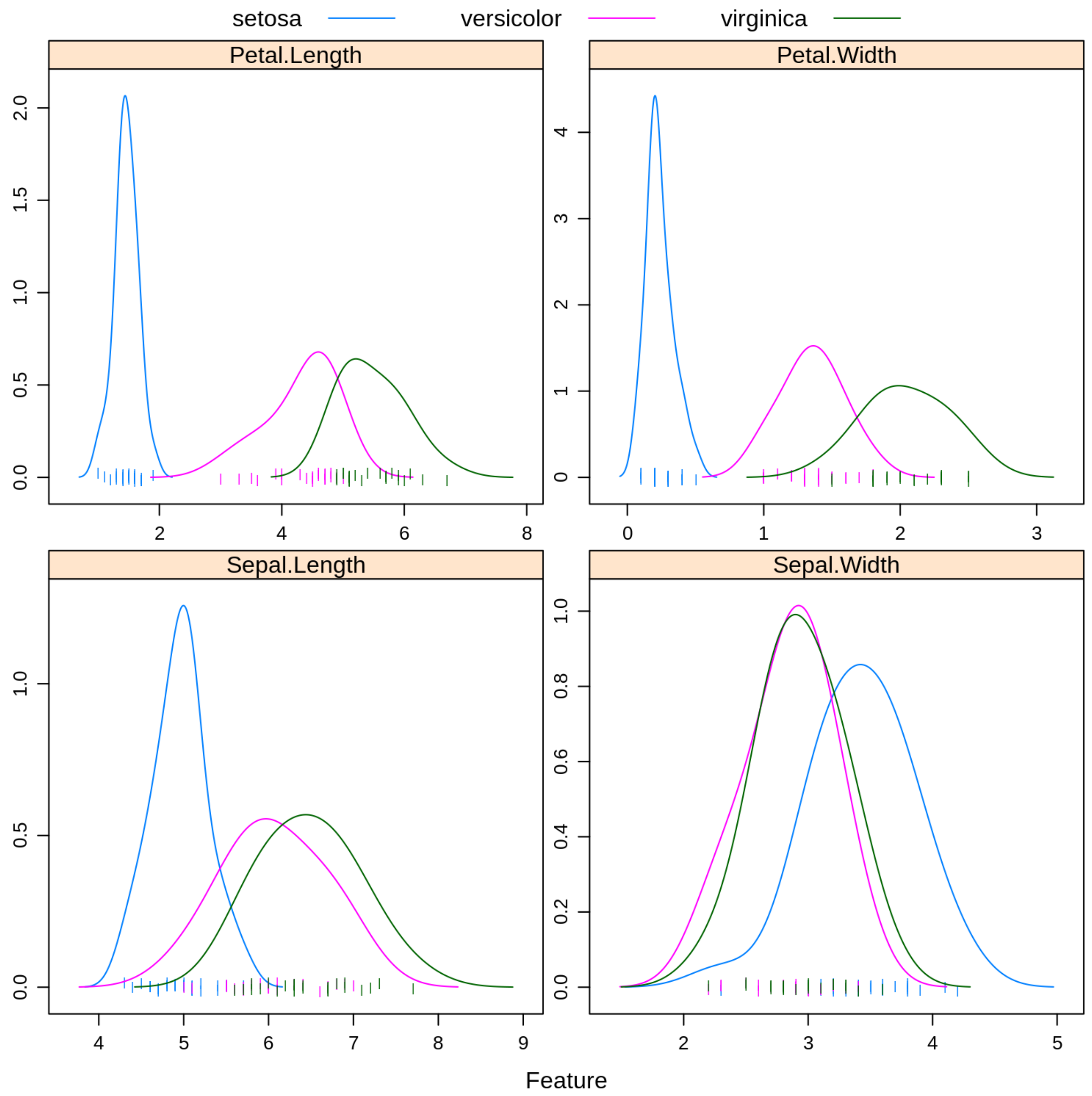
Quadratic discriminant analysis

- Model $P(X = x \mid Y = k)$

  - $X = \begin{bmatrix} sepal\ length \\ sepal\ width \\ petal\ length \\ petal\ width \end{bmatrix}$

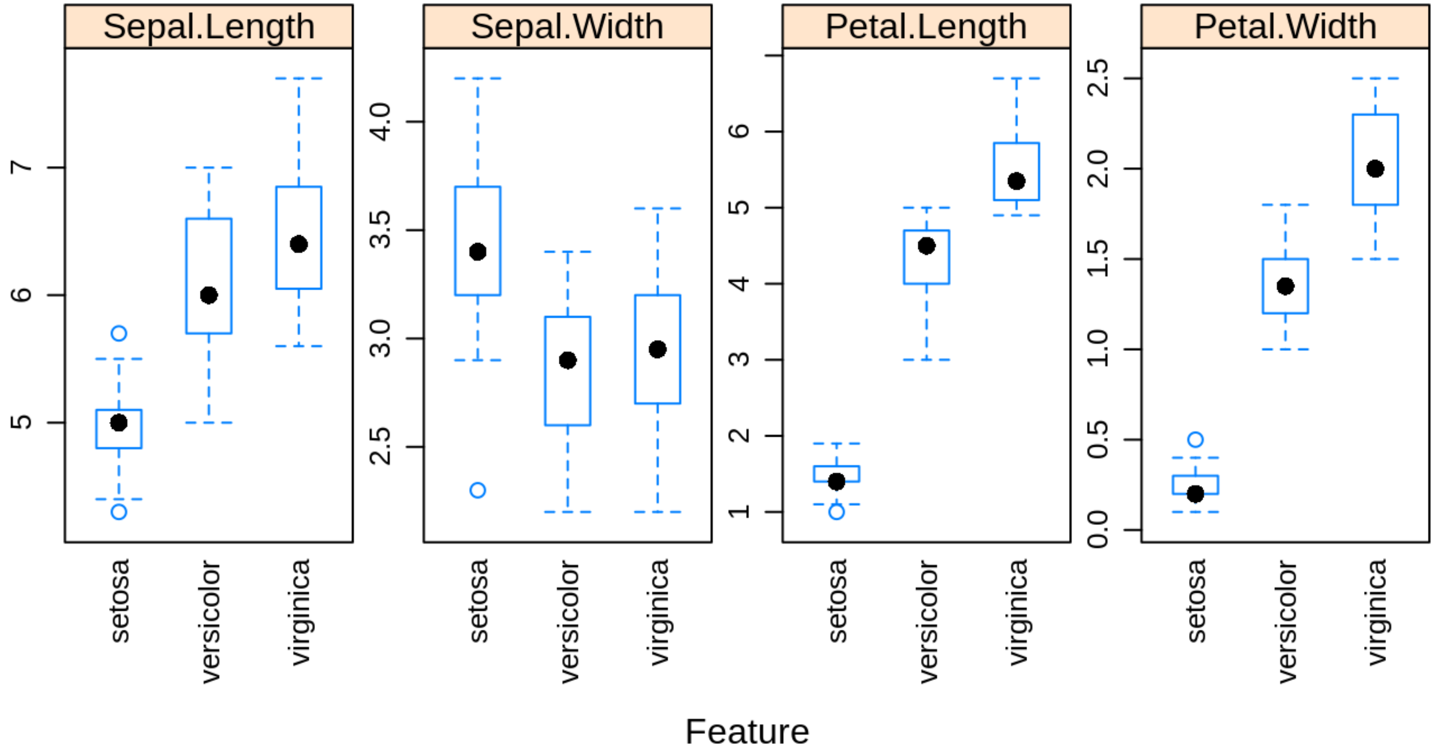
  - $Y \in \{versicolor, setosa, virginica\}$

by a *Multivariate Normal Distribution* $N(\mu_k, \Sigma_k)$ with mean μ_k, covariance matrix Σ_k



Σ_k in Quadratic discriminant analysis

- Σ_{setosa}
  - Diagonal entries equal to variance of each feature for setosa
  - Off-diagonal entries equal to covariance between two features for setosa



Estimating the covariance Σ_k in QDA

- Estimate the covariance Σ_k

$$\hat{\Sigma}_k = \frac{1}{n_k - 1} \sum_{i:y_i=k} (x_i - \hat{\mu}_k) \cdot (x_i - \hat{\mu}_k)^\top$$

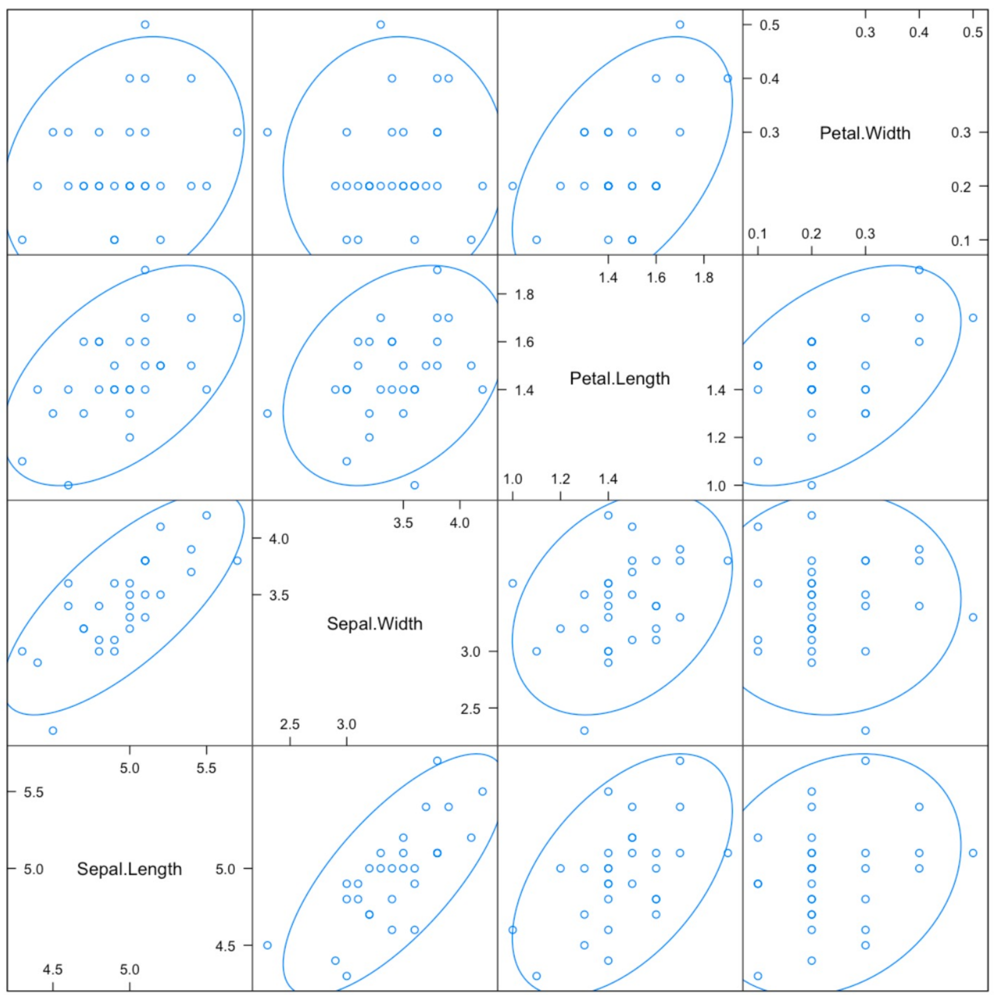
where $n_k = \#\{i: y_i = k\}$

- Example: Σ_{setosa}

```r
iris_trn_setosa <- iris_trn[iris_trn$Species == "setosa",]
cov(iris_trn_setosa[,c(1:4)])
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length  0.103226601 0.095172414  0.031798030 0.007697044
## Sepal.Width   0.095172414 0.160985222  0.025172414 0.001687192
## Petal.Length  0.031798030 0.025172414  0.035369458 0.009125616
## Petal.Width   0.007697044 0.001687192  0.009125616 0.009581281
```

setosa



Scatter Plot Matrix

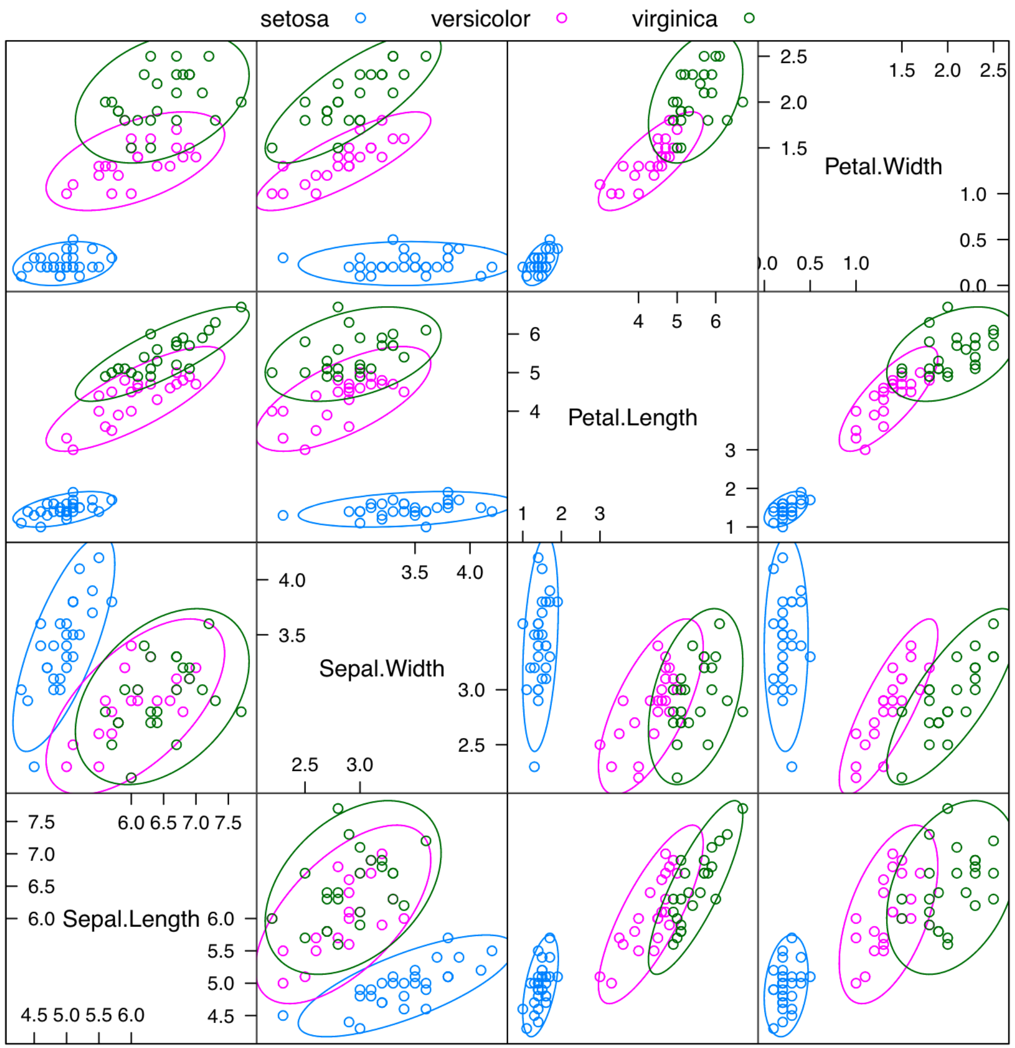
Estimating the covariance Σ in LDA

- Estimate the covariance Σ

$$\hat{\Sigma} = \sum_{k=1}^{K} \frac{n_k - 1}{n - K} \cdot \hat{\Sigma}_k$$

where $n_k = \#\{i: y_i = k\}$

- Example: $\hat{\Sigma} = \frac{n_{setosa}-1}{n-3} \cdot \hat{\Sigma}_{setosa} + \frac{n_{versicolor}-1}{n-3} \cdot \hat{\Sigma}_{versicolor} + \frac{n_{virginica}-1}{n-3} \cdot \hat{\Sigma}_{virginica}$



Scatter Plot Matrix

Summary of LDA

- For each class k, we model $P(X = x|Y = k) = f_k(x)$ as a *Multivariate Normal Distribution* $N(\mu_k, \Sigma)$ with mean μ_k and covariance matrix Σ

- We estimate $\hat{P}(X = x|Y = k)$ as $N(\hat{\mu}_k, \hat{\Sigma})$ and $\hat{P}(Y = k) = \hat{\pi}_k$

- We apply to Bayes theorem to obtain $P(Y = k \mid X = x)$

$$\hat{P}(Y = k \mid X = x) = \frac{\hat{P}(Y = k, X = x)}{\hat{P}(X = x)} = \frac{\hat{P}(X = x \mid Y = k)\hat{P}(Y = k)}{\sum_j \hat{P}(X = x \mid Y = j)\hat{P}(Y = j)}$$

Summary of QDA

- For each class k, we model $P(X = x|Y = k) = f_k(x)$ as a *Multivariate Normal Distribution* $N(\mu_k, \Sigma_k)$ with mean μ_k and covariance matrix Σ_k

- We estimate $\hat{P}(X = x|Y = k)$ as $N(\hat{\mu}_k, \hat{\Sigma}_k)$ and $\hat{P}(Y = k) = \hat{\pi}_k$

- We apply to Bayes theorem to obtain $P(Y = k \mid X = x)$

$$\hat{P}(Y = k \mid X = x) = \frac{\hat{P}(Y = k, X = x)}{\hat{P}(X = x)} = \frac{\hat{P}(X = x \mid Y = k)\hat{P}(Y = k)}{\sum_j \hat{P}(X = x \mid Y = j)\hat{P}(Y = j)}$$

Comparison between LDA and QDA

- **Decision boundary**: the set of points in which 2 classes do just as well
  - LDA has *linear* decision boundary
  - QDA has *quadratic* decision boundary
- **Bias-variance tradeoff**
  - LDA is less flexible but has a smaller variance. Small sample size n: LDA
  - QDA requires more parameters. Large sample size n: QDA