Bayesian Classifier
Linear Disciminant Analysis (LDA)
Quadratic Discriminant Analysis (QDA)

Ref: Chapter 4, *Introduction To Statistical Learning, Gareth James et al*

Bayesian Classifier Example

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Bayesian Classifier Example

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

$$P(y|x_1, ..., x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

$$y = argmax_y P(y)\prod_{i=1}^{n} P(x_i|y)$$

Bayesian Classifier Example

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	[illegible]	[illegible]	[illegible]	[illegible]	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

```
today = (Sunny, Hot, Normal, False)
```

Bayesian Classifier Example

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Bayesian Classifier Example

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

$$P(Yes|today) \propto \frac{2}{9}\cdot\frac{2}{9}\cdot\frac{6}{9}\cdot\frac{6}{9}\cdot\frac{9}{14} \approx 0.0141$$

and

$$P(No|today) \propto \frac{3}{5}\cdot\frac{2}{5}\cdot\frac{1}{5}\cdot\frac{2}{5}\cdot\frac{5}{14} \approx 0.0068$$

$$P(Yes|today) = \frac{0.0141}{0.0141+0.0068} = 0.67$$

and

$$P(No|today) = \frac{0.0068}{0.0141+0.0068} = 0.33$$

Bayes Classifier

- Bayes Classifier
  - The ideal classifier
  - A Generative Classifier
    - Assumes that the probability distributions of oservations are known
    - And they satisfy certain assumptions

Bayes Classifier

- Bayes Classifier
  - The ideal classifier
  - A Generative Classifier
    - Assumes that the probability distributions of oservations are known
    - And they satisfy certain assumptions

$$p(Y = k \mid X = x_0)$$

In a two class classifier

$$p(Y = 1 \mid X = x_0) > 0.5 \Rightarrow$$

assigned classification = 1

else, 0

Bayes Classifier

Many classifiers are based on the principles of Bayes Classifier

- They use Bayes Theorem

π_k = overall prior probability that randomly chosen probability belongs to class k

ie: $P(Y=k)$

$f_k(x) = P(X=x \mid Y=k)$ ...

then,

$$P(Y=k \mid X=x) = \frac{P(X=x \mid Y=k) \cdot P(Y=k)}{P(X=x)}$$

Linear Discriminant Analysis

then,

$$P(Y=k \mid X=x) = \frac{P(X=x \mid Y=k) \cdot P(Y=k)}{P(X=x)}$$

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

Assumptions made ...

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)$$

let us further assume that $\sigma_1^2 = \ldots = \sigma_K^2$

Therefore

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

Linear Discriminant Analysis

The following expression ...

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

Can be simplified to ...

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)$$

Linear Discriminant Analysis: Univariate

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)$$

LDA estimates the following:

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

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{k=1}^{K} \sum_{i:y_i=k} (x_i - \hat{\mu}_k)^2$$

$$\hat{\pi}_k = n_k/n$$

$$\boxed{\hat{\delta}_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k)}$$

How to use LDA?

$$\hat{\delta}_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k)$$

- Given a candidate observation ‘x’
- Find out the value of the above expression for every class ‘k’
- Assign the observation ‘x’ to that class ‘k’ for which the above expression evaluates to the largest value

When is LDA used?

- Instead of Logistic Regression
  - When the output states / response classes are **more than 2**

- Number of observations are small and approximately normally distributed (each class)
  - Multivariate Gaussian Normal Distribution

- When the classes are well separated
  - Logistic Regression **parameter estimates are unstable**
  - LDA does not have this problem

LDA Assumptions

- Normally distributed X in every class
- Same variance across classes

LDA : Multivariate Case

- Univariate
  - Distribution of x ~ N(μ_x, σ_x)
- Multivariate
  - Distribution of **X** ~ N(**μ**, **Σ**)

- In the multivariate case:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \boldsymbol{\Sigma}^{-1}(x-\mu)\right)$$

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

LDA : Multivariate Case

- Univariate
  - Distribution of x ~ N(μ_x, σ_x)
- Multivariate
  - Distribution of **X** ~ N(**μ**, **Σ**)

- In the multivariate case:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \boldsymbol{\Sigma}^{-1}(x-\mu)\right)$$

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

Assumptions

- Class specific μ
- Covariance matrix **"Σ" common to all K classes**

LDA Multivariate Case

- Bayes Decision Boundary is defined by:

$$\delta_k(x) = \delta_\ell(x)$$

$$x^T \mathbf{\Sigma}^{-1} \mu_k - \frac{1}{2} \mu_k^T \mathbf{\Sigma}^{-1} \mu_k = x^T \mathbf{\Sigma}^{-1} \mu_l - \frac{1}{2} \mu_l^T \mathbf{\Sigma}^{-1} \mu_l$$

  – Even in this case the boundary is linear in X

Can we relax some of the assumptions?

QDA results from the covariance relaxation

Quadratic Discriminant Analysis

- Assumption that remains: Gaussian Distribution
- Relaxed
  - Each class can now have its own covariance matrix

- In case of QDA
  - Distribution of **X** ~ N($\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$)

$$
\begin{aligned}
\delta_k(x) &= -\frac{1}{2}(x-\mu_k)^T \boldsymbol{\Sigma}_k^{-1}(x-\mu_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| + \log \pi_k \\
&= -\frac{1}{2}x^T \boldsymbol{\Sigma}_k^{-1} x + x^T \boldsymbol{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\mu_k^T \boldsymbol{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| + \log \pi_k
\end{aligned}
$$

Quadratic Discriminant Analysis

- Assumption that remains: Gaussian Distribution
- Relaxed: Each class has its own covariance matrix

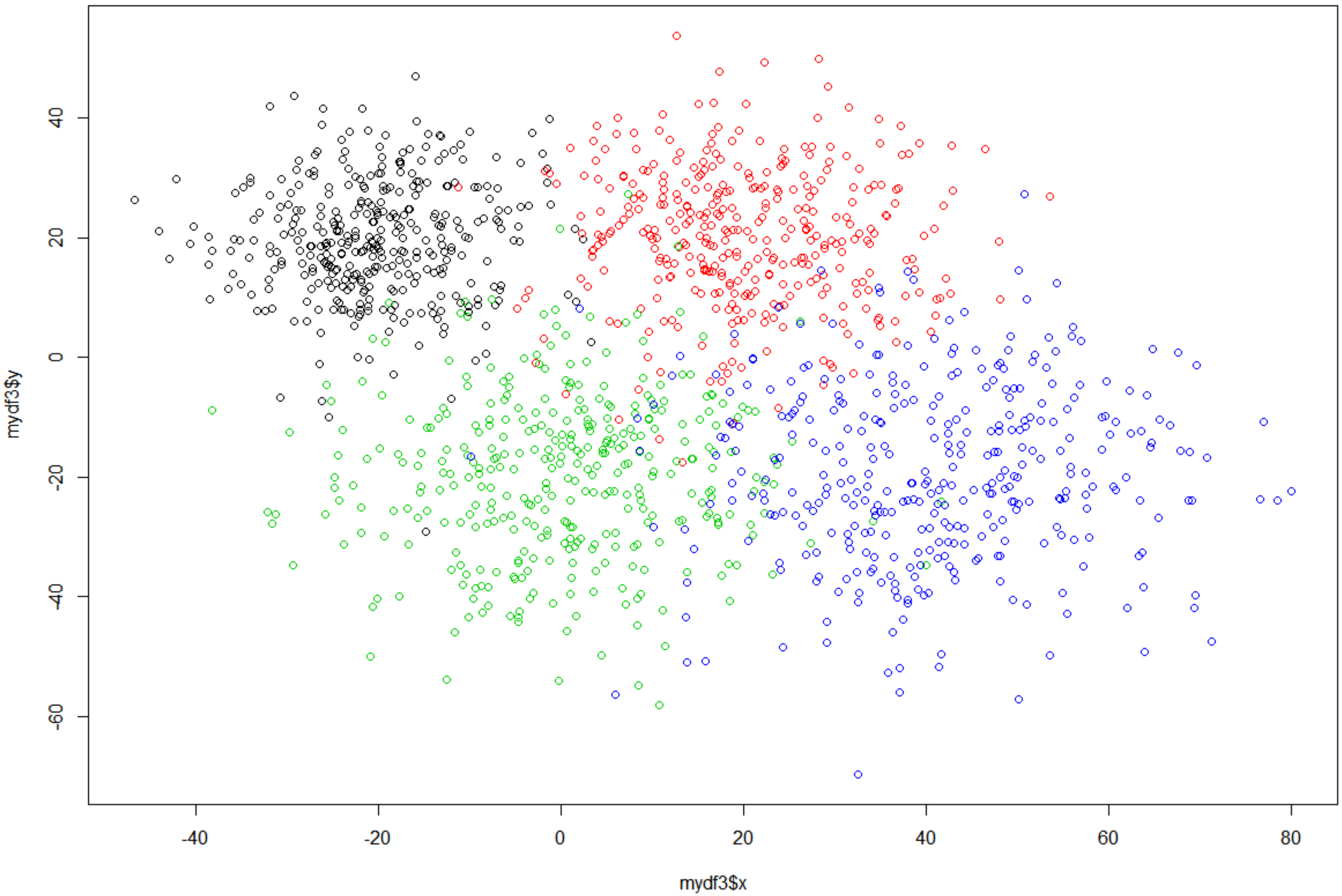
- In case of QDA
  - Distribution of **X** ~ N($\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$)

$$
\begin{aligned}
\delta_k(x) &= -\frac{1}{2}(x-\mu_k)^T \boldsymbol{\Sigma}_k^{-1}(x-\mu_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| + \log \pi_k \\
&= -\frac{1}{2}x^T \boldsymbol{\Sigma}_k^{-1} x + x^T \boldsymbol{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\mu_k^T \boldsymbol{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| + \log \pi_k
\end{aligned}
$$

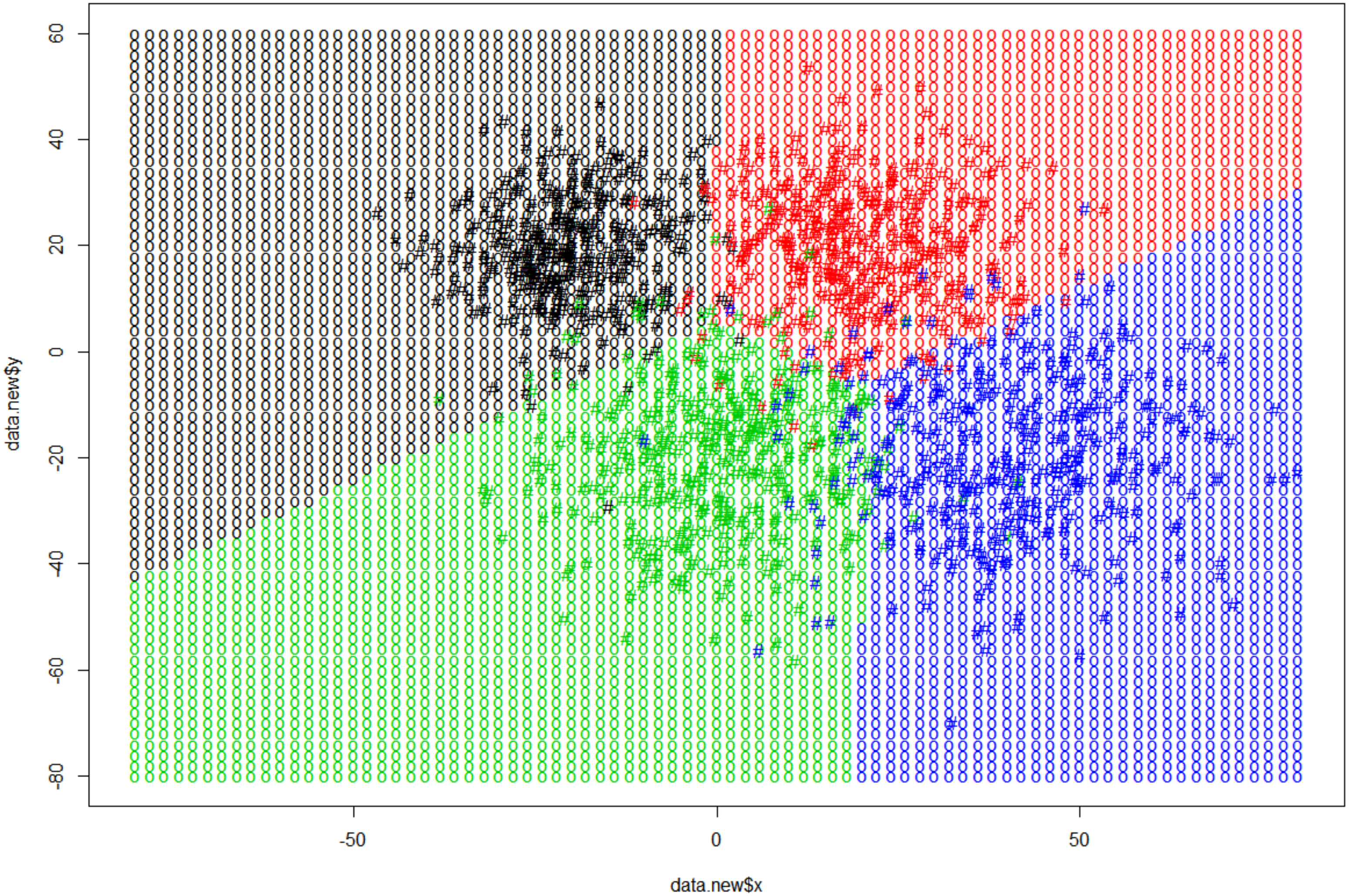
QUADRATIC TERM

hence QDA

LDA v/s QDA: Sample Data Set



LDA Boundaries



QDA Boundaries