09 MARCH 2017

ISTANBUL TECHNICAL UNIVERSITY

BLG 454E LEARNING FROM DATA

HOMEWORK 1

TEAM: BEE MEISTER
EMRE ÖZDİL – 150120138
MERVE ECEVİT – 150140115

BLG 454E Learning From Data HW1

Merve Ecevit 150140115
Emre Özdil 150120138

Q1)

T = Terrorist
S = Scanned Terrorist

$P(T) = 0,01 \quad P(\neg T) = 0,99$

$P(S/T) = 0,95 \quad P(\neg S/\neg T) = 0,95$

$P(\neg S/T) = 0,05 \quad P(S/\neg T) = 0,05$

$$P(T/S) = \frac{P(S/T)\,P(T)}{P(S)} \Rightarrow P(S) = P(S/T)\,P(T) + P(S/\neg T)\,P(\neg T)$$

$$P(T/S) = \frac{P(S/T)\,P(T)}{P(S/T)\,P(T) + P(S/\neg T)\,P(\neg T)} = \frac{0,95 \cdot 0,01}{0,95 \cdot 0,01 + 0,05 \cdot 0,99} = \frac{0,0095}{0,059}$$

$$= 0,161$$

Q2) $U(c(x^*)) = \sum_{c^{true}} U(c^{true}, c(x^*))\, p(c^{true} \mid x^*)$

the optimal decision is that which maximise the expected utility which is in above.

$$\text{Matrix} = \begin{bmatrix} 5 & 3 & 1 \\ 0 & 4 & -2 \\ -3 & 0 & 10 \end{bmatrix}$$

(c^{true} over the columns, c^{pred} over the rows)

$p(c=1|x) = 0.7$
$p(c=2|x) = 0.2$
$p(c=3|x) = 0.1$

$U = 5 \cdot 0.7 + 0 - 3 \cdot (0.7) = 1.4$

$U = 3 \cdot (0.2) + 4 \cdot (0.2) + 0 = 1.4$

$U = 1 \cdot (0.1) - 2 \cdot (0.1) + 10 \cdot (0.1) = 0.9$

The maximum expected utility value equals to 1.4 Therefore choosing both first and second column is best decision to take.

3) Probability density function (pdf) is:

$$P(x) = \frac{1}{\sqrt{2\pi\Theta^2}} e^{-\frac{1}{2\Theta^2}(x-\mu)^2}$$

Likely function is:

$$L(\mu, \Theta) = (2\pi\Theta^2)^{-n/2} \cdot e^{\left(-\frac{1}{2\Theta^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

$$\frac{\partial L(\mu,\Theta)}{\partial \mu} \longrightarrow \sum_{i=1}^{n} x_i = N\mu$$

$$\boxed{\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i} \rightarrow \text{Maximum likelihood estimator}$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{10}(115+122+130+127+149+160+152+138+149+180)$$

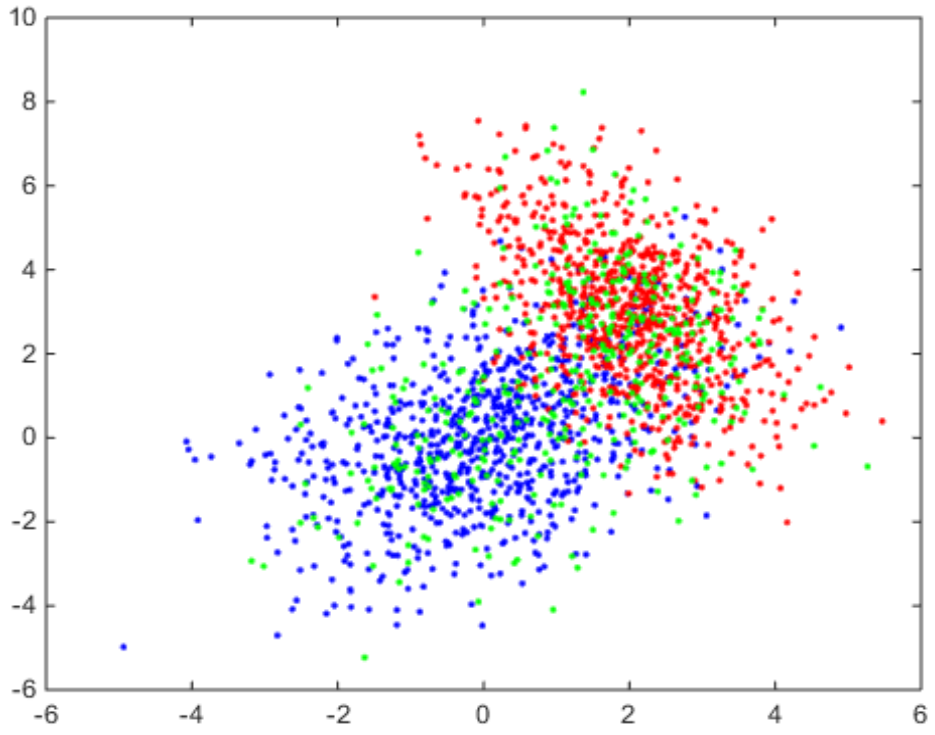
$$= 142.2 \rightarrow \text{mean, maximum likelihood estimate}$$

$$\frac{\partial L(\mu,\Theta)}{\partial \Theta^2} \rightarrow \boxed{\Theta^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{x})^2} \rightarrow \text{maximum likelihood estimator}$$

$$= \frac{1}{10}\left((180-142.2)^2 + \dots + (115-142.2)^2\right)$$

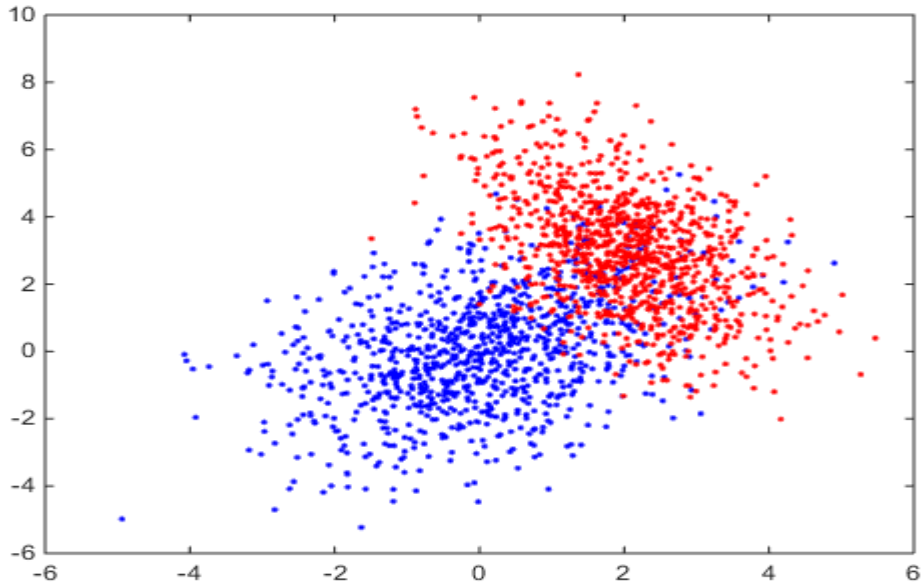
$$= 347.96 \rightarrow \text{variance}$$

Q4) Training Set and Test Set



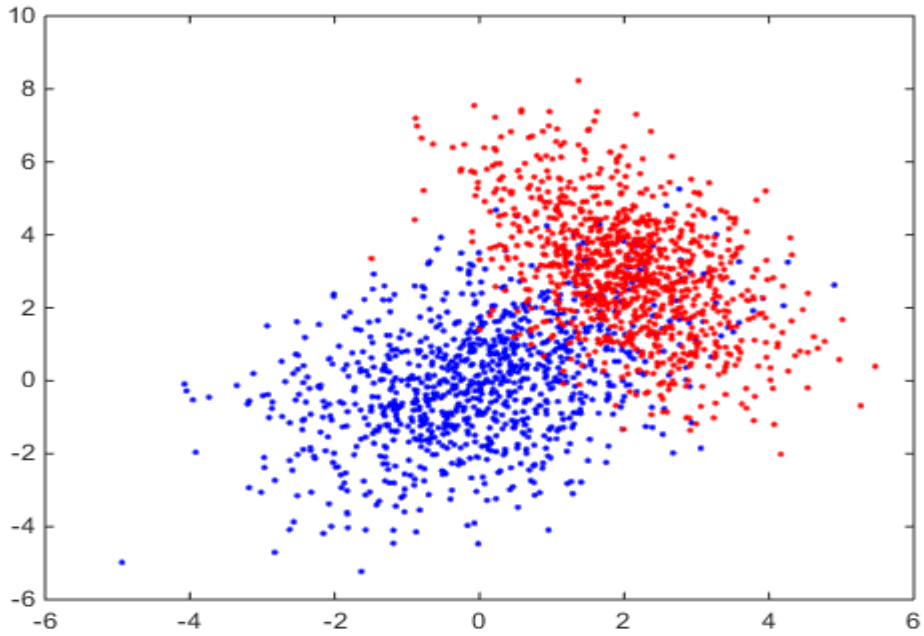
- Elements of class 0 -> blue points
- Elements of class 1 -> red points
- Test elements -> green points

Using KNN3 Classification



- Elements of class 0 -> blue points
- Elements of class 1 -> red points

Using KNN5 Classification



- Elements of class 0 -> blue points
- Elements of class 1 -> red points

Accuracy of KNN3 and KNN5

Since KNN5 uses 5 point to decide the class of point, it gives better result than KNN3.

```
>> LFD454E_Hw1
Accuracy of KNN3 = 0.8875

Accuracy of KNN5 = 0.8900
```