IAML Past Paper 2014-2015

1. Naive Bayes classification

Gaussian Distribution $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

Dataset

user session	1	2	3	4	5	6
rate of ls	4	1	1	$-\sqrt{3}$	$\sqrt{3}$	0
rate of cd	0	$\sqrt{3}$	$-\sqrt{3}$	3	3	0
	from genuine users			from intruders		

a. use Naive Bayes Classifier. Demonstrate that the following is true.

$p_- = 0.5 \quad \mu_{-,ls} = 2 \quad \mu_{-,cd} = 0 \quad \sigma^2_{-,ls} = 2 \quad \sigma^2_{-,cd} = 2$

$p_+ = 0.5 \quad \mu_{+,ls} = 0 \quad \mu_{+,cd} = 2 \quad \sigma^2_{+,ls} = 2 \quad \sigma_+^2{}_{cd} = 2$

$p_- = 6/12 = 0.5$

$p_+ = 6/12 = 0.5$

$\mu_{-,ls} = \frac{4+1+1}{3} = 2 \qquad \mu_{+,ls} = \frac{-\sqrt{3}+\sqrt{3}+0}{3} = 0$

$\mu_{-,cd} = \frac{\sqrt{3}-\sqrt{3}}{3} = 0 \qquad \mu_{+,cd} = \frac{3+3+0}{3} = 2$

$\sigma^2_{-,ls} = \frac{(4-2)^2 + (1)^2 + (1)^2}{3} = 2$

$\sigma^2_{-,cd} = \frac{(\sqrt{3})^2 + (\sqrt{3})^2}{3} = 2$

$\sigma^2_{+,ls} = \frac{(\sqrt{3})^2 + (\sqrt{3})^2}{3} = 2$

$\sigma_{+,cd} = \frac{(3-2)^2 + (3-2)^2 + (2)^2}{3} = \frac{1+1+4}{3} = 2$

Checks out!

b. Classify $[2, \sqrt{3}]$ using the classifier.

$$p(2|+) = \frac{1}{\sqrt{2\pi 2^2}} \exp\left\{-\frac{\cancel{(2-0)^2}}{2\cancel{(2)^2}}\right\}$$

$$= \frac{1}{\sqrt{2\pi 2^2}} e^{-1/2} = 0.12$$

$$p(\sqrt{3}|+) = \frac{1}{\sqrt{2\pi(2)^2}} \exp\left\{-\frac{(\sqrt{3}-2)^2}{2(2)^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi(2)^2}} \exp(-7+4\sqrt{3}) = 0.185$$

$$p(2|-) = \frac{1}{\sqrt{2\pi(2)^2}} \exp\{0\} = 0.199$$

$$p(\sqrt{3}|-) = \frac{1}{\sqrt{2\pi(2)^2}} \exp\left\{-\frac{(\sqrt{3}-0)^2}{2(2)}\right\}$$

$$= \frac{1}{\sqrt{2\pi(2)^2}} \exp\{-3/4\} = 0.0942$$

$$p(x|+) = e^{-1/2} \times e$$

c. Yes. Probabilities multiplied by 1.
Ignore that Is.

d. Will the classifier in (a) correctly classify all sessions?

plug in all the values

e. DB

The place where you can't decide where an inst. belongs

$P(K=+1 | X)=0.5$

$P(K=-1 X)=0.5$

$DB \rightarrow \frac{P(-1|x)}{P(+1|x)} = 1$

f. Hacker wants to remove session

Remove the most important

not 6 because (0, 0)
Remove 1 because it changes mean
$\Rightarrow$ moves the Gaussian
$\mu \rightarrow 2 \quad \mu' = 2/3$

g. Relax Independence Assumption

$P(k=-1 \; x_1, x_2) = P(x_1, x_2 | k=-)$

$P(AB|C) = P(B|AC)\,P(A|C) = P(B|C) \quad |C)$

$\hookrightarrow$ Bayesian Rule ♡

2. Clustering and PCA

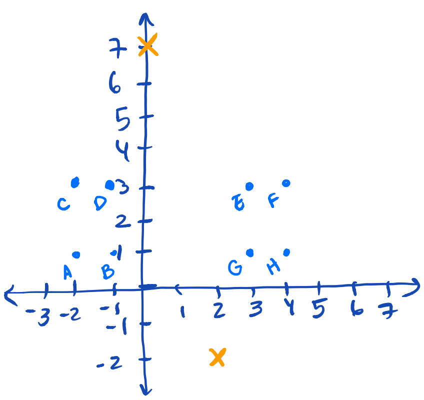
Consider a dataset with 2-real valued attributes and 8 instances:

6.32 a $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ b $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ c $\begin{bmatrix} -2 \\ 3 \end{bmatrix}$ d $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ e $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$ f $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$

g $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ h $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$

a. Run K-means clustering algorithm to convergence.
Initial centroids:

$$m_1 = [0, 7] \quad m_2 = [2, \overset{5}{-2}]$$



$C_1 \rightarrow$ C, D, E, F

$C_2 \rightarrow$ A, B, G, H