JichenDai

SID: 10448922

1. Short Questions

a. **False**

b. **True**

c. **True**

d. **False**

e. **b**

f. **c**

g. **h**

h. **b**

i. **a, b, d**

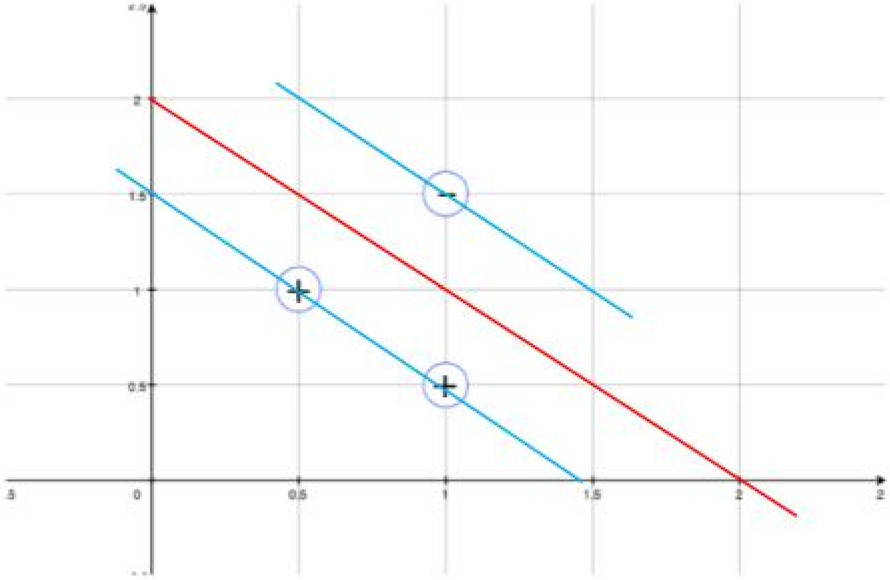
j. **c**

k. **Yes,**

k-means gives hard assignment while Gaussian mixture clustering gives soft assignment to each data point.

So, if the variance of Gaussian mixture is large, those points which are on the edge of clusters may be assigned to different groups in the Gaussian mixture solution.

2. SVM



(b) **2** support vectors

(c)

1. SVM is robust and the possibility of over-fitting is less
2. If there is a clear margin of separation between classes, SVM can work very well.
3. SVM is effective when the number of dimensions is greater than the number of samples.
4. It scales relatively well to high dimensional data.

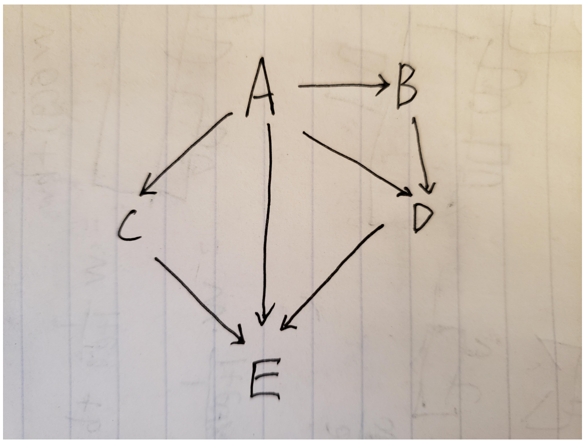
3. AdaBoost

(1) $\alpha = \frac{1}{2}\log_2\frac{1-\varepsilon}{\varepsilon} = \frac{1}{2}\log_2\frac{7/8}{1/8} = \frac{1}{2}\log_2 7$

(2) False， when the data are not linear separable using a group of weak classifiers.

4. Bayesian Network

(a)



(b) 1 + 2 + 2 + 4 + 8 = **17** independent parameters

(c) P(A,B,C,D,E) = P(A)P(B|A)P(C|A,B)P(D|A,B,C)P(E|A,B,C,D)

(d) 1 + 2 + 4 + 8 + 16 = **31** independent parameters

5. Decision Tree

(1) $P(Y=T) = \frac{2}{3}$

$P(Y=F) = \frac{1}{3}$

$H(Y) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \doteq 0.9333$

(2) $P(Y=T|GPA=L) = \frac{1}{2}$ $\quad$ $P(Y=F|GPA=L) = \frac{1}{2}$

$P(Y=T|GPA=M) = \frac{1}{2}$ $\quad$ $P(Y=F|GPA=M) = \frac{1}{2}$

$P(Y=T|GPA=H) = 1$ $\quad$ $P(Y=F|GPA=H) = 0$

$H(Y|GPA)$

$$= -\frac{1}{3}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{1}{3}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{1}{3}\left(1\log 1 + 0\log 0\right)$$

$$\doteq \frac{2}{3}$$

(3) $P(Y=T|S=T) = 1$ $\quad$ $P(Y=F|S=T) = 0$

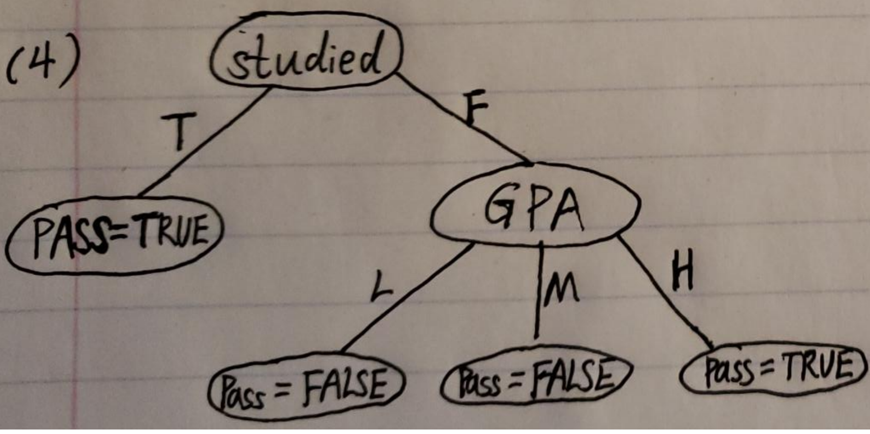
$P(Y=T|S=F) = \frac{1}{3}$ $\quad$ $P(Y=F|S=F) = \frac{2}{3}$

$H(Y|\text{Studied})$

$$= -\frac{1}{2}\left(1\log 1 + 0\log 0\right) - \frac{1}{2}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right)$$

$$= 0.4667$$

(4)



6. Neural Network

(1) This a regression problem.

$$\hat{y} = Wh + b_w$$

$$= W\frac{1}{1+e^{-g}} + b_w$$

$$= W\frac{1}{1+e^{-(Vx+b_v)}} + b_w$$

(2)

numWeight
= Size_input * Size_hidden + Size_hidden * Size_output
= 2*3 + 3*1
= 9

numBias
= Size_hidden + Size_output
= 3 + 1
= 4

Total_param = 9 + 4 = 13

(3)

$$\frac{\partial J}{\partial \hat{y}} = (y-\hat{y})(-1) = (\hat{y}-y)$$

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial W} = (\hat{y}-y)h$$