CS 559 final exam

10457625

Chenxu Wang

1. (a) : False

(b) : True

(c) : True

(d) : False

(e) : (b)

(f) : (c)

(g) : (h)

(h) : (b) (c)

(i) : (a)(b)(d)

(j) : (c)

(k) : Yes. K-means is a hard assignment approach, each data is associated uniquely with one cluster. GMM is a soft assignment, based on the posterior probabilities. K-means is as a special case when there is the same variance for each Gaussian component.

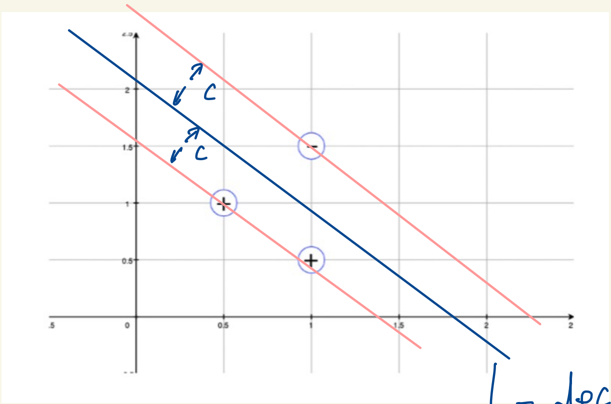
g.

$X \backslash Y$	0	1
0	0.1	0.3
1	0.2	0.4

when $X=1$. $P(X=1) = 0.3 + 0.4 = 0.7$

$$P(Y=1 \mid X=1) = \frac{P(X=1, Y=1)}{P(X=1)} = \frac{0.4}{0.7} = \frac{4}{7}$$

2. (a)



↳ decision boundary.

(b)

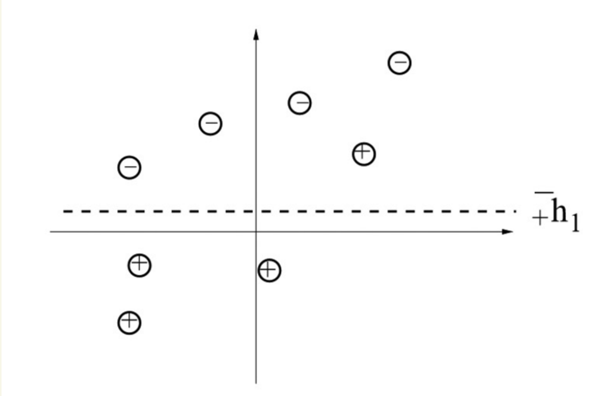
There are 3 support vector which are on the two pink lines.

(c). 1. It can obtain an unique result.

2. There is a good performance for classification.

3. Using appropriate kernel function, it can solve more complex problem. in the high dimension.

3.



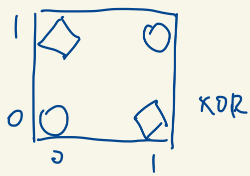
(a) The error rate of h_1:

$$\varepsilon_1 = \frac{\sum_n u_n^{(1)} \delta(h_1(x_n) \neq y_n)}{Z_1}, \quad Z_1 = \sum_n u_n^{(1)},$$

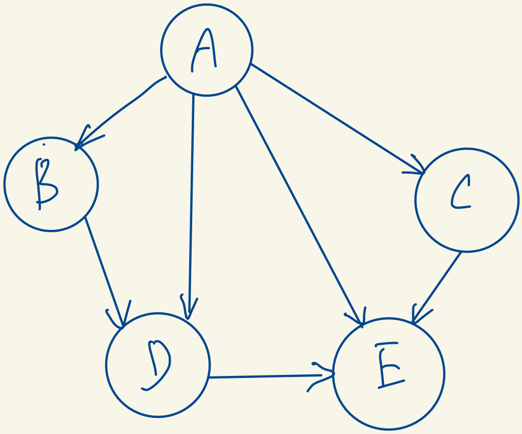
$$\therefore \text{LHS} = \frac{1}{8}$$

$$\alpha = \frac{1}{2} \cdot \ln\left(\frac{1-\varepsilon_1}{\varepsilon_1}\right) = \frac{1}{2}\left(\ln \frac{1-\frac{1}{8}}{\frac{1}{8}}\right) = \frac{1}{2} \cdot \ln 7. = 0.97.$$

(b). False, Non-linear separable dataset cannot computed by Adaboost. like XOR problem.



4. (a).



(b) $$P(A,B,C,D,E) = \underbrace{P(A)}_{1}\cdot\underbrace{P(B|A)}_{2}\cdot\underbrace{P(C|A)}_{2}\cdot\underbrace{P(D|A,B)}_{4}\cdot\underbrace{P(E|A,C,D)}_{8}$$

$$= 1+2+2+4+8 = 17$$

(c) $$P(A,B,C,D,E) = \underbrace{P(A)}_{1}\cdot\underbrace{P(B|A)}_{2}\cdot\underbrace{P(C|A,B)}_{4}\cdot\underbrace{P(D|A,B,C)}_{8}\cdot\underbrace{P(E|A,B,C,D)}_{16}$$

(d)

$$1+2+4+8+16 = 31.$$

5. (a). $P(Y=T)=\frac{2}{3}$ $P(Y=F)=\frac{1}{3}$

$$H(Y) = -[P(Y=T)\log P(Y=T) + P(Y=F)\cdot\log P(Y=F)]$$

$$= -[\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}] = -\frac{1}{3}\log 3^{-1} - \frac{2}{3}(\log 3^{-1} + \log_2 2)$$

$$= 0.533 + 0.4 = 0.9333$$

(b) $P(GPA=L)=\frac{1}{3}$ $P(GPA=M)=\frac{1}{3}$ $P(GPA=H)=\frac{1}{3}$

$$H(Y|GPA) = -[\frac{1}{3}(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}) + \frac{1}{3}(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}) + \frac{1}{3}(1\cdot\log 1 + 0)] = \frac{2}{3}$$

$$= 0.6667$$

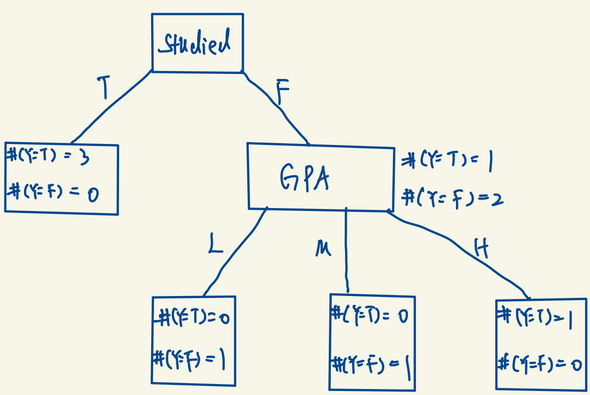
(c) $P(\text{Studied}=T)=\frac{1}{2}$ $P(\text{Studied}=F)=\frac{1}{2}$

$$H(Y|\text{Studied}) = -[\frac{1}{2}(1\cdot\log 1 + 0) + \frac{1}{2}(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3})] = 0.4667$$

(d) $IG(GPA) = 0.9333 - \frac{2}{3} = 0.2663$ $IG(GPA) < IG(\text{Studied})$

$IG(\text{Studied}) = 0.9333 - 0.4667 = 0.466$

$\therefore$ We make the first decision by "studied"



6. (a) $\underset{3\times1}{g} = \underset{3\times2}{V} \cdot \underset{2\times1}{X} + bv_{3\times1}$

$h = \sigma(g)_{3\times1}$

$\hat{y} = \underset{1\times3}{W} \cdot \underset{3\times1}{h} + bw_{1\times1}$

$\therefore \hat{y} = W \cdot \sigma(VX + bv) + bw$

(b). Since the input layer has 2 units & a hidden layer with 3 units. $\therefore$ shape. $V = 3\times2$. shape. $bv = 3\times1$

Since the output layer with 1 unit. $\Rightarrow$ $\therefore$ # weight $= 6+3 = 9$

$\therefore$ shape. $W = 1\times3$ shape. $bw = 1\times1$ # bias $= 3+1 = 4$.

(c). Using Chain Rule:

$$\frac{\partial J}{\partial W_{1\times3}} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} = \frac{1}{2} \cdot 2(y - \hat{y}) \cdot (-1) \cdot h^T$$

$$= (\hat{y} - y) \cdot h^T$$