Solver : Tan Wei Jie, Chester

Email Address : C150153@e.ntu.edu.sg

FINISH STRONG!

CE4041 : Machine Learning, PYP : 17/18 S2

1ai False.

1aii False.

1aiii False.

1aiv True.

1av False.

1avi False.

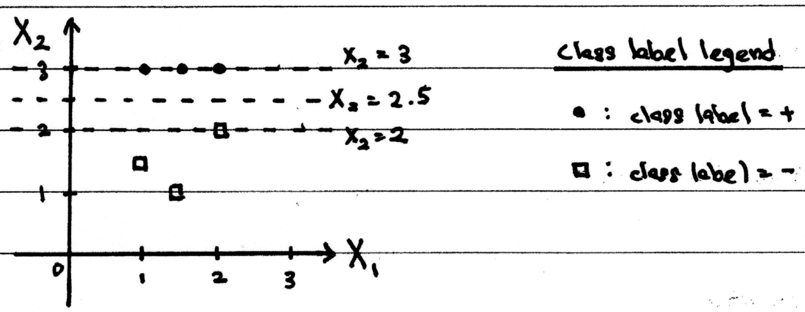
1avii False.

1aviii False.

1aix True.

1ax True.

1bi



Decision boundary : $X_2 = 2.5$

Parallel hyperplanes passing through closest data points : $X_2 = 2$ and $X_2 = 3$

Support vectors = (1, 3), (1.5, 3), (2, 3), (2, 2)

1bii Test data point = (2, 2.2)

Euclidean distance between data points $\bar{x}_i$ and $\bar{x}_j$: $d(\bar{x}_i, \bar{x}_j) = \sqrt{\sum_{k=1}^{d} (x_{ik} - x_{jk})^2}$

Data point	Class label	Distance to test data point	K where data point is K-NN
P1	+	0.8	2
P2	-	1.3	6
P3	+	1.28	5
P4	+	0.94	3
P5	-	0.2	1
P6	-	1.22	4

3-NN and majority voting:

Votes for class label + = 2

Votes for class label - = 1

Classify test data point as class label +.

1biii 3-NN and distance-weight voting:

Distance-weight votes for class label + = $\frac{1}{0.8^2} + \frac{1}{0.94^2} = 2.69$

Distance-weight votes for class label - = $\frac{1}{0.2^2} = 25$

Classify test data point as class label -.

FINISH STRONG!

2a $P(Y=1 \mid X_1=1, X_2=1, X_3=1)$

$= \frac{P(X_1=1, X_2=1, X_3=1 \mid Y=1)P(Y=1)}{P(X_1=1, X_2=1, X_3=1)}$

$= \frac{P(X_1=1 \mid Y=1)P(X_2=1 \mid Y=1)P(X_3=1 \mid Y=1)P(Y=1)}{P(X_1=1, X_2=1, X_3=1)}$

$= \frac{0.3 \times (1-0.4) \times (1-0.6) \times 0.5}{0.1}$

$= 0.36$

$P(Y=0 \mid X_1=1, X_2=1, X_3=1)$

$= 1 - P(Y=1 \mid X_1=1, X_2=1, X_3=1)$

$= 1 - 0.36$

$= 0.54$

2b $A=0$: patient A does not have cancer, $A=1$: patient A has cancer, $P(A=1) = 0.1$

$\lambda_{00} = 0$

$\lambda_{11} = 0$

$\lambda_{10} = 0.1$

$\lambda_{01} = 1$

$\lambda_{2} = 0.08$

$R(a_0)$

$= \lambda_{00} P(A=0) + \lambda_{01} P(A=1)$

$= 0.1$

$R(a_1)$

$= \lambda_{10} P(A=0) + \lambda_{11} P(A=1)$

$= 0.1 \times (1-0.1)$

$= 0.09$

$R(a_2)$

$= \lambda_2$

$= 0.08$

$R(a_2) = 0.08 < R(a_1) = 0.09 < R(a_0) = 0.1$

Take the reject action.

2c $P(C=High \mid A=Pos, B=Neg)$

$= \dfrac{P(A=Pos, B=Neg \mid C=High)\,P(C=High)}{P(A=Pos, B=Neg)}$

$= \dfrac{P(A=Pos, B=Neg \mid C=High)\,P(C=High)}{P(A=Pos, B=Neg, C=Low) + P(A=Pos, B=Neg, C=High)}$

$= \dfrac{P(A=Pos, B=Neg \mid C=High)\,P(C=High)}{P(A=Pos, B=Neg \mid C=Low)\,P(C=Low) + P(A=Pos, B=Neg \mid C=High)\,P(C=High)}$

$= \dfrac{P(A=Pos \mid C=High)\,P(B=Neg \mid C=High)\,P(C=High)}{P(A=Pos \mid C=Low)\,P(B=Neg \mid C=Low)\,P(C=Low) + P(A=Pos \mid C=High)\,P(B=Neg \mid C=High)\,P(C=High)}$

$= \dfrac{0.7 \times (1-0.9) \times P(C=High)}{0.1 \times (1-0.1) \times P(C=Low) + 0.7 \times (1-0.9) \times P(C=High)}$

$= \dfrac{0.07P(C=High)}{0.09\,P(C=Low) + 0.07\,P(C=High)}$

$P(C=High)$

$= P(C=High, G=Yes, D=Healthy) + P(C=High, G=Yes, D=Unhealthy)$

$+ P(C=High, G=No, D=Healthy) + P(C=High, G=No, D=Unhealthy)$

$= P(C=High \mid G=Yes, D=Healthy)\,P(G=Yes, D=Healthy)$

$+ P(C=High \mid G=Yes, D=Unhealthy)\,P(G=Yes, D=Unhealthy)$

$+ P(C=High \mid G=No, D=Healthy)\,P(G=No, D=Healthy)$

$+ P(C=High \mid G=No, D=Unhealthy)\,P(G=No, D=Unhealthy)$

$= P(C=High \mid G=Yes, D=Healthy)\,P(G=Yes)\,P(D=Healthy)$

$+ P(C=High \mid G=Yes, D=Unhealthy)\,P(G=Yes)\,P(D=Unhealthy)$

$+ P(C=High \mid G=No, D=Healthy)\,P(G=No)\,P(D=Healthy)$

$+ P(C=High \mid G=No, D=Unhealthy)\,P(G=No)\,P(D=Unhealthy)$

$= 0.8\times0.1\times0.2 + 0.9\times0.1\times(1-0.2) + 0.1\times(1-0.1)\times0.2 + 0.5\times(1-0.1)\times(1-0.2)$

$= 0.466$

$P(C=High \mid A=Pos, B=Neg)$

$$= \frac{0.07\times0.466}{0.09\times(1-0.466)+0.07\times0.466}$$

$= 0.404$

3a

Car Value	Parent
Low	3
High	5

Split on Engine:

Car Value	Engine = Basic	Engine = Good
Low	2	1
High	2	3

Split on Years_Used:

Car Value	Years_Used < 5	Years_Used > 5
Low	0	3
High	3	2

Entropy at node $t = -\sum_{y_k} P(Y=y_k \mid t)\log_2 P(Y=y_k \mid t)$

$0 \log_2 0 = 0$

Entropy (Parent)

$= -\left[\frac{3}{8}\log_2\frac{3}{8} + \frac{5}{8}\log_2\frac{5}{8}\right]$

$= 0.9544$

Entropy (Engine = Basic)

$= -\left[\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right]$

$= 1$

Entropy (Engine = Good)

$= -\left[\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right]$

$= 0.8113$

Entropy (Engine)

$= \frac{4}{8} \times$ Entropy (Engine = Basic) $+ \frac{4}{8} \times$ Entropy (Engine = Good)

$= \frac{4}{8} \times 1 + \frac{4}{8} \times 0.8113$

$= 0.9057$

Entropy (Years_Used < 5)

$= -\left[\frac{0}{3}\log_2\frac{0}{3} + \frac{3}{3}\log_2\frac{3}{3}\right]$

$= 0$

Entropy (Years_Used > 5)

$= -\left[\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right]$

$= 0.9710$

Entropy (Years_Used)

$= \frac{3}{8} \times$ Entropy (Years_Used < 5) $+ \frac{5}{8} \times$ Entropy (Years_Used > 5)

$= \frac{3}{8} \times 0 + \frac{5}{8} \times 0.9710$

$= 0.6069$

FINISH STRONG!

Info Gain (Engine)

= Entropy (Parent) - Entropy (Engine)

= 0.9544 - 0.9057

= 0.0487

Info Gain (Years_Used)

= Entropy (Parent) - Entropy (Years_Used)

= 0.9544 - 0.6069

= 0.3475

Choose to split on Years_Used.

3b Another approach is multi-way split by discretization.

3c $y = \text{sign}\left(\sum_{i=1}^{d} w_i X_i - \theta\right)$

$\lambda = 0.1$, $\theta = 0.5$

$w_1 = 1$, $w_2 = 0.5$, $w_3 = -0.5$

$w(t+1) = w(t) + \lambda (y_i - h_i) \bar{x}_i$

Apply P1 ($X_1 = 1$, $X_2 = -1$, $X_3 = 1$, $y = 1$):

$h = \text{sign}(1 \times 1 + 0.5 \times (-1) + (-0.5) \times 1 - 0.5) = -1$

Prediction is incorrect, perform backward pass.

$w_1 = 1 + 0.1(1-(-1))1 = 1.2$

$w_2 = 0.5 + 0.1(1-(-1))(-1) = 0.3$

$w_3 = -0.5 + 0.1(1-(-1))1 = -0.3$

Apply P2 ($X_1 = -1$, $X_2 = 1$, $X_3 = 1$, $y = -1$):

$h = \text{sign}(1.2 \times (-1) + 0.3 \times 1 - 0.3 \times 1) = -1$

Prediction is correct, no backward pass.

3d Training error rate = $\frac{1}{2} = 0.5$

Prediction for P1 = -1

Prediction for P2 = -1

4a Data matrix $\bar{X}$

$$= \begin{bmatrix} -8 & -3 \\ -9 & 1 \\ -3 & -1 \\ 2 & 7 \\ 5 & -6 \\ 13 & 2 \end{bmatrix}$$

Mean vector $\hat{\mu}$

$$= \begin{bmatrix} \frac{-8-9-3+2+5+13}{6} & \frac{-3+1-1+7-6+2}{6} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Data matrix $\bar{X}$ is centered.

Centered matrix $\tilde{X} = \bar{X}$

Covariance matrix $\tilde{\Sigma}$

$$= \frac{1}{n-1} \tilde{X}^T \tilde{X}$$

$$= \frac{1}{5} \begin{bmatrix} -8 & -9 & -3 & 2 & 5 & 13 \\ -3 & 1 & -1 & 7 & -6 & 2 \end{bmatrix} \begin{bmatrix} -8 & -3 \\ -9 & 1 \\ -3 & -1 \\ 2 & 7 \\ 5 & -6 \\ 13 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 70.4 & 5.6 \\ 5.6 & 20 \end{bmatrix}$$

4b To find 1-dimensional representation, select top principal component (eigenvector) to construct projection matrix:

Projection matrix
$= \begin{bmatrix} -0.99 \\ -0.11 \end{bmatrix}$

1-dimensional representation of P1
$= \begin{bmatrix} -8 & -3 \end{bmatrix} \begin{bmatrix} -0.99 \\ -0.11 \end{bmatrix}$

$= [8.25]$

4c Let P1, P3 and P6 be initially assigned to cluster A.

Let P2, P4 and P5 be initially assigned to cluster B.

Centroid of cluster A $= \left(\frac{-8-3+13}{3}, \frac{-3-1+2}{3}\right) = (0.6667, -0.6667)$

Centroid of cluster B $= \left(\frac{-9+2+5}{3}, \frac{1+7-6}{3}\right) = (-0.6667, 0.6667)$

Iteration 1:

Data point	Euclidean distance to centroid of cluster A	Euclidean distance to centroid of cluster B	Cluster assigned to
P1	8.9753	8.1989	B
P2	9.8093	8.3400	B
P3	3.6818	2.8674	B
P4	7.7817	6.8718	B
P5	6.8718	8.7496	A
P6	12.6183	13.7316	A

Centroid of cluster A $= \left(\frac{5+13}{2}, \frac{-6+2}{2}\right) = (9, -2)$

Centroid of cluster B $= \left(\frac{-8-9-3+2}{4}, \frac{-3+1-1+7}{4}\right) = (-4.5, 1)$

Iteration 2:

Data point	Euclidean distance to centroid of cluster A	Euclidean distance to centroid of cluster B	Cluster assigned to
P1	17.0294	5.3151	B
P2	18.2483	4.5	B
P3	12.0416	2.5	B
P4	11.4018	8.8460	B
P5	5.6569	11.8004	A
P6	5.6569	17.5285	A

Centroids do not change.

Final clustering result:

P5 and P6 are assigned to cluster A. P1, P2, P3 and P4 are assigned to cluster B.

Final centroids:

Centroid of cluster A = (9, -2). Centroid of cluster B = (-4.5, 1).

4d Non-parametric density estimation approaches do not assume any forms for the underlying density. They assume that similar inputs have similar outputs (use local information).

Parametric density estimation approaches assume a form for the probability density function defined for a set of parameters to be found.