ISTANBUL TECHNICAL UNIVERSITY

COMPUTER ENGINERING DEPARTMENT

BLG 527E MACHINE LEARNING

CRN: 13817

Instructor: Zehra Çataltepe

Homework #5

January 10, 2018

Tuğrul Yatağan

504161551

STUDENT ASSESSMENT

Q 1	**Q 2**	**Q 3**	**Q 4**	**Q 5**	**Q 6**	**Q 7**	**Q 8**	**Q 9**	**Q 10**	**Q 11**	**Q 12**	**Q 13**	**Q 14**	**Q 15**	**Q 16**	**Q 17**	**Q 18**	**Q 19**	**Q 20**
0	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	1	0	1	0

Answers

Q1)

Q2)

$X1 \to X2 \to X3 \to X4$

$P(X1) = 0.2$

$P(Xi|Xi-1) = 0.8$

$P(Xi|\sim Xi-1) = 0.1$

$P(X3|X1,\sim X4) = ?$

$P(X2|X1) = 0.8$

$P(X3|X2) = 0.8$

$P(X3|X3) = 0.8$

$P(X2|\sim X1) = 0.1$

$P(X3|\sim X2) = 0.1$

$P(X4|\sim X3) = 0.1$

$P(X3|X1,\sim X4) = P(X3|\sim X4) = P(\sim X4|X3)\ P(X3) / P(X4)$

Q3)

a)

$p_i = p_i^1 - p_i^2$

m = $\frac{\sum_{i=1}^{10} p_i}{10}$ = (0 – 0.07 + 0 + 0 + 0 + 0 + 0.06 + 0 + 0 + 0)/10 = -0.001

S^2 = $\frac{\sum_{i=1}^{10} (p_i - m)^2}{10-1}$ = 0.0005297

$$\frac{\sqrt{9}m}{S} = 0.1373994560916$$

Q4)

$P(D|B) = P(A,C,D|B) + P(\sim A,C,D|B) + P(A,\sim C,D|B) + P(\sim A,\sim C,D|B)$

$P(A,C,D|B) = P(A,B,C,D)/P(B) = P(A)P(B)P(C|A,B)P(D|A,C)/P(B) = P(A)P(C|A,B)\,P(D|A,C)$

$P(\sim A,C,D|B) = P(\sim A,B,C,D)/P(B) = P(\sim A)P(B)P(C|\sim A,B)P(D|\sim A,C)/P(B) = P(\sim A)\,P(C|\sim A,B)\,P(D|\sim A,C)$

$P(A,\sim C,D|B) = P(A,B,\sim C,D)/P(B) = P(A)P(B)P(\sim C|A,B)P(D|A,\sim C)/P(B) = P(A)P(\sim C|A,B)\,P(D|A,\sim C)$

$P(\sim A,\sim C,D|B) = P(\sim A,B,\sim C,D)/P(B) = P(\sim A)P(B)P(\sim C|\sim A,B)P(D|\sim A,\sim C)/P(B) = P(\sim A)\,P(\sim C|\sim A,B)P(D|\sim A,\sim C)$

$P(D|B) = P(A)P(C|A,B)\,P(D|A,C) + P(\sim A)\,P(C|\sim A,B)\,P(D|\sim A,C) + P(A)P(\sim C|A,B)\,P(D|A,\sim C) + P(\sim A)$
$P(\sim C|\sim A,B)P(D|\sim A,\sim C)$

$P(D|B) = 0.1*0.1*0.5 + 0.9*0.2*0.1+0.1*0.9*0.2+0.9*0.8*0.1 = 0.113$

Q5)

$P(S=1) = 0.01$

$P(G=0|S=1) = 0.07$

$P(G=1|S=1) = 0.93$

$P(G=1|S=0) = 0.09$

$P(G=0|S=0) = 0.91$

$P(G=1) = P(G=1|S=1)P(S=1) + P(G=1|S=0)P(S=0) = 0.93*0.01 + 0.09*0.99 = 0.0984$

$P(S=1|G=1) = 0.93*0.01/0.0984 = 0.09451$

$P(S=0|G=1) = 0.90549$

$R(Apply|G=1) = 10*P(S=1|G=1) + 60*P(S=0|G=1) = 10*0.09451 + 60*0.90549 = 55.2745$

$R(Do\ nothing|G=1) = 100*P(S=1|G=1) + 0*P(S=0|G=1) = 100*0.09451 + 0*0.90549 = 9.451$

We shouldn’t apply the therapy.

Q6)

Q7)

$P(X(t)=H) = 2/5$

$P(X(t)=T) = 3/5$

$P(X(t)=H|X(t-1)=H) = 2/5$

$P(X(t)=T|X(t-1)=H) = 3/5$

$P(X(t)=H|X(t-1)=T) = 2/5$

$P(X(t)=T|X(t-1)=T) = 3/5$

$X(t)$ and $X(t-1)$ are independent since $P(X(t)|X(t-1)) = P(X(t))$

Q8)

a)

PCA projects inputs into a lower dimensional space. It is a dimension reduction technique. Backward feature selection is a feature selection method.

PCA tries to project the inputs according to the directions in which there is maximum variance. Backward feature selection chooses the feature whose validation error is the minimum as the feature to leave out

b)

mRMR is much faster than forward-backward selection methods

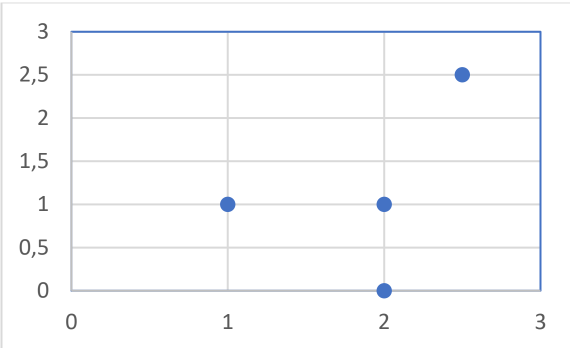
c)

First principal component for PCA is always along the maximum variance for the whole data. PCA do not use the labels of the data.

LDA line tries to maximize the distance between the projected means of the two classes while minimizing the sum within class variances

Q9)

a)



Distance matrix:

	x1	**x2**	**x3**	**x4**
x1	0	1	3	3
x2	1	0	1	1
x3	3	1	0	1
x4	3	1	1	0

G1 = {x1,x2}

G2 = {x3,x4}

G3 = {G1, G2}

b-c)

kmeans has complexity O(n)*(no of iterations) whereas hierarchical clustering has complexity $O(n^2)$. Hence k-means is faster. Both methods are trying to find the optimum selection of cluster membership among a set of k^N possibilities. k-means clustering assumes spherical shaped, non overlapping clusters.

Q10)

The expectation-maximization algorithm is used in maximum likelihood estimation where the problem involves two sets of random variables of which one, X, is observable and the other, Z, is hidden. The goal of the algorithm is to find the parameter vector Φ that maximizes the likelihood of the observed values of X, $L(\Phi|X)$.

Q11)

Q12)

Q13)

Q14)

Q15)

Q16)

Q17)

a)

Bagging assigns constant 1/N for probability of selection of each training instance, Adaboost changes the probability of selection of each instance based on how much error is made on that instance.

Weight of each classifier output is constant ın bagging where as it depends on classifier performance in Adaboost.

Q18)

Q19)

a)

multi-layer perceptron (MLP) to be a subset of deep neural networks (DNN). MLP is subset of DNN. While DNN can have loops and MLP are always feed-forward

b)

CNN is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. CNNs use a variation of multilayer perceptions designed to require minimal preprocessing

Q20)

a)

b)

c)

d)

e)

f)