EndSem Examination
Statistical Machine Learning [CSE 342/ECE 356]

Section A [4 marks each]

Q1) There are two boys and two girls standing at different locations as described below:

Boy1: (2,4)
Boy2: (4,3)
Girl1: (7,9)
Girl2: (8,7)

A camera is located at (16,-5). The cameraman wants to take the photo such that, in the photo, the distance between the two genders is maximized and the distance between the persons of the same gender is minimized. Give the optimal direction (as unit vector) towards which the camera should face. Use an appropriate dimensionality reduction algorithm to solve this.

Solution:
We can solve this problem using FDA, which tries to achieve the same goal on a projection (a photo in this case).
Upon applying FDA, we will get the direction of the projection as [0.7071 0.7071]. Since the projection is perpendicular to what the camera faces and the camera is located at (16,-5), the direction to which the camera should face is [-0.7071 0.7071].

It's fine if one arrives at the same answer using LDA.

3 marks for finding the direction of projection.
1 mark for finding the direction of the camera.

Q2) For a certain dataset with binary class labels, a summary has been given below, i.e. the centroid and the size of the classes:

	Class-0	Class-1
Centroid	(3,5)	(9,7)
Number of samples	4	12

After applying LDA on the data, the centroids become 8 and 16, respectively. Where can the global centroid be located in the original space given that it passes through the line joining the two centroids?

Solution:
Let the optimal vector on which the projections are computed be $[x\ y]^T$.
Thus, the projection value of the first centroid becomes 3x+5y, which has been already given as 8. Similarly, the projection value of the second centroid becomes 9x+7y, which has been already given as 16.

So,
3x+5y=8
9x+7y=16

It's clear that x=1 and y=1. Thus the optimal vector is $[1\ 1]^T$.

We know that the projection of global centroid is the weighted mean of the projections of the centroids, so its value is 0.25*8+0.75*16=14.

Let's assume the global centroid to be (u,v).
Thus, $[1\ 1]x\ [u\ v]^T=14$
It gives us u+v=14.

The line passing through the two centroids is 3v=u+12.
Solving the two equations, we get u=7.5 and v=6.5. Thus the location of the global centroid in (7.5,6.5).

1.5 marks for computing the direction vector of the projection.
1 mark for computing the projection of the global centroid.
1.5 marks for computing the global centroid

Give full marks if one simply takes the weighted average of the two centroids in the original space.

Q3) MSE loss observed for a Linear Regression Model, i.e. $\widehat{y_i}=B_0+B_1x_i$, turns out to be 5, and MSE loss for the base model $\widehat{y_i}=\overline{y}$ on the same training data happens to be 10. Given that there were a total of 50 training samples, compute R2 and RMSE scores of the two models on the training data.

Soln:

For LR model,

RSS=NxMSE$_{LR}$=250

TSS=NxMSE$_{Base}$=500

Thus R2=1-250/500=0.5 **[1 Mark]**

RMSE=$\sqrt{MSE}$=$\sqrt{5}$ **[1 Mark]**

For base Model,

RSS=TSS=NxMSE$_{Base}$=500

Thus R2=0 **[1 Mark]**

RMSE=$\sqrt{10}$ **[1 Mark]**

Section B [2 marks each]

Q4) Prove that a logistic regression model can also be seen as a linear regression model that predicts the log of odds.

Soln:

If $z=B_0+B_1x$, the logistic regression model can be represented as

$p=\sigma(z)$.

$p/(1-p)=\sigma(z)/(1-\sigma(z))=\exp(z)$

$\log(p/(1-p))=z$

I.e. $\log(p/(1-p))=B_0+B_1x$

Full proof: 2 Marks
Incomplete: 1 Mark

Give 0.5 marks if someone has attempted and wrote something sensible related to linear/logistic regression.

Q5) While building a decision tree, it so happened that all five features given below provided the same information gain, but there was a clear winner in terms of the gain ratio. Which feature must have been that?

F1	F2	F3	F4	F5
1	1	0	1	0
0	1	1	1	0
2	0	0	2	1
0	1	0	0	1
1	1	2	2	1
2	2	1	0	2

Solution:

The feature having the lowest entropy will be the required feature.

E_{F1}=Entropy(2,2,2)=1.0986

E_{F2}=Entropy(1,4,1)=0.8676

E_{F3}=Entropy(3,2,1)=1.0114

E_{F4}=Entropy(2,2,2)=1.0986

E_{F5}=Entropy(2,3,1)=1.0114

Thus, F2 is the required feature.

If the student computes intrinsic information correctly, give 0.25 marks for each feature.
[0.25x5=1.25 marks]
The answers may change based on the base of the logarithm chosen.

If the student arrives at F2, give another 0.75 mark

Q6) What assumptions of Naive Bayes' algorithm make it so naive? Do these assumptions matter for the dataset below? What will be the prediction of a Naive Bayes' model for a Sunny day?

Weather	Play
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Sunny	No
Rainy	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Sunny	No
Sunny	Yes
Rainy	No
Overcast	Yes
Overcast	Yes

Solution:

Assumptions **[0.5 Marks]**:

1) The features are independent of each other
2) All features contribute equally in making the prediction

No, since there is only one feature **[0.5 Marks]**

P(Yes|Sunny)=3/14 **[0.25 Marks]**
P(No|Sunny)=2/14 **[0.25 Marks]**
Prediction for Sunny=Yes **[0.5 Marks]**

Q7) Which one do you think is more parallelizable, Boosting or Bagging? Why? Provide an example of an ensemble learning algorithm for each approach.

Solution:

Bagging is more parallelizable because the models in bagging are developed independent of each other, whereas in boosting the models are sequentially developed, depending on the previous one.

[0.5+0.5=1 Marks]

Bagging Example: Random Forest **[0.5 Marks]**
Boosting Example: AdaBoost/Gradient Boosting **[0.5 Marks]**