CS550: Machine Learning Tierce Examination1

Name:

Roll Number:

Instructions

a. Read all questions carefully before you start to answer them. In case of any doubt, you can write your assumptions before answering the question.
b. This is a CLOSED BOOK exam.
c. You can use a calculator.
d. Use of Phone, INTERNET or Google Searches etc. is NOT ALLOWED.
e. Discussion with any other person is NOT ALLOWED. Any unfair means used will result in D or F grade and will be reported to the competent authority.
f. You will get 90 minutes to solve all the questions. We won't be giving extra time. Please submit whatever you have done at the end of the allotted time. Anyone found writing after completion time will get -20 marks.
g. PLEASE ALWAYS PROVIDE A JUSTIFICATION FOR YOUR ANSWERS. CORRECT SOLUTIONS WITHOUT ANY JUSTIFICATION WILL NOT GET FULL CREDITS. THEY MAY ALSO BE CONSIDERED AS POTENTIAL UNFAIR MEANS.
h. Make sure to write your name and roll number on the top of your sheet(s) when submitting.
i. Please write the answers in the EXAM PAPER itself. The justifications and calculations can be in the answer sheet. We will look at answer sheet only if needed.

Q	Topic	Max. Marks	Marks Obtained		
1	ML General Concepts	20			
2	Regression	25			
3	Classification	20			
4	Metrics + Ensemble	25			
Total		90			

Q1. ML General Concepts (2 marks each)

A. Using regularization will …….. training accuracy but is expected to …….. test accuracy.

B. We can estimate the test accuracy of the model using the process of ……..

C. A simpler model has …….. bias but …….. variance and does not learn the …… in the data.

D. An ensemble of weak models (low accuracy) will become strong (high accuracy) provided that the models are ……..

E. In a classification problem, the examples that are most difficult to classify lie …….. the decision boundary.

F. The mean square loss function used in regression analysis is …….. and therefore, gradient descent will …….. lead to the optimal solution.

G. In a binary classification problem, hard margin Support Vector Machines aim to …….. the margin between the two classes.

For a soft margin SVM, in their objective function, the term $w^T w$ is for …….. and CZ_i is for ……..

H. In short, when will you use the following to solve a regression problem:

a. Matrix Methods (Normal Form, PseudoInverse):

b. Stochastic Gradient Descent:

c. Gradient Descent:

I. Which features should be selected for modeling? (Select all that apply)

a. Those which are highly correlated to each other
b. Those which are not correlated to each other
c. Those which are correlated with the output variable
d. Those which have a large non-zero coefficient in a model

Q2. Predicting salaries of Google Engineers

The salaries of Google Engineers working in CA was scraped from social media and the following table was prepared.

Role	# of Years Experience (Y)	Salary ($K/year)
SDE	3.5	128
SDE	4.5	133
SDE	1	97
SDE	2.5	118
SDE	10	164
SDE	6	145
Data Scientist	2	118
Data Scientist	3	142
Data Scientist	1	100
Data Scientist	2.5	130

1. **(15 marks)** Please use matrix methods (or any other technique you like) to determine the regression line for predicting the salary of Data Scientists working at Google.

   Salary of Data Scientist = ……. + Y* …….

   Now fill the table below using the above equation

Role	Y	Salary ($K/year)	Comments
Data Scientist	3		
Data Scientist	7		
Data Scientist	0		

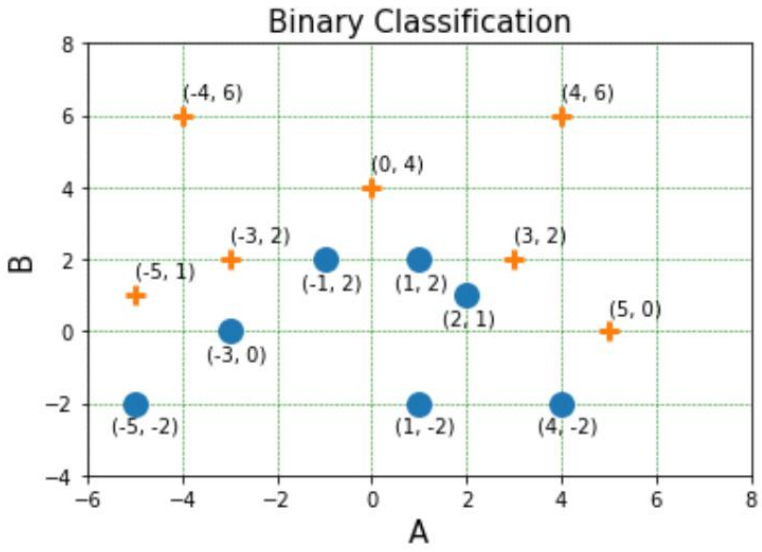
2. **(6 marks)** Using 3 nearest neighbors, predict the salary of following employees. You are free to choose the weights you want to give to the neighbors.

Role	Y	Salary ($K/year)	Comments
SDE	3		
SDE	7		
SDE	0		

   How did you assign weights to the neighbors?

3. **(4 marks)** Examining the predictions above, who is more highly paid at Google, SDE or Data Scientists? Can you say this confidently for experienced engineers?

Q3. [20 marks] A dataset for predicting low quality chips was collected and cleaned. We have two features, A and B which can be used for binary classification into +ve (shown with +) and -ve class (indicated by o) in the plot below.



a. [2 marks] Can we separate this data with a linear classifier in original features (A, B)? Why or why not?

b. [3 marks] Is it possible to separate this data with a classifier that is polynomial in A and B? What is the degree of the polynomial that should suffice? Why or why not?

c. [15 marks] Please write the equation of a classifier that separates this data perfectly?
Hint: You can use polynomial or piece-wise linear function. Please make sure that all points are correctly classified.

Q4. (25 marks) Glaucoma is a serious disease that is the leading cause of irreparable blindness. An AI Fundus Camera has been developed to screen patients at large for Glaucoma, based on the image of the Retina. The positive patients will be referred to expert eye doctors for further assessment.

3 Models, were trained for this purpose and it turned out that each classifier had its own strengths and weaknesses. They sometimes agreed with each other on their predictions and sometimes differed.

We are given the confusion matrices for each of the models on the Validation set.

Model (M1)	Actual	
Predicted	Yes	No
Yes	9	5
No	1	105

Model (M2)	Actual	
Predicted	Yes	No
Yes	8	2
No	2	108

Model (M3)	Actual	
Predicted	Yes	No
Yes	10	8
No	0	102

Model	**Precision**	**Recall**	**F1 Score**
M1			
M2			
M3			

Which of these models is best suited for Glaucoma Screening? Why?

Ensemble Model:

The table below tells you the probability of Glaucoma in the eye based on its image for each model. Combine the predictions from these models using hard, soft and weighted voting strategies. Which metric would you use for determining weights?

Test Example	M1	M2	M3	Hard Voting	Soft Voting	Weighted Voting
Right Eye	0.4	0.4	0.8			
Left Eye	0.3	0.6	0.7			

Which of these strategies you feel would work the best?