Mid-term Examination Solution
2024
CSN 382
IIT Roorkee

Date: 24th Feb

Write your Roll Number and Name here -

Note:

1. Do not write anything on the question paper except your name and roll number.
2. Any case of copying will be treated strictly as per the institute rules.

Que1. [4 Marks]

You have the following data points with one feature and corresponding y variable.
(x,y)=[(2,5),(4,10),(6,15),(8,20)]

Derive the value of the optimal Hypothesis function using algebra (Gradient Descent is not required).

To derive the hypothesis function for linear regression, we'll use the formula for the slope (θ1) and the y-intercept (θ0):

Therefore, the hypothesis function for the given data points is:

$h\theta(x) = 0 + 2.5 \cdot x$

$\theta 0 = 0$

$\theta 1 = 2.5$

a. What are the following scenarios for θ (weight parameters) in the gradient ascent algorithm?
   1. 'Make progress with each example' - *Stochastic Gradient Ascent*
   2. 'make progress with each iteration' - *Batch gradient ascent.*

b. Show formulation for both the above scenarios using any ML algorithm's setup which you are familiar with. - Logistic regression formulation with batch and stochastic update in theta.

Que-3. [2+2+2 Marks]

a. Give an example of concave shape cost function.
   log(x) or any other example

b. Justify why logistic regression adapts maximum likelihood estimation of thetas.

Because square error may end up in local minima, log likelihood ensures global minima and it brings the output in the range of 0 to 1.

c. **Let's assume a scenario of Linear Regression. Do we have the same error term (hθ(x)-y) and learning rate α for update in all thetas in an iteration of gradient descent? If yes, what is the deciding factor in the formulation which differentiates all thetas from each other? If no, write 'Not valid statement'.**

- xj term in the gradient descent formulation makes the difference in the updation of thetas.

Que-4. [2+2 Marks]

a. **Provide the value of the functions given below.**
   i. **1{2+2=4} evaluates to what?** - *1*
   ii. **1{2+2=5} evaluates to what?** - *0*

b. **Define LASSO and Ridge Regularization. Justify which one is more suitable if we want to retain selected features only.**

LASSO is L1 regularization - have absolute terms of theta - more suitable to suppress the features.

Ridge is L2 regularization - square terms of theta

Que-5. [10 Marks]

There are three different types of language speakers in a group, 98 English, 89 Spanish and 40 French. Your model classified English speakers, 86 as English, 10 as Spanish and 2 as French. Similarly, it classified Spanish speakers, 5 English, 79 Spanish, and 5 French. French speakers are classified as, 10 English, 5 Spanish, and 25 French.

a. **Make a confusion matrix for the classification model.**
b. **Demonstrate which class values pairs where the model is most confused.**
c. **Compute Precision, Recall and F1-score of the model.**
d. **Compute Accuracy of the model.**
e. **Which metric is preferable to evaluate the model and why?**

	English	Spanish	French	Total
Predicted | English	86	10	2	98
Spanish	5	79	5	89
French	10	5	25	40
Total	101	94	32	227

Now, let's compute Precision, Recall, and F1-score for each class:

- Precision for English (P_English): $\frac{86}{86+5+10}$
- Recall for English (R_English): $\frac{86}{86+10+2}$
- F1-score for English (F1_English): $\frac{2 \cdot P_English \cdot R_English}{P_English + R_English}$

Similarly, compute Precision, Recall, and F1-score for Spanish and French.

- Precision for Spanish (P_Spanish): $\frac{79}{10+79+5}$
- Recall for Spanish (R_Spanish): $\frac{79}{5+79+5}$
- F1-score for Spanish (F1_Spanish): $\frac{2 \cdot P_Spanish \cdot R_Spanish}{P_Spanish + R_Spanish}$
- Precision for French (P_French): $\frac{25}{2+5+25}$
- Recall for French (R_French): $\frac{25}{10+5+25}$
- F1-score for French (F1_French): $\frac{2 \cdot P_French \cdot R_French}{P_French + R_French}$

Next, compute the Accuracy of the model:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Finally, regarding which metric is preferable to evaluate the model, it depends on the specific goals of the application. If the classes are imbalanced, you might prefer to look at metrics like F1-score, which balances precision and recall. Accuracy can be misleading in imbalanced datasets.