Week 2 Quiz Answers

Why does Logistic Regression Algorithm has the term "Regression" in its name?   1 point

○ It is used for the task of Regression

○ It was a mistake, done by early ML Practitioners, and is now a well accepted terminology

◉ It's underlying technique is quite the same as Linear Regression. and it takes the regression output to pass as an input to the logistic function (Sigmoid).

Clear selection

Consider the following statements-S1: Logistic regression mainly used for Regression S2: It is possible to apply a logistic regression algorithm on a 3-class classification problem   1 point

○ Both S1 and S2 are true

○ Both S1 and S2 are false

○ S1 is true and S2 is false

◉ S1 is false and S2 is true

Clear selection

Why is linear regression not suitable for Classification tasks?   1 point

○ It can only predict Continous Numerical Values

○ It cannot capture Complex Decision Boundaries

○ It does not provide probability estimates.

◉ All of the Above

Clear selection

Which of the following is True about Optimization Approach of Normal Equation in Linear Regression?   2 points

- ☐ It is Slow, when the Number of Features is Large.
- ☐ It can be applied to other algorithms too
- ☐ It requires no Iterations, in contrast to Gradient Descent
- ☑ None of the Above

Below are the three scatter plot(A,B,C left to right) and hand drawn decision boundaries for logistic regression.   2 points

A   B   C
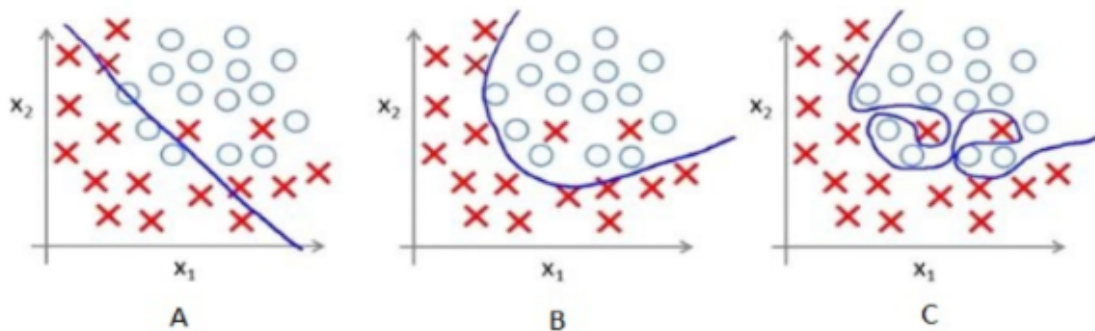
- ☐ The best model for this regression problem is the last (third) plot because it has minimum training error (zero).
- ☑ The second model is more robust than first and third because it will perform best on unseen data.
- ☑ The third model is overfitting more as compare to first and second.
- ☐ All will perform same because we have not seen the testing data.

How will the bias change on using High (Infinite) Regularisation   1 point

- ⦿ Bias will be high
- ○ Bias will be low
- ○ Can't say
- ○ None of these

Clear selection

For Ridge Regression, if the regularization parameter = 0, what does it mean? 1 point

○ Large coefficients are not penalized

◉ Overfitting problems are not accounted for

○ The loss function is as same as the ordinary least square loss function

○ All of these

Clear selection

What are the limitations of Lasso Regression? 2 points

☑ If the number of features (p) is greater than the number of observations (n), Lasso will pick at most n features as non-zero, even if all features are relevant

☑ If there are two or more highly collinear feature variables, then LASSO regression selects one of them randomly which is not good for the interpretation of the data

☐ Lasso can be used to select important features of a dataset

☐ The difference between ridge and lasso regression is that lasso tends to make coefficients to absolute zero as compared to Ridge which never sets the value of the coefficient to absolute zero

Which of the following is true about Regularized Linear Regression model? 1 point

◉ Increase in regularization parameter (lambda) will make the model to underfit the data and the validation error will go up.

○ Decrease in regularization parameter (lambda) will make the model to overfit the data and the training error go up

○ Increase in regularization parameter (lambda) will make the model to underfit the data and the training error go down

○ All of the above are true

Clear selection

Is it true that the L1 term in Lasso has the following purposes: performing feature selection, compensating for overfitting, and smoothing?   1 point

◉ True

◯ False

Clear selection

What does 'Naive' mean in Naive Bayes?   1 point

◯ The Bayes' Theorem makes estimating the probabilities easier. The 'naïve' in the name of classifier comes from this ease of probability calculation.

◉ The model assumes that the input features are statistically independent of one another. The 'naïve' in the name of classifier comes from this naïve assumption.

◯ The full Bayes' Theorem is not used. The 'naive' in naive bayes specifies that a simplified version of Bayes' Theorem is used.

◯ None of these

Clear selection

You are in possession of a bag of words spam classifier trained on a large   2 points
corpus of emails. Below is a table of some estimated word probabilities.
You are given a new email to classify, with only two words: "perfect note".
Select all values of P(Y = spam) which are possible for which the classifier
will give "spam" as the most likely label for the given email.

W	note	to	self	become	perfect
$P(W \mid Y = \text{spam})$	1/6	1/8	1/4	1/4	1/8
$P(W \mid Y = \text{ham})$	1/8	1/3	1/4	1/12	1/12

☑ 0.4

☐ 0.2

☑ 1

☑ 0.8

Which distribution should you use with the Naïve Bayes Algorithm for continuous features?

1 point

○ Multinomial

○ Poisson

○ Bernoulli

◉ Gaussian

Clear selection

In the Naive Bayes Algorithm, suppose that prior for class W1 is greater than class W2, would the decision boundary shift towards region R1 (region for deciding W1) or towards R2 (region for deciding W2)?

1 point

○ towards region R1.

◉ towards region R2.

○ No shift in decision boundary.

○ It depends on the exact value of priors.

Clear selection

Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?

1 point

○ AUC-ROC

○ Accuracy

○ LogLoss

◉ Mean Squared Error

Clear selection

If TP=9 FP=5 FN=26 TN=70 then Error rate will be   1 point

○ 45 percentage

○ 99 percentage

◉ 28 percentage

○ 20 perentage

                                                                Clear selection

Imagine you are working on a project which is a binary classification   2 points
problem. You trained a model on the training dataset and get the below
confusion matrix on the validation dataset

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

☑ Accuracy is ~0.91

☐ Misclassification rate is ~ 0.91

☐ False positive rate is ~0.95

☑ True positive rate is ~0.95

Calculate MAE, MSE, RMSE from the given table.   2 points

Write the answer in the same order as given in the question separated by comma,
without any spaces in between
(For example :- 200,100,100 implies that MAE=200, MSE=100 and RMSE=100)

Actual Value	Predicted Value
10	14
21	19
34	43
108	97
10000	10400

85.2,32044.4,179.09

What is Log-Loss?   2 points

- [ ] It is a Loss function, for Regression Problems.
- [x] It is an Evaluation Metric, for Classification Porblems
- [x] It is a Loss function, for Classification Problems.
- [ ] It is an Evaluation Metric, for Regression Porblems

Imagine, you are solving classification problems with a highly imbalanced   2 points
class. The majority class is observed 99% of time in the training data. Your
model has 99% accuracy after taking the predictions on test data. Which of
the following is true in such a case?

- [x] Accuracy metric is not a good idea for imbalanced class problems.
- [ ] Accuracy metric is a good idea for imbalanced class problems.
- [x] Precision and recall metrics are good for imbalanced class problems.
- [ ] Precision and recall metrics aren't good for imbalanced class problems.

Suppose, You applied a Logistic Regression model on a given data and got a 2 points
training accuracy X and testing accuracy Y. Now, you want to add a few new
features in the same data. Select the option(s) which is/are correct in such a
case.

Note: Consider remaining parameters are same.

- ☑ Training accuracy increases

- ☐ Training accuracy increases or remains the same

- ☐ Testing accuracy decreases

- ☑ Testing accuracy increases or remains the same