

Question 17.**Marks : 1.00**

You are in possession of a bag of words spam classifier trained on a large 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 $P(Y = \text{spam})$ is possible for which the bag of words with these word probabilities will give "spam" as the most likely label.

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

Answer☐ 0.2☒ 0.4☒ 1☒ 0.8

Question 16.**Marks : 1.00**

Suppose, You applied a Logistic Regression model on a given data and got a 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)

Answer

- ☒ Testing accuracy increases or remains the same
- ☐ Training accuracy increases or remains the same
- ☐ Testing accuracy decreases
- ☒ Training accuracy increases

Question 15.

Marks : 1.00

How will the bias change on using high(infinite) regularisation

Answer

☐ Bias will be low

☐ None of These

☒ Bias will be high

Question 14.

Marks : 1.00

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

Answer

☒ False

☐ True

Question 13.

Marks : 1.00

What does 'naive' mean in Naive Bayes?

Answer

- ☐ None of these
- ☐ The full Bayes' Theorem is not used. The 'naive' in naive bayes specifies that a simplified version of Bayes' Theorem is used.
- ☒ 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 Bayes' Theorem makes estimating the probabilities easier. The 'naïve' in the name of classifier comes from this ease of probability calculation.

Question 12.

Marks : 1.00

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

Answer

☒ Gaussian

☐ Poisson

☐ Multinomial

☐ Bernoulli

Question 11.

Marks : 1.00

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

Answer

☒ Mean Squared Error

☐ AUC-ROC

☐ Accuracy

☐ LogLoss

Question 10.

Marks : 1.00

Which of the following statements about regularisation is not correct?

Answer

- ☐ Using too large a value of lambda can cause your hypothesis to underfit the data.
- ☐ Using a very large value of lambda cannot hurt the performance of your hypothesis
- ☒ None of these
- ☐ Using too large a value of lambda can cause your hypothesis to overfit the data

Question 9.**Marks : 1.00**

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)?

Answer

- ☐ towards region R1.
- ☐ It depends on the exact value of priors.
- ☒ towards region R2.
- ☐ No shift in decision boundary.

Question 8.

Marks : 1.00

Which of the following is true about regularized linear regression model?

Answer

- ☒ Increase in regularization parameter (λ) will make the model to underfit the data and the validation error will go up.
- ☐ Decrease in regularization parameter (λ) will make the model to overfit the data and the training error go up
- ☐ Increase in regularization parameter (λ) will make the model to underfit the data and the training error go down
- ☐ All of the above are true

Question 7.

Marks : 1.00

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

Answer

☐ 99 percentage

☐ 20 percentage

☐ 45 percentage

☒ 28 percentage

Question 6.

Marks : 1.00

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

Answer

- ☐ Large coefficients are not penalized
- ☐ Overfitting problems are not accounted for
- ☒ All of these
- ☐ The loss function is as same as the ordinary least square loss function

Question 5.

Marks : 1.00

What are the limitations of Lasso Regression?

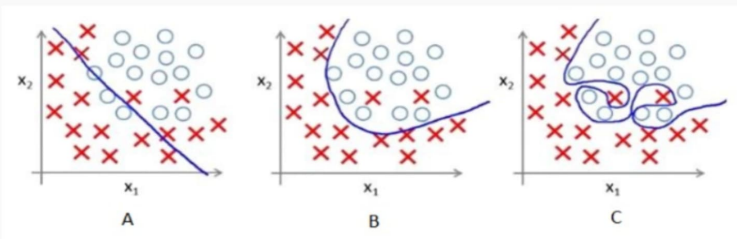
Answer

- ☒ 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
- ☐ None
- ☒ 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

Question 4.

Marks : 1.00

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



Answer

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

Question 3.

Marks : 1.00

Imagine, you are solving classification problems with a highly imbalanced 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?

Answer

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

Question 2.

Marks : 1.00

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

Answer

- ☐ S1 is true and S2 is false
- ☐ Both S1 and S2 are true
- ☐ Both S1 and S2 are false
- ☒ S1 is false and S2 is true

Question 1.

Marks : 1.00

Imagine you are working on a project which is a binary classification 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	

Answer

☐ False positive rate is ~0.95

☐ Misclassification rate is ~ 0.91

☒ Accuracy is ~0.91

☒ True positive rate is ~0.95