MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A) Least Square Error B) Maximum Likelihood
C) Logarithmic Loss D) Both A and B

Answer: (A) Least Square Error

Explanation: To get the best fit line, we take into consideration the Mean square error which is the difference between the predicted and actual data points. With the least Mean Square error, we achieve the best-fit line for the model.

----

2. Which of the following statement is true about outliers in linear regression?
A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers
C) Can't say D) None of these

Answer: (A) Linear regression is sensitive to outliers

Explanation: we reduce the mean square error to get the best fit line but an outlier will increase the MSE therefore to reduce MSE best fit line will deviate.

----

3. A line falls from left to right if a slope is _______?
A) Positive B) Negative
C) Zero D) Undefined

Answer: (A) Positive

Explanation: line falls from left to right will be like **"/ ' . Y= mx+c,** Lets suppose line passing through origin so c=0, ,Y= mx, slope (m)= y2-y1/x2-x1.

----

4. Which of the following will have symmetric relation between dependent variable and independent variable?
A) Regression B) Correlation
C) Both of them D) None of these

Answer: (C) Both Regression and Correlation

Explanation: symmetric relation means vice and versa like a==b so b=a. If Y= dependent variable and x= Independent variable so Y=Mx+C represents the relation here.
Correlation is also a degree of association between independent and dependent variables is measured by correlation coefficient (r) -1 to +1.

----

5. Which of the following is the reason for over fitting condition?
A) High bias and high variance B) Low bias and low variance
C) Low bias and high variance D) none of these

Answer: (C) Low Bias and High Variance

Explanation: Over-fitting means the Model is doing well with the train dataset results in less error (Low Bias) while not doing well with test dataset results in a high error(High Variance)

---

6. If output involves label then that model is called as:
A) Descriptive model B) Predictive modal
C) Reinforcement learning D) All of the above

Answer: (A) Descriptive model

A label provides Description for the output

---

7. Lasso and Ridge regression techniques belong to __________?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization

Answer: (D) Regularization

Explanation: L1 (Lasso) and L2 (Ridge) are regularization techniques to prevent overfitting. It penalizes the model to get a generalized model.

---

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE

Answer: (D) SMOTE: Synthetic Minority over-sample technique

Explanation: Imbalanced dataset in classification means Label has more outputs of 1 class and very less of class 2 (80-20) (class 1 and 2 are the output like Yes, No). In this case, the Model will be given less knowledge to evaluate class 2 predictions. So probably, it may predict class 1 for most cases.

The model should be given enough data to analyze the relation between inputs and outputs. SMOTE will create Synthetic samples for minority class (Class 2 in this case) so the model gets trained enough to predict class 2 results as well.

---

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make a graph?
A) TPR and FPR B) Sensitivity and precision
C) Sensitivity and Specificity D) Recall and precision

Answer: (A) TPR and FPR

Explanation: ROC curve created for True Positive Rate VS False Positive Rate.
TPR= TP / (TP+FN)
FPR= FP / (TP +TN)
----

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True B) False

Answer: (B) False

Explanation: IN AUCROC, More Area under Curve means the best model. It is also useful to compare the different model's performance. Which model has a high AUC will be the best one.

----

11. Pick the feature extraction from below:
A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Answer: (B) Apply PCA to project high dimensional data.

Explanation: Feature extraction means reducing the no. of features in the dataset. Principal Component Analysis is used to handle high dimensional data. It reduces the features to (n_components=) In such a way, that a reduced set of features should be able to summarize most of the information contained in the original dataset.

Before applying PCS, we need to scale data first.

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
A) We don't have to choose the learning rate.
B) It becomes slow when number of features is very large.
C) We need to iterate.
D) It does not make use of dependent variable.

Answer: (A) and (B)

Explanation: We don't have to choose the learning rate. It is done by Gradient Descent generally it takes less learning rate so not to skin global minima.

Also, a large features number means more dimensions and more coefficients to be calculated which slows the model.

ASSIGNMENT – 39

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in the linear regression equation?

MACHINE LEARNING

13. Explain the term regularization?

Explanation:
In general, regularization means to make things regular or acceptable.

Regularization is a technique that makes slight modifications to the learning algorithm to make the model generalizes better. This in turn improves the model's performance on the unseen data as well.

Let's understand why Regularization is required:

Under-fitting---------------------best point-------------------Overfitting

Under-fitting means High Bias, the model is not learning enough from the data. Also, giving errors to predict for unseen data, High Variance.

Overfitting means the Model learned data completely (remember), Low Bias but giving errors to predict for unseen data, High Variance. Overall the model is not flexible enough.

Here we try to achieve the best point to generalize the model so that it could do best with Train data as well as with test data.

How it does: In a linear equation Y= C+ m1x1+m2x2+m3x3+…..mnxn

Here, C is intercept and Xn are data points with mn coefficients.

We don't want huge coefficients as a small change in coefficient can make a large difference for the dependent variable Therefore regularization shrink the coefficients of features to avoid overfitting basically by adding a penalty.

14. Which particular algorithms are used for regularization?

Explanation: As we have already explained the Regularization in the previous question. It penalizes the model to reduce the weights of coefficients to get a generalized model or to say to avoid overfitting.

How the model works: To get the best fit line we use the cost function reduced to 0.
Sum of Residuals= $\sum \quad (y - y^\wedge)$**2 = 0 (cost function)**
All data points are on the best fit line so Sum of Residuals =0.

It is overfitting here because when the test data comes it gives a high error (High Variance). Now, we need to regularize the model with the Ridge and Lasso Technique. To reduce High Variance -→ Low Variance.
In Linear Regression, we tend to reduce the cost function only.
<u>L2 Regularization:</u> In this **Ridge Regression**, we will add some penalties long with a cost function. Penalty is equivalent to the square of the magnitude of coefficients.

= $\sum \quad (y - y^\wedge)$**2 +**$\boldsymbol{\lambda}$ **(slope)2**
Will try to minimize the whole equation. $\boldsymbol{\lambda}$ is any positive number.

As the $\sum \quad (y - y^\wedge)$**2** is already =0, we are remaining with $\boldsymbol{\lambda}$ **(slope)2.**

The slope is basically the coefficients of features here. To make this equation to 0. We have to reduce the coefficients as much as possible.

In this process, the features are not contributing much to predicting the label. Ridge will give less importance to those features and their coefficients will shrink down to nearly 0. Along with the coefficients of features which are highly contributing to predicting the label shrink down to some extent.

<u>L1 Regularization:</u> In this **LAsso Regression**, we will add some penalties long with a cost function. The penalty is equivalent to the absolute value of the magnitude of coefficients.

= $\sum \quad (y - y^\wedge)$**2 +**$\boldsymbol{\lambda}$ **|slope|**

As the $\sum \quad (y - y^\wedge)$**2** is already =0, we are remaining with $\boldsymbol{\lambda}$ **|slope|.**

In this process, the features are not contributing much to predicting the label. Lasso will give 0 importance to those features and shrink down the coefficients to 0.

Hyperparameter: $\boldsymbol{\lambda}$ best value can be selected by LassoCV and RidgeCV.

---

15. Explain the term error present in the linear regression equation?

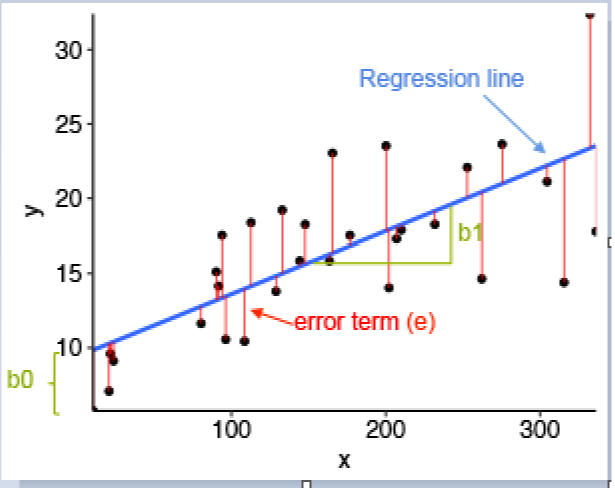
Explanation: Regression analysis is a set of statistical processes for estimating the relationships among features. It may be Simple or Multiple regression.

Let's take simple regression, where it has 1 independent feature and the model has to find the linear relationship of it with the dependent feature.
Equation is $y = c + mx + E$

E= Regression residual error

Linear regression models main aim is to find the best fit linear line with optimal values of intercept (C) and coefficients(m) such that error is minimized.

Error is the difference between the actual value and the predicted value. And the goal is to reduce this difference.



Blueline is the best fit line (Regression Line) predicted by Model
X is our independent variable
Y is our dependent variable

Here, the vertical difference between the data point and the regression line is known as error or residual.

Error= actual value- predicted value
Sum of error= sum(actual – predicted)

Square of sum of error= (sum(actual-predicted values))2

SSE= $\sum \quad (y - y\text{^})2$

Bottom line is to minimize SSE.
