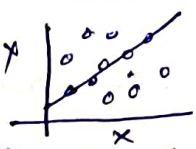
FLIP ROBO

Assignment-4

ASSIGNMENT – 39

MACHINE LEARNING



→ Best fit Line

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

$y = a + bx + e$ (e: error; a: intercept; b: coefficient / slope; x: independent variable; y: dependent)

(2) Which of the following statement is true about outliers in linear regression?

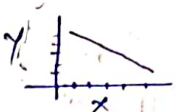
A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can't say D) none of these

(3) A line falls from left to right if a slope (coefficient) is _____?

A) Positive B) Negative

C) Zero D) Undefined



(4) Which of the following will have symmetric relation between dependent variable and independent variable?

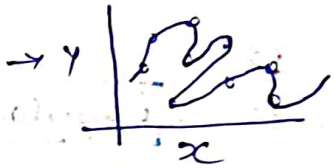
A) Regression B) Correlation

C) Both of them D) None of these

(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



(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

(7) Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

(8) To overcome with imbalance dataset which technique can be used?

A) Cross validation B) Regularization

C) Kernel D) SMOTE

(9) The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

A) TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

(recall / sensitivity) ↑

TPR - True positive rate

FPR - false positive rate ↓

(1 − specificity)

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

A) True B) False

(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

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

$y = a + bx + e$

(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.

Ans-8 SMOTE – synthetic Minority Over Sampling Technique.

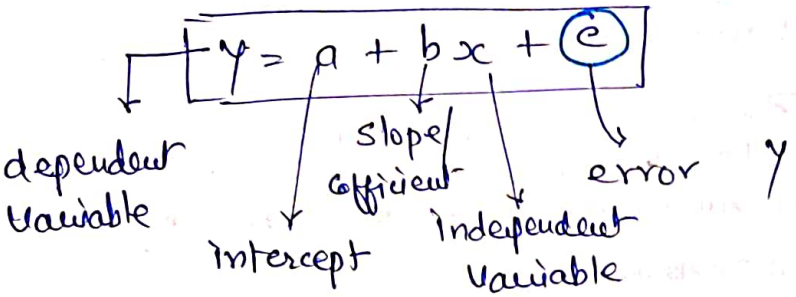
(equally distribute Y values.)

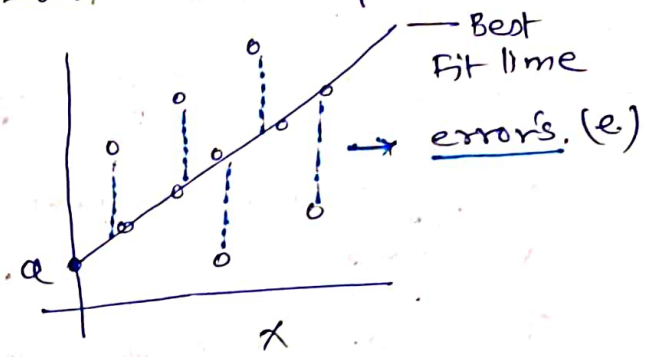
imbalance data set → SMOTE → balanced data set

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 linear regression equation? — Diff. b/w the Actual Answers and the predicted Answer.





Answer-13
Answer 14 → **Regularization** is one of the way to handle Bias_Variance_Trade off

```
from sklearn.linear_model import Lasso, Ridge, ElasticNet
```

Here in Regularization we can use three algorithms :-

Lasso
- Omits Un-necessary columns which are not much contributable

Ridge
- it reduces the coefficient differences and trying to reduce gap between two.

ElasticNet
- Combination of Lasso + Ridge Regression.

Here, alpha is the attribute, whose values could be
(alpha = .0001, .001, .01, .1, 1)

(B) Perceptron
(C) Recurrent Neural network