MACHINE LEARNING

ANS1- a- Least Square Method

ANS2- a- Linear Regression is sensitive to outliers

ANS3- b- Negative

ANS4- b- Correlation

ANS5- c- Low Bias and High Variance

ANS6- b-Predictive modal

ANS7- d- Regularization

ANS8- d- SMOTE

ANS9- c- Sensitivity and Specificity

ANS10- b-false

ANS11- c-Removing stop words

ANS12- a- We don't have to choose the learning rate

b- It becomes slow when the no. of feature is very large

ANS13- REGULARIZATION:

Regularization is a technique used to reduce the errors by fitting the functions appropriately on the given training set and avoid overfitting. This technique converts a complex model into a simpler one, so as to avoid the risk of overfitting and shrinks the coefficients, for lesser computational cost.

The commonly used regularization techniques are –

a- LASSO – Least Absolute Shrinkage and Selection Opening
b- Ridge Regression
c- Elastic net Regression

ANS14- DIFFERENT REGULARIZATION ALGORITHMS :

1- Ridge Regression- It is a method that is used to analyze the data that suffers from multi-collinearity. It is also called L2 regression.

$$Loss = \sum_{i=1}^{n} (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^{n} w_i^2$$

Ridge Regression adds “square magnitude” of coefficient as penalty term to the loss function(L). The regularization parameter (λ) regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

λ → 0, the penalty term has no effect, and the estimates produced by ridge regression will be equal to least-squares i.e. the loss function resembles the loss function of the Linear Regression algorithm. Hence, a lower value of λ will resemble a model close to the Linear regression model.

λ → ∞, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will approach zero (coefficients are close to zero, but not zero).

2-LASSO Regression- It is a regression analysis that performs both feature selection and regularization in order to enhance the prediction ability of the model.

$$Loss = \sum_{i=1}^{n} (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^{n} |w_i|$$

It adds “absolute value of magnitude” of coefficient a penalty term to the loss function. LASSO regression converts coefficients of less important features to zero, which indeed helps in feature selection, and it shrinks the coefficients of remaining features to reduce the model complexity, hence avoiding overfitting. It is also called L1 Regression.

3-Elastic Net Regression- Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

$$Loss = \sum_{i=0}^{n} (y_i - (w_i x_i + c))^2 + \lambda_1 \sum_{i=0}^{n} |w_i| + \lambda_2 \sum_{i=0}^{n} w_i^2$$

ANS15- The linear regression model contains an error term that is represented by ε. The error term is used to account for the variability in y that cannot be explained by the linear relationship between x and y. If ε were not present, that would mean that knowing x would provide enough information to determine the value of y.