Question Number	Answers
1	A
2	A
3	B
4	B
5	C
6	B
7	D
8	A
9	A
10	A
11	B
12	A B C

13. Explain term regularization?

Regularization is the process of adding tuning parameter (penalty term) to a model to induce smoothness in order to prevent overfitting. This is most often done by adding a constant multiple to an existing weight vector. This constant is often the L1 (Lasso - |I|) or L2 (Ridge - I). The model predictions should then minimize the loss function calculated on the regularized training set.

14. Which particular algorithms are used for regularization?

1. Ridge Regression
2. LASSO (Least Absolute Shrinkage and Selection Operator) Regression

Ridge Regression

- Ridge regression is also known as the L2 Regularization. Ridge regression is a method for analyzing data that suffer from multi-collinearity. Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.

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

Loss Function for Ridge Regression

Ridge regression adds a penalty (L2 penalty) to the loss function that is equivalent to the square of the magnitude of the coefficients. The regularization parameter (λ) regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

- $\lambda \to 0$, 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.

- $\lambda \to \infty$, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will approach zero (coefficients are close to zero, but not zero).

LASSO

LASSO regression is also known as the L1 Regularization (L1 penalty).

LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model.

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

Loss Function for LASSO Regression

LASSO regression adds a penalty (L1 penalty) to the loss function that is equivalent to the magnitude of the coefficients. In LASSO regression, the penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the regularization parameter λ is sufficiently large. 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.

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

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

$$\mathbf{Y=\alpha X + \beta\rho + \epsilon}$$

Where α, β =Constant parameters X, ρ =Independent variables ϵ =Error term

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.