# 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**

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 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

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:

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.**

ASSIGNMENT – 39MACHINE LEARNING
Q13 and Q15 are subjective answer type questions, Answer them briefly.
**13. Explain the term regularization?**
Regularization is a technique used in machine learning and statistical modeling to prevent overfitting and improve the generalization of the model. Overfitting occurs when a model learns not only the underlying pattern from the training data but also noise and random fluctuations. This causes the model to perform well on the training data but poorly on new, unseen data.
Regularization introduces a penalty term to the model's objective function to discourage complex models that are more likely to overfit. The idea is to find a balance between fitting the training data well and keeping the model simple enough to generalize to new data effectively.

There are different types of regularization techniques commonly used:

1. **L1 Regularization (Lasso):** This adds a penalty proportional to the absolute value of the magnitude of coefficients. It encourages sparsity in feature selection by shrinking some coefficients to zero.

2. **L2 Regularization (Ridge):** This adds a penalty proportional to the square of the magnitude of coefficients. It tends to shrink the coefficients of correlated features towards each other.

3. **Elastic Net:** This combines both L1 and L2 regularization penalties. It balances between the sparsity-inducing ability of L1 regularization and the grouping effect of L2 regularization.

4. **Dropout:** In neural networks, dropout is a technique where randomly selected neurons are ignored during training. This prevents neurons from co-adapting too much to the training data and acts as a form of regularization.

14. **Which particular algorithms are used for regularization?**
Regularization techniques can be applied to various machine learning algorithms across different domains. Here are some common algorithms where regularization is often used:

1. **Linear Regression:**
   - Ridge Regression: Uses L2 regularization to penalize large coefficients.
   - Lasso Regression: Uses L1 regularization to induce sparsity in the coefficients.
   - Elastic Net: Combines L1 and L2 regularization.

2. **Logistic Regression:**
   - Regularization techniques similar to those used in linear regression can be applied to logistic regression to prevent overfitting.

3**. Support Vector Machines (SVM):**
   - SVMs can benefit from regularization to control the margin and prevent overfitting, particularly using L2 regularization.

4. **Neural Networks:**
   - **Dropout:** A form of regularization where random neurons are ignored during training to prevent over-reliance on specific neurons.
   - **Weight Decay:** Equivalent to L2 regularization, where a penalty is applied to the squared magnitude of weights.
   - **Batch Normalization:** While primarily used for accelerating training and improving performance, it also acts as a regularizer by normalizing activations between layers.

5. **Decision Trees:**
   - Regularization in decision trees can involve limiting the depth of the tree (pruning) or controlling the minimum number of samples required to split a node (min_samples_split).

6. **Ensemble Methods (e.g., Random Forests, Gradient Boosting Machines):**
   - Regularization can be applied to individual base learners within ensembles to prevent overfitting, similar to their respective base algorithms (e.g., decision trees in Random Forests).

7. **Bayesian Models:**
   - Bayesian regularization techniques can be employed to control model complexity and prevent overfitting by specifying prior distributions over model parameters.

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

In the context of linear regression, the "error" refers to the difference between the actual observed value of the dependent variable (target variable) and the value predicted by the linear regression model. This error is also known as the residual.

Let's break down the components of a linear regression equation to understand this better:

1. **Linear Regression Model Equation**
   For a simple linear regression with one independent variable (feature), the model can be represented as:
   $$ y = \beta_0 + \beta_1 x + \epsilon $$
   - $y$: Dependent variable (target variable) we are trying to predict.
   - $x$: Independent variable (feature) that we use to predict $y$.
   - $\beta_0$ and $\beta_1$: Coefficients or parameters of the linear regression model. $\beta_0$ is the intercept (the value of $y$ when $x = 0$), and $\beta_1$ is the slope (the change in $y$ for a unit change in $x$).
   - $\epsilon$: Error term or residual, representing the difference between the predicted value $\hat{y}$ and the actual observed value $y$.

2. **Error (Residual):**
   - The error $\epsilon$ is the difference between the actual value $y$ and the predicted value $\hat{y}$:
     $$ \epsilon = y - \hat{y} $$
   - Ideally, a good linear regression model minimizes this error across all data points. In other words, it aims to make the sum of squared residuals (RSS) as small as possible.

3. **Purpose of the Error Term:**
   - The error term $\epsilon$ captures the influence of all other factors that affect $y$ but are not included in the model $\beta_0 + \beta_1 x$.

- These factors can include measurement errors, omitted variables, or random disturbances in the data that affect $y$ but are not explained by $x$.

4. **Assumptions of Linear Regression:**
   - The error term $\epsilon$ is assumed to have a mean of zero ($E[\epsilon] = 0$), indicating that, on average, the model does not systematically over- or under-predict $y$.
   - The errors are also assumed to be independently and identically distributed (i.i.d.), and typically follow a normal distribution ($\epsilon \sim N(0, \sigma^2)$).