# 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?
   1. Least Square Error B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

1. Which of the following statement is true about outliers in linear regression?
   1. Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can’t say D) none of these

1. A line falls from left to right if a slope is ?
   1. Positive B) Negative

C) Zero D) Undefined

1. Which of the following will have symmetric relation between dependent variable and independent variable?
   1. Regression B) Correlation

C) Both of them D) None of these

1. Which of the following is the reason for over fitting condition?
   1. High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none of these

1. If output involves label then that model is called as:
   1. Descriptive model B) Predictive modal

C) Reinforcement learning D) All of the above

1. Lasso and Ridge regression techniques belong to ?
   1. Cross validation B) Removing outliers

C) SMOTE D) Regularization

1. To overcome with imbalance dataset which technique can be used?
   1. Cross validation B) Regularization

C) Kernel D) SMOTE

1. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses to make graph?
   1. TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
   1. True B) False
2. Pick the feature extraction from below:
   1. Construction bag of words from a email
   2. Apply PCA to project high dimensional data
   3. Removing stop words
   4. Forward selection

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

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
   1. We don’t have to choose the learning rate.
   2. It becomes slow when number of features is very large.
   3. We need to iterate.
   4. It does not make use of dependent variable.

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

1. **Explain the term regularization?**

Regularization is a technique used in machine learning and statistics to prevent overfitting and improve the generalization of a model. It adds a penalty to the model's complexity, which helps to ensure that the model does not become too complex or fit the noise in the training data.

Here’s a detailed explanation:

### Why Regularization?

* **Overfitting**: When a model is too complex, it may fit the training data very well but perform poorly on new, unseen data. This is because the model may have learned the noise and specific patterns of the training data rather than the underlying data distribution.
* **Generalization**: Regularization helps to improve the model's ability to generalize to new data by discouraging overly complex models.

### Common Regularization Techniques

1. **L1 Regularization (Lasso)**:
   * **Penalty**: Adds a penalty equal to the absolute value of the magnitude of coefficients.
   * **Effect**: Can lead to sparsity in the coefficients, meaning some coefficients may be exactly zero. This can be useful for feature selection.
   * **Formula**: Cost Function=Loss Function+λ∑i∣wi∣\text{Cost Function} = \text{Loss Function} + \lambda \sum\_{i} |w\_i|Cost Function=Loss Function+λ∑i​∣wi​∣
   * **Where**: λ\lambdaλ is the regularization parameter, and wiw\_iwi​ are the model coefficients.
2. **L2 Regularization (Ridge)**:
   * **Penalty**: Adds a penalty equal to the square of the magnitude of coefficients.
   * **Effect**: Helps to shrink the coefficients but does not lead to sparsity. It generally reduces the impact of less important features.
   * **Formula**: Cost Function=Loss Function+λ∑iwi2\text{Cost Function} = \text{Loss Function} + \lambda \sum\_{i} w\_i^2Cost Function=Loss Function+λ∑i​wi2​
   * **Where**: λ\lambdaλ is the regularization parameter, and wiw\_iwi​ are the model coefficients.
3. **Elastic Net Regularization**:
   * **Penalty**: Combines both L1 and L2 penalties.
   * **Effect**: Provides a balance between Lasso and Ridge regularization. Useful when dealing with many features.
   * **Formula**: Cost Function=Loss Function+λ1∑i∣wi∣+λ2∑iwi2\text{Cost Function} = \text{Loss Function} + \lambda\_1 \sum\_{i} |w\_i| + \lambda\_2 \sum\_{i} w\_i^2Cost Function=Loss Function+λ1​∑i​∣wi​∣+λ2​∑i​wi2​

### Regularization Parameter (λ\lambdaλ)

* **Role**: Controls the strength of the penalty. A higher value of λ\lambdaλ increases the penalty, leading to more regularization and potentially more bias. A lower value reduces the penalty, potentially leading to less regularization and more variance.

### Summary

Regularization helps to balance the trade-off between fitting the training data well and maintaining a model that generalizes well to unseen data. By adding a regularization term to the loss function, we can control model complexity and improve overall performance.

1. **Which particular algorithms are used for regularization?**

Regularization techniques are integrated into various machine learning algorithms to control model complexity and improve generalization. Here are some specific algorithms that use regularization:

### 1. **Linear Regression**

* **Lasso Regression (L1 Regularization)**: Adds a penalty equal to the absolute value of the magnitude of coefficients, leading to sparse models where some coefficients may be zero.
* **Ridge Regression (L2 Regularization)**: Adds a penalty equal to the square of the magnitude of coefficients, which shrinks coefficients but does not set them to zero.
* **Elastic Net Regression**: Combines both L1 and L2 penalties, providing a balance between Lasso and Ridge regression.

### 2. **Logistic Regression**

* **L1 Regularization**: Helps with feature selection by driving some coefficients to zero.
* **L2 Regularization**: Shrinks coefficients to reduce overfitting while keeping all features in the model.
* **Elastic Net Regularization**: Combines L1 and L2 penalties for a balanced approach.

### 3. **Support Vector Machines (SVM)**

* **Regularization Parameter (C)**: Controls the trade-off between achieving a low training error and a low testing error. A high value of C emphasizes a low training error, while a low value of C emphasizes a low testing error.

### 4. **Neural Networks**

* **L1 and L2 Regularization**: Applied to the weights of the network to prevent overfitting.
* **Dropout**: A technique where randomly selected neurons are ignored during training to prevent overfitting and ensure that the model does not rely too heavily on any particular neuron.

### 5. **Generalized Linear Models (GLM)**

* **L1 Regularization**: For feature selection and sparse models.
* **L2 Regularization**: To prevent overfitting by penalizing large coefficients.

### 6. **Gradient Boosting Machines (GBM)**

* **Regularization Techniques**: Such as shrinkage (learning rate) and subsampling, are used to control overfitting and improve model performance.

### 7. **Decision Trees and Random Forests**

* **Regularization Parameters**: Such as maximum depth, minimum samples split, and minimum samples leaf are used to prevent trees from growing too deep and capturing noise in the training data.

These algorithms incorporate regularization to improve model robustness and performance by managing complexity and preventing overfitting.

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

In linear regression, the term "error" refers to the difference between the observed values and the values predicted by the linear model. It is a crucial concept for understanding how well the model fits the data. Here's a detailed explanation:

### Error in Linear Regression

1. **Definition of Error**:
   * **Error (or Residual)**: For a given data point, the error is the difference between the actual value of the dependent variable (observed value) and the value predicted by the linear regression model.

Errori=yi−y^i\text{Error}\_i = y\_i - \hat{y}\_iErrori​=yi​−y^​i​

where:

* + yiy\_iyi​ is the observed value for the iii-th data point.
  + y^i\hat{y}\_iy^​i​ is the predicted value for the iii-th data point.

1. **Types of Errors**:
   * **Residual Error**: The error for each individual data point, representing how far off the model's prediction is from the actual value.
   * **Sum of Squared Errors (SSE)**: The sum of the squares of all residual errors. It is used to quantify the total discrepancy between the observed values and the model's predictions.

SSE=∑i=1n(yi−y^i)2\text{SSE} = \sum\_{i=1}^{n} (y\_i - \hat{y}\_i)^2SSE=i=1∑n​(yi​−y^​i​)2

* + **Mean Squared Error (MSE)**: The average of the squared errors, which provides a measure of the average discrepancy between the observed and predicted values.

MSE=1n∑i=1n(yi−y^i)2\text{MSE} = \frac{1}{n} \sum\_{i=1}^{n} (y\_i - \hat{y}\_i)^2MSE=n1​i=1∑n​(yi​−y^​i​)2

* + **Root Mean Squared Error (RMSE)**: The square root of the MSE, giving an error measure in the same units as the dependent variable.

RMSE=MSE\text{RMSE} = \sqrt{\text{MSE}}RMSE=MSE​

* + **Mean Absolute Error (MAE)**: The average of the absolute differences between the observed and predicted values.

MAE=1n∑i=1n∣yi−y^i∣\text{MAE} = \frac{1}{n} \sum\_{i=1}^{n} |y\_i - \hat{y}\_i|MAE=n1​i=1∑n​∣yi​−y^​i​∣

1. **Significance of Error**:
   * **Model Fit**: The size and distribution of errors give insights into how well the model fits the data. Smaller errors generally indicate a better fit.
   * **Performance Evaluation**: Errors are used to evaluate the performance of the regression model and to compare different models. Lower error values typically indicate better model performance.
2. **Error Components**:
   * **Model Error**: The difference between the true relationship and the model’s approximation.
   * **Measurement Error**: Errors arising from inaccuracies in data collection or measurement.

In summary, errors in linear regression are the deviations of predicted values from actual values, and understanding these errors helps in assessing model performance and making improvements.