* **Linear regression**

Linear regression is a commonly used statistical technique for modeling the relationship between a dependent variable and one or more independent variables. It is used to predict the value of the dependent variable based on the values of the independent variables. In linear regression, a linear relationship is assumed between the dependent variable and the independent variable(s), and the aim is to find the best-fitting line that minimizes the sum of the squared differences between the predicted values and the actual values.

Linear regression can be used for both simple and multiple regression models. Simple linear regression involves one independent variable, while multiple linear regression involves two or more independent variables. The technique is widely used in various fields such as finance, economics, engineering, and social sciences for making predictions and understanding relationships between variables. A reference for further reading on linear regression can be found in the book "Applied Linear Regression" by Sanford Weisberg (Wiley, 2014).

Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are three commonly used metrics to evaluate the performance of a regression model. They are measures of the average difference between the predicted values and the actual values.

MSE is the average of the squared differences between the predicted values and the actual values. It is given by:

where y\_pred is the predicted value, y\_actual is the actual value, and n is the number of data points.

MAE is the average of the absolute differences between the predicted values and the actual values. It is given by:

where |x| represents the absolute value of x.

RMSE is the square root of the MSE. It is given by:

The RMSE is often used as a preferred metric because it has the same unit as the dependent variable and is more interpretable than the MSE.

In general, a lower value of these metrics indicates better performance of the model. However, the choice of metric depends on the specific problem and the goals of the analysis.

* **Neural Nwtwork**

Neural networks are a class of machine learning algorithms that are inspired by the structure and function of the human brain. They consist of interconnected nodes, called neurons, that work together to perform complex computations. Neural networks are capable of learning patterns and relationships in data, making them suitable for a wide range of applications such as image and speech recognition, natural language processing, and autonomous control. The ability to learn from data makes neural networks useful in situations where traditional rule-based systems would be impractical or impossible to implement.

Neural networks can be trained using various algorithms, including backpropagation and stochastic gradient descent. The training process involves adjusting the weights and biases of the neurons in the network to minimize the error between the predicted output and the actual output. A reference for further reading on neural networks can be found in the book "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (MIT Press, 2016).

Neural network structure is optimized in this analysis via hyperparameter tuning to have more precise prediction. Hyperparameter tuning is a critical step in building neural network models that involves selecting the optimal values for the hyperparameters. Hyperparameters are parameters that are set before the training process begins, and they affect the behavior and performance of the neural network.

Common hyperparameters that require tuning include the learning rate, number of hidden layers, number of neurons in each layer, activation function, and regularization strength. The optimal values for these hyperparameters depend on the specific problem, dataset, and architecture of the neural network. Hyperparameter tuning can be performed using various methods, including grid search, random search, Bayesian optimization, and genetic algorithms. These methods involve searching the hyperparameter space to find the optimal combination of hyperparameters that maximize the performance of the neural network on a validation set.

A reference for further reading on hyperparameter tuning in neural networks can be found in the book "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron (O'Reilly Media, 2019).

* **Random Forest**

Random forests are a popular machine learning algorithm that is used for classification and regression tasks. They are an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of the model. In a random forest, multiple decision trees are trained on random subsets of the data and random subsets of the features. This randomness helps to reduce overfitting and improve the generalization ability of the model. The final prediction is made by averaging the predictions of all the trees in the forest.

Random forests are widely used in various fields such as finance, healthcare, and image analysis, where they have shown to be effective in handling high-dimensional data and dealing with missing or noisy data. A reference for further reading on random forests can be found in the paper "Random Forests" by Leo Breiman (Machine Learning, 2001).

* **Decision Tree**

Decision trees are a popular machine learning algorithm that is used for classification and regression tasks. They are a type of supervised learning algorithm that works by recursively partitioning the data into subsets based on the values of the input features, until a stopping criterion is met. In a decision tree, each internal node represents a test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label or a numerical value. The structure of the tree is learned from the training data using various algorithms, such as the ID3, C4.5, or CART algorithm.

Decision trees are widely used in various fields such as finance, healthcare, and image analysis, where they have shown to be effective in handling high-dimensional data and dealing with missing or noisy data. A reference for further reading on decision trees can be found in the book "Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei (Morgan Kaufmann, 2011).

* **Classification metrics**

Precision, recall, and F1 score are common metrics used to evaluate the performance of classification models.

Precision measures the proportion of true positive predictions out of all positive predictions. It is calculated as:

precision = true positives / (true positives + false positives)

Recall measures the proportion of true positive predictions out of all actual positives. It is calculated as:

recall = true positives / (true positives + false negatives)

The F1 score is the harmonic mean of precision and recall and provides a single metric that balances both measures. It is calculated as:

F1 score = 2 \* (precision \* recall) / (precision + recall)

These metrics are useful for evaluating the performance of classification models in different scenarios. For example, high precision is important when the cost of false positives is high, while high recall is important when the cost of false negatives is high.