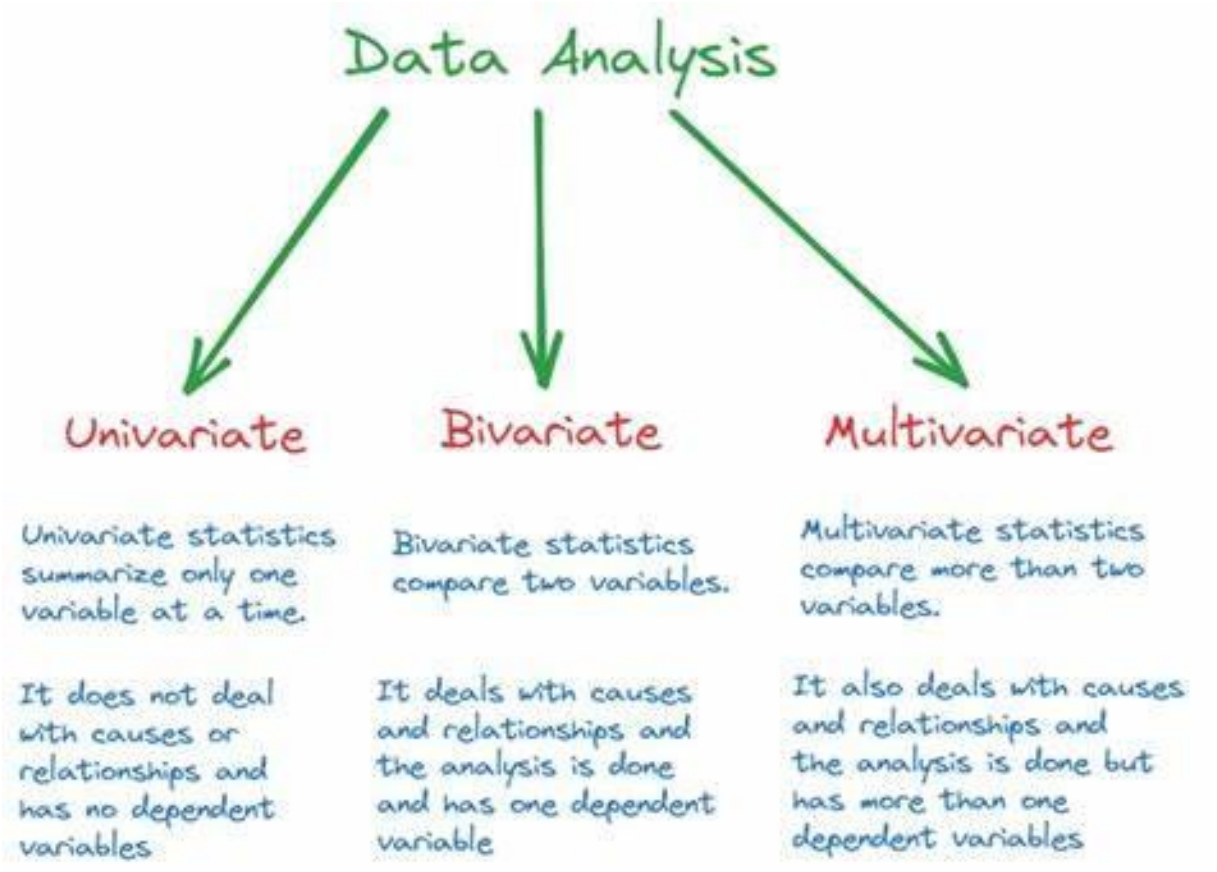
Artificial Intelligence Fullstack
[Course]
Week 8 – Machine Learning - Important Concepts

It is not about Theory, it is 20% Theory and 80% Practical – Technical/Development/Programming [Mostly Python based]

Univariate vs Bivariate vs Multivariate



Examples

Reference:

https://www.geeksforgeeks.org/univariate-bivariate-and-multivariate-data-and-its-analysis/
https://medium.com/analytics-vidhya/univariate-bivariate-and-multivariate-analysis-8b4fc3d8202c
https://www.modernanalyst.com/Careers/InterviewQuestions/tabid/128/ID/4904/Describe-the-difference-between-univariate-bivariate-and-multivariate-analysis.aspx

ML | Bais and Variance

Bias is simply defined as the inability of the model because of that there is some difference or error occurring between the model's predicted value and the actual value.

Let Y be the true value of a parameter, and let Y^ be an estimator of Y based on a sample of data. Then, the bias of the estimator Y^ is given by:

$$Bias(\hat{Y}) = E(\hat{Y}) - Y$$

where E(Y^) is the expected value of the estimator Y^. It is the measurement of the model that how well it fits the data.

Variance is the measure of spread in data from its mean position.

Let Y be the actual values of the target variable, and Y^ be the predicted values of the target variable. Then the variance of a model can be measured as the expected value of the square of the difference between predicted values and the expected value of the predicted values.

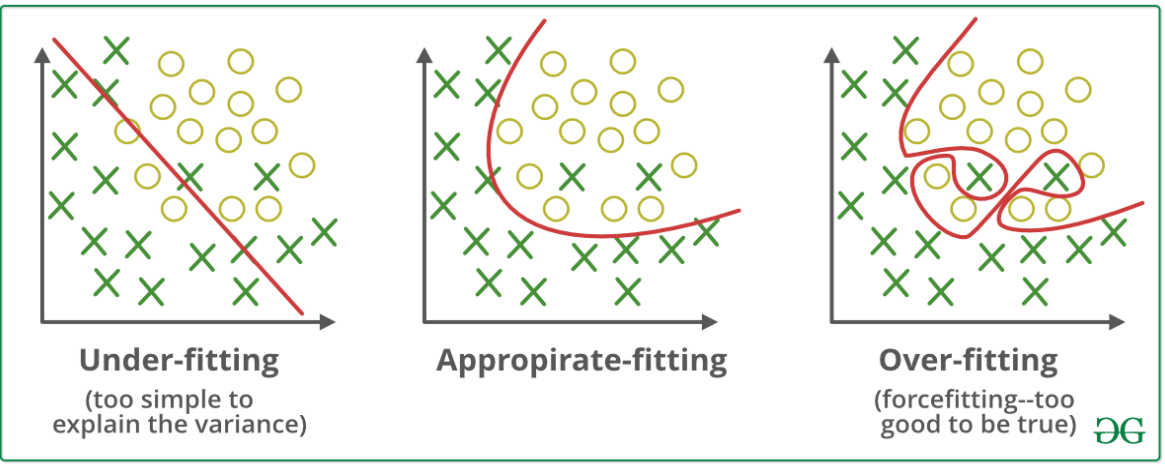
$$Variance = E[(\hat{Y} - E[\hat{Y}])2]$$

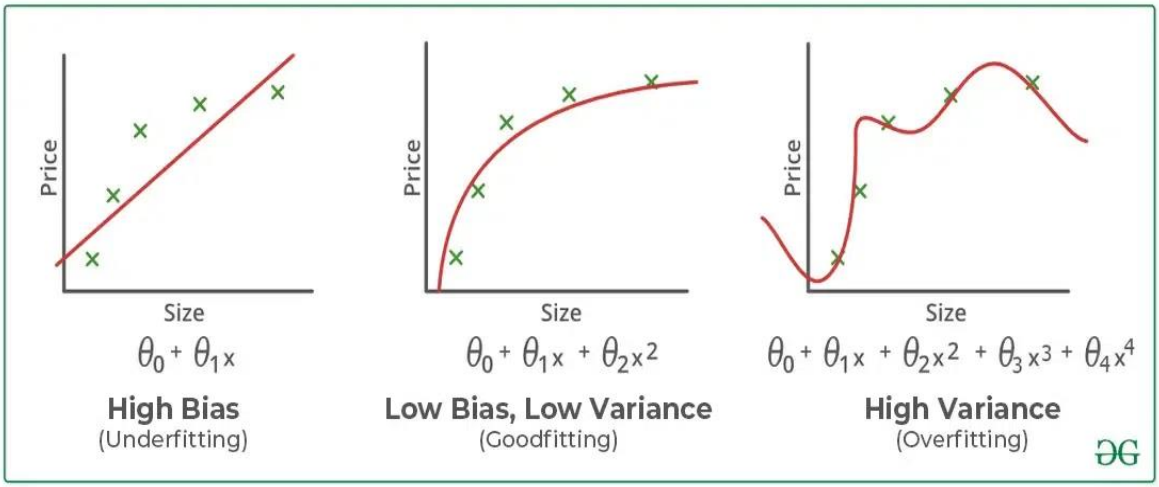
where $E[\bar{Y}]$ is the expected value of the predicted values. Here expected value is averaged over all the training data.

Reference:

https://www.geeksforgeeks.org/bias-vs-variance-in-machine-learning/
https://www.bmc.com/blogs/bias-variance-machine-learning/
https://www.tutorialspoint.com/machine_learning/machine_learning_bias_and_variance.htm
https://www.analyticsvidhya.com/blog/2020/08/bias-and-variance-tradeoff-machine-learning/
https://www.mastersindatascience.org/learning/difference-between-bias-and-variance/

ML | Underfitting and Overfitting





Balance Between Bias and Variance

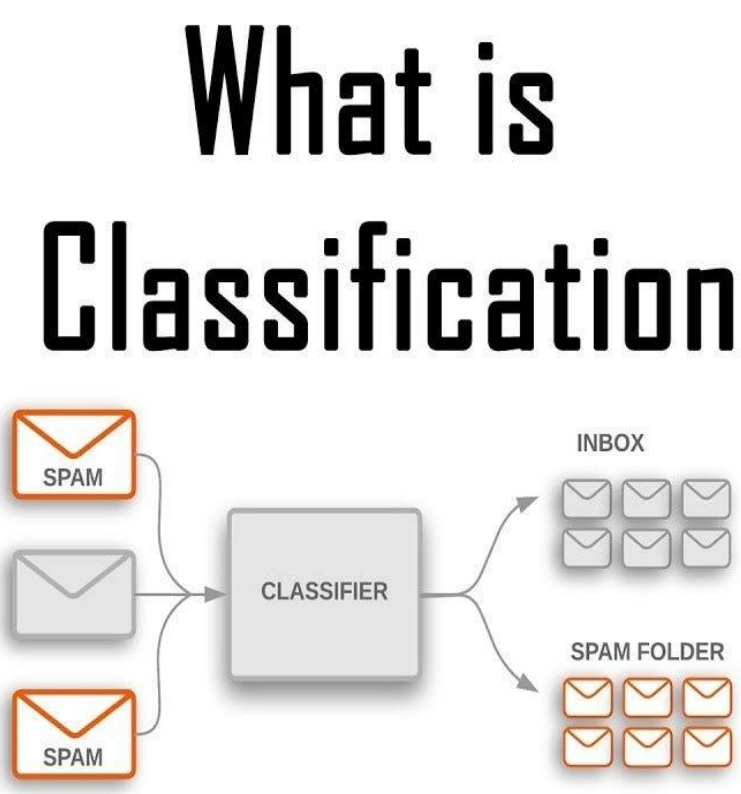
Reference:

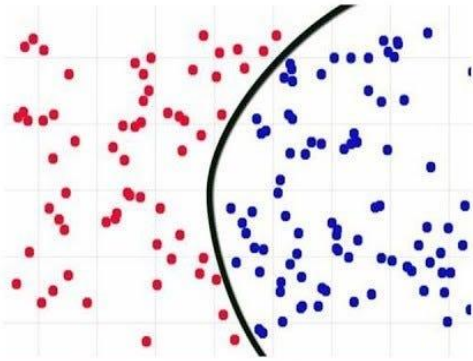
https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/
https://www.coursera.org/articles/overfitting-vs-underfitting?msockid=2ec7dedc79d36ed91db7cd9078c76f83
Underfitting and Overfitting in Machine Learning | Baeldung on Computer Science

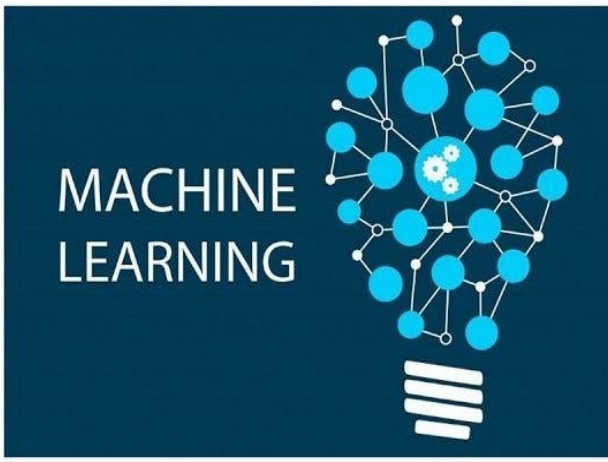
https://www.geeksforgeeks.org/bias-vs-variance-in-machine-learning/

ML | Classification

Classification is a supervised machine learning method where the model tries to predict the correct label of a given input data. In classification, the model is fully trained using the training data, and then it is evaluated on test data before being used to perform prediction on new unseen data.

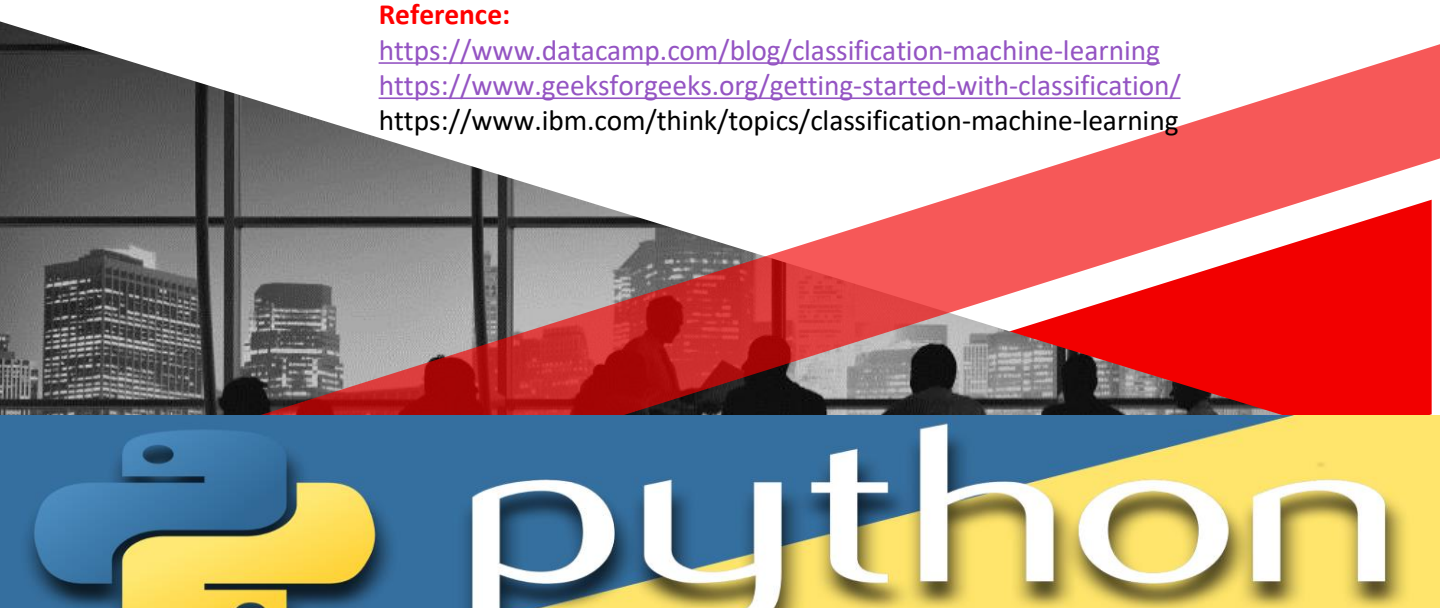






Reference:

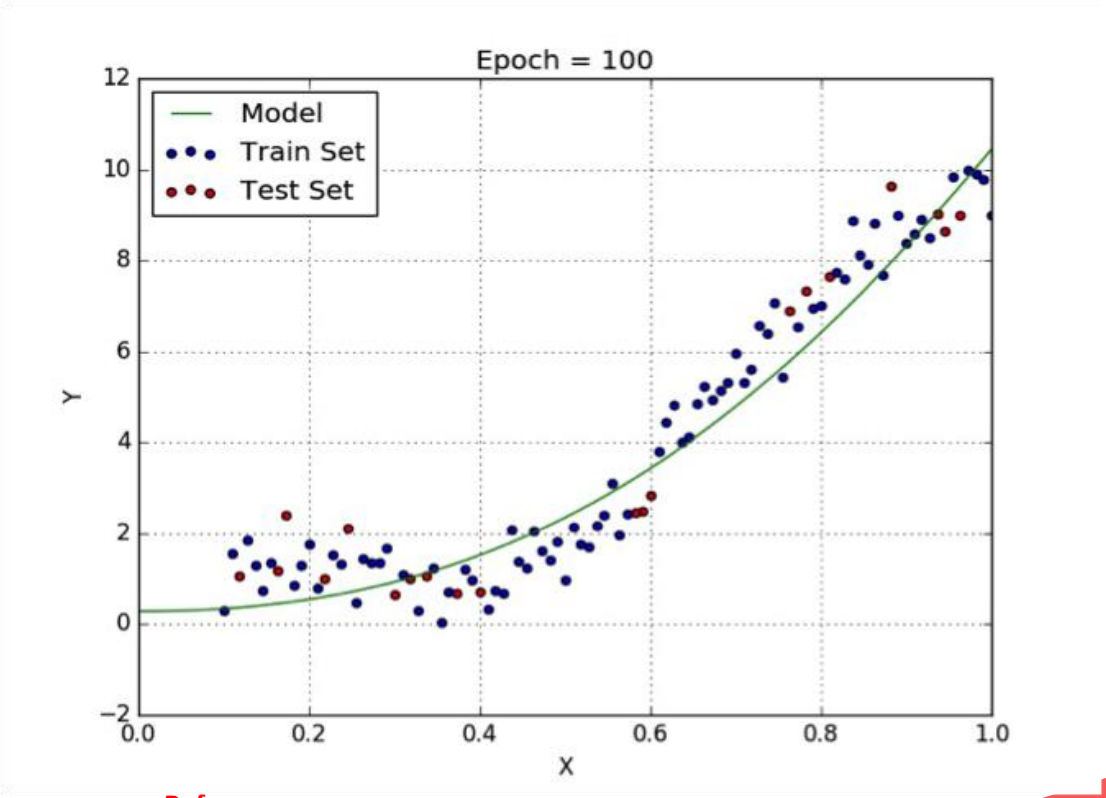
https://www.datacamp.com/blog/classification-machine-learning
https://www.geeksforgeeks.org/getting-started-with-classification/
https://www.ibm.com/think/topics/classification-machine-learning



ML | Regression

Regression in machine learning refers to a **supervised learning** technique where the goal is to predict a continuous numerical value based on one or more independent features. It find relationships between variables so that predictions can be made. w have two types of variables present in regression:

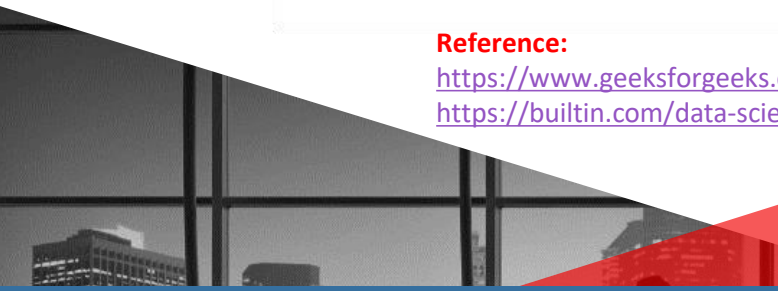
- **Dependent Variable (Target)**: The variable we are trying to predic e.g house price.
- **Independent Variables (Features)**: The input variables that influence the prediction e.g locality, number of rooms.



Reference:

https://www.geeksforgeeks.org/regression-in-machine-learning/

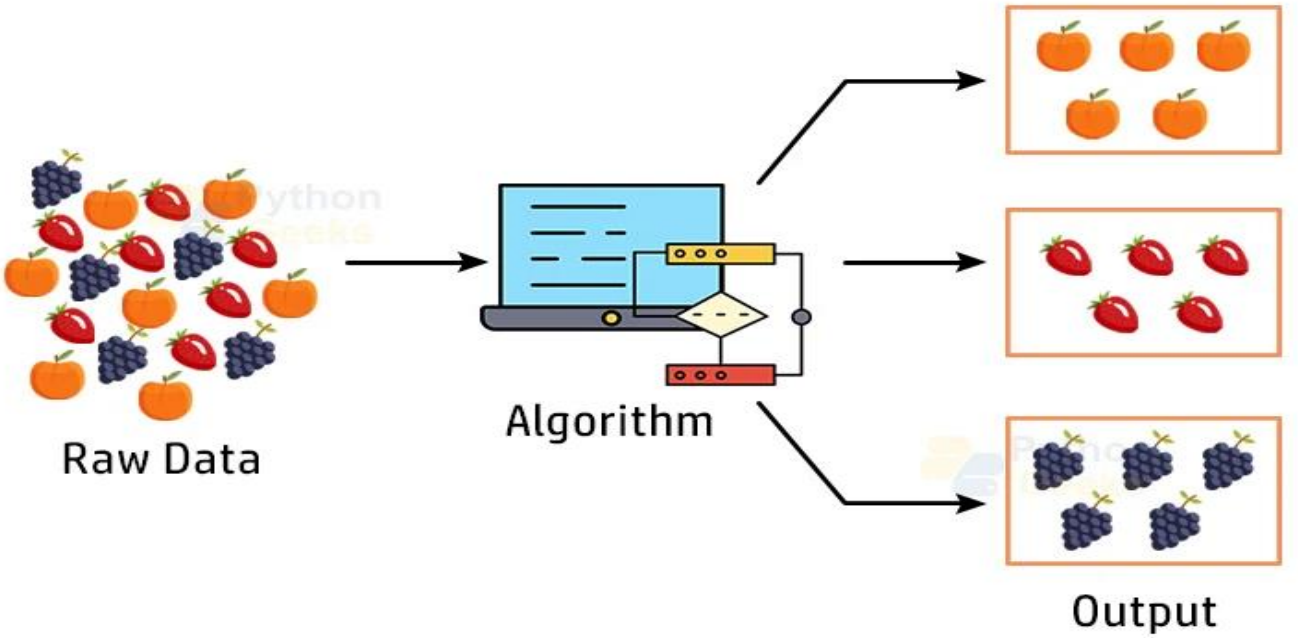
https://builtin.com/data-science/regression-machine-learning



python

ML | Clustering

The task of **grouping data points based on their similarity with each other is called Clustering or Cluster Analysis**. This method is defined under the branch of unsupervised learning, which aims at gaining insights from unlabelled data points.

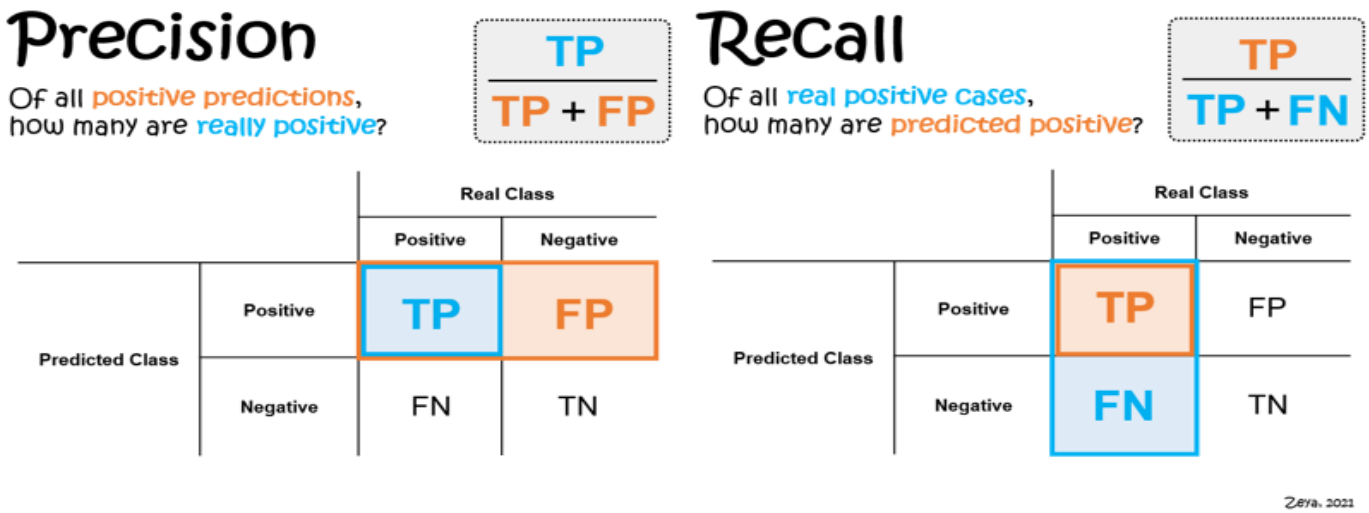


Reference:

https://www.geeksforgeeks.org/clustering-in-machine-learning/
https://www.datacamp.com/blog/clustering-in-machine-learning-5-essential-clustering-algorithms
https://www.ibm.com/think/topics/clustering

ML | Precision and Recall

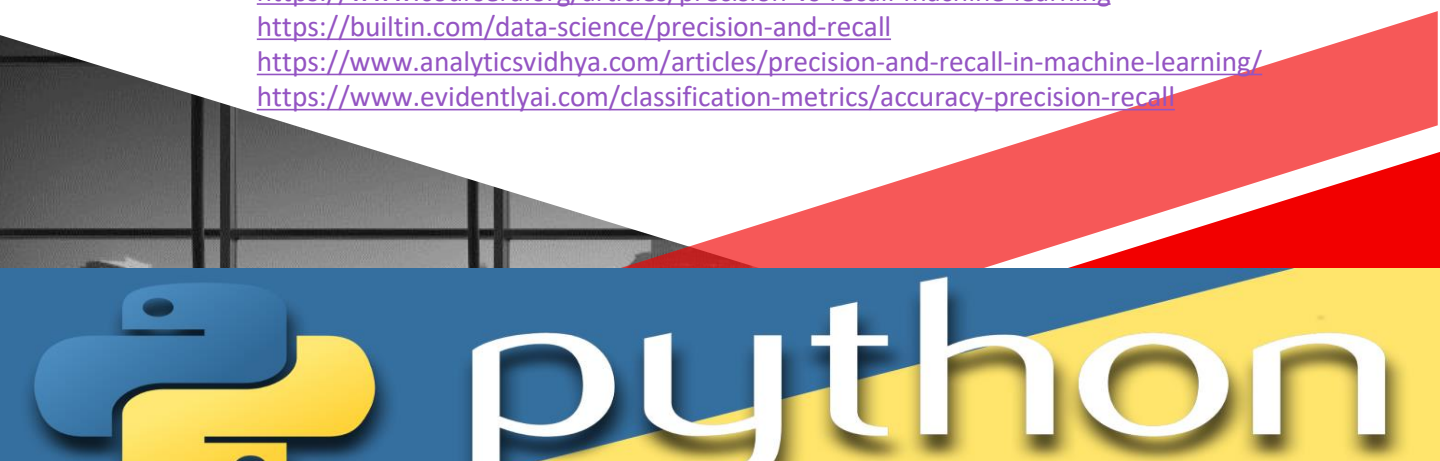
Precision and recall are two important measures of accuracy in a machine learning model. Precision is the ratio of a model's classification of all positive classifications as actually positive. Recall (also known as the true positive rate) is the ratio of all actual positives classified correctly as positives.



- ***True Positives (TP)**: Correctly predicted positive instances.*
- ***False Positives (FP)**: Incorrectly predicted positive instances.*
- ***True Negatives (TN)**: Correctly predicted negative instances.*
- ***False Negatives (FN)**: Incorrectly predicted negative instances.*

Reference:

https://towardsdatascience.com/precision-and-recall-made-simple-afb5e098970f/
https://www.coursera.org/articles/precision-vs-recall-machine-learning
https://builtin.com/data-science/precision-and-recall
https://www.analyticsvidhya.com/articles/precision-and-recall-in-machine-learning/
https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall



ML | F1 score

The F1 score (or F-measure) combines precision and recall into one metric so that you can optimize for the best precision and recall at the same time. Once you calculate your precision and recall from the confusion matrix, you can calculate your F1 score using the formula:

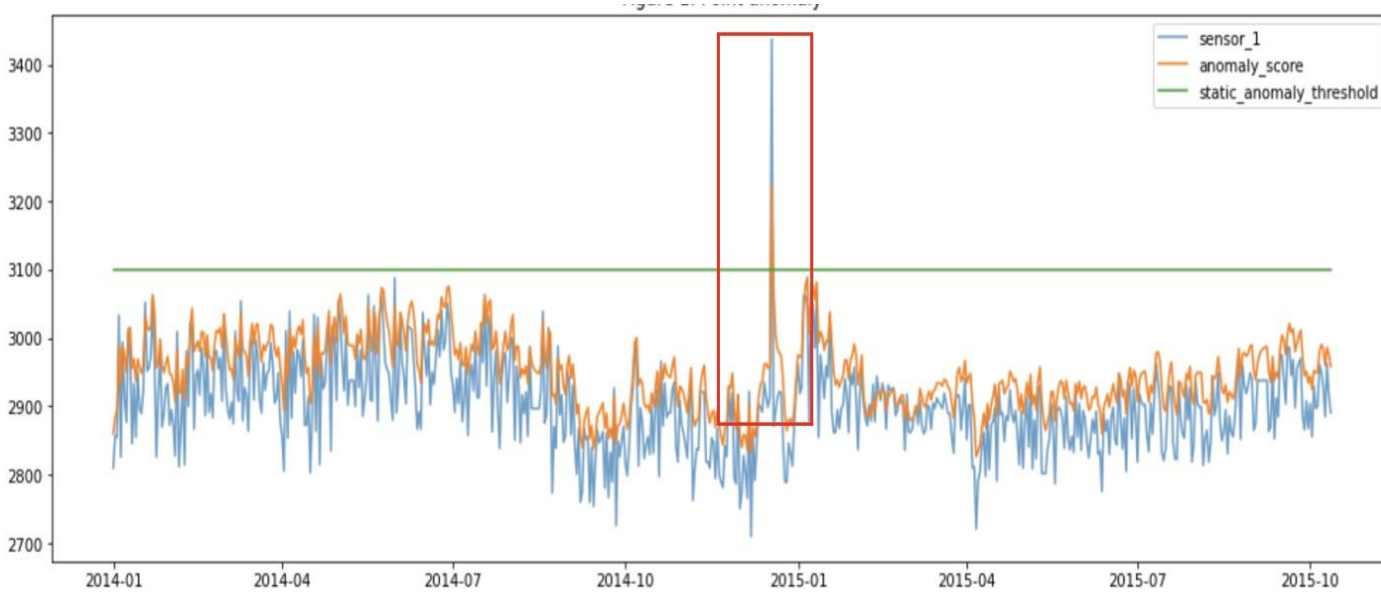
F1 score= 2*(precision*recall)/ precision + recall

This value represents the harmonic mean between precision and recall and is a common metric in imbalance classification problems. A perfect F1 score is 1.0, indicating perfect precision and recall, while the worst score possible is 0.0. When optimizing a machine learning model for precision and recall, you want to maximize your F1 score to achieve this balance.

Reference:
https://www.geeksforgeeks.org/f1-score-in-machine-learning/
https://encord.com/blog/f1-score-in-machine-learning/

ML | Anomaly Detection

Anomaly detection is examining specific data points and detecting rare occurrences that seem suspicious because they're different from the established pattern of behaviors. Anomaly detection isn't new, but as data increases manual tracking is impractical.

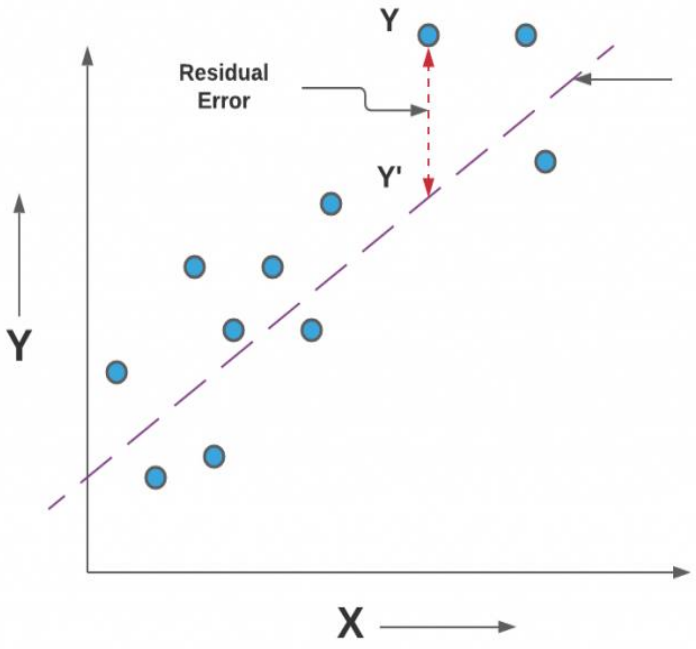


Reference:

https://aws.amazon.com/what-is/anomaly-detection/
https://www.ibm.com/think/topics/anomaly-detection
https://www.datacamp.com/tutorial/introduction-to-anomaly-detection
https://www.geeksforgeeks.org/what-is-anomaly-detection/

ML | Loss Function

The loss function, also referred to as the error function, is a crucial component in machine learning that quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.



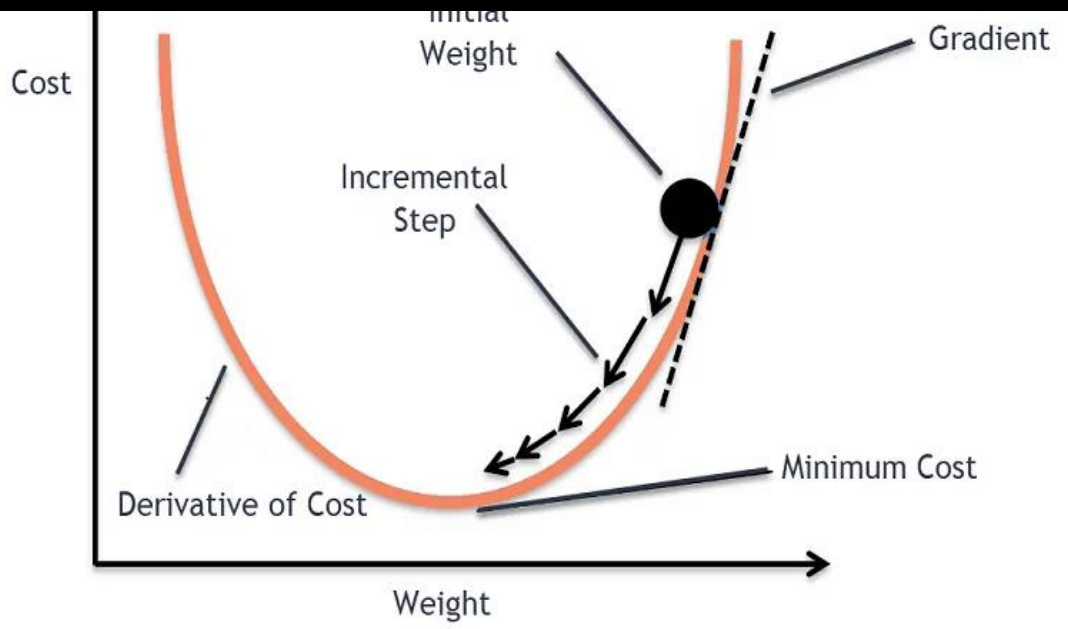
$$MSE = \frac{\text{Sum } (Y - Y')^2}{N}$$

Reference:

https://www.datacamp.com/tutorial/loss-function-in-machine-learning
https://www.geeksforgeeks.org/ml-common-loss-functions/
https://www.ibm.com/think/topics/loss-function
https://builtin.com/machine-learning/common-loss-functions

ML | Gradient Descent

Gradient descent is the backbone of the learning process for various algorithms, including linear regression, logistic regression, support vector machines, and neural networks which serves as a fundamental optimization technique to minimize the cost function of a model by iteratively adjusting the model parameters to reduce the difference between predicted and actual values, improving the model's performance.



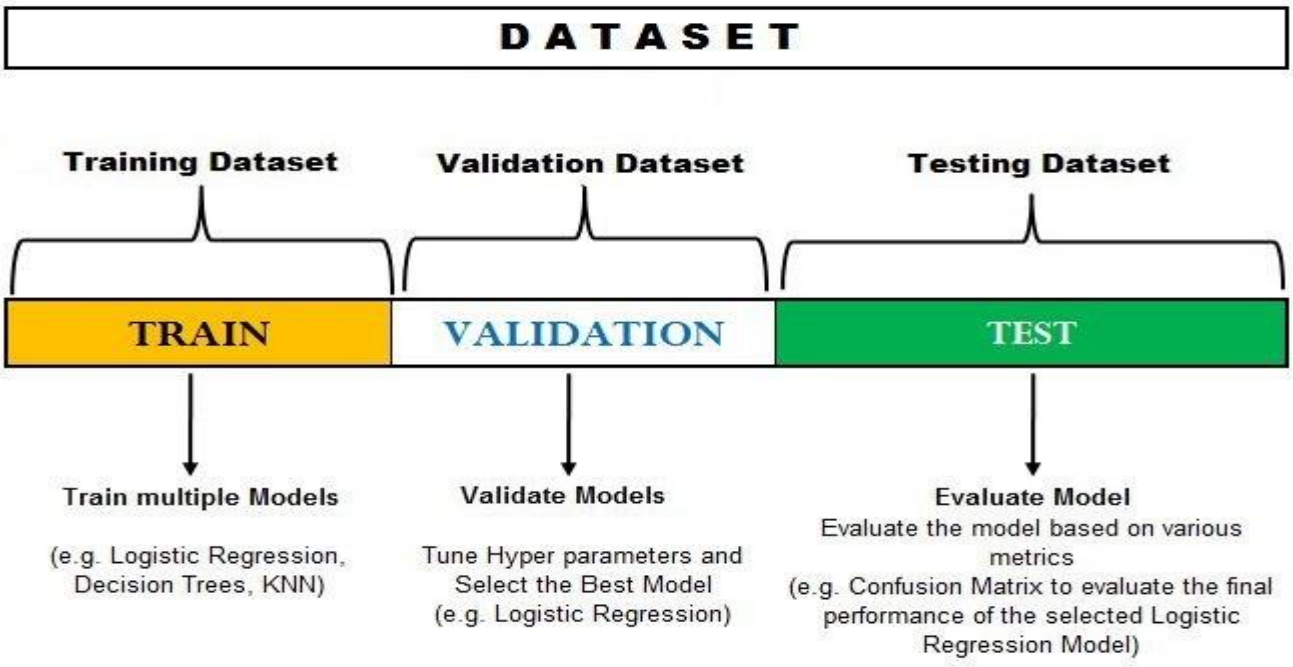
Reference:
https://www.datacamp.com/tutorial/tutorial-gradient-descent
https://www.ibm.com/think/topics/gradient-descent
https://builtin.com/data-science/gradient-descent
https://www.khanacademy.org/math/multivariable-calculus/applications-of-multivariable-derivatives/optimizing-multivariable-functions/a/what-is-gradient-descent
https://www.geeksforgeeks.org/gradient-descent-algorithm-and-its-variants/
https://www.analyticsvidhya.com/blog/2020/10/how-does-the-gradient-descent-algorithm-work-in-machine-learning/

ML | Training vs Testing vs Validation Sets

Training Set : This is the actual dataset from which a model trains .i.e. the model sees and learns from this data to predict the outcome or to make the right decisions.

This dataset is independent of the training set but has a somewhat similar type of probability distribution of classes and is used as a benchmark to evaluate the model, used only after the training of the model is complete.

The validation set is used to fine-tune the hyperparameters of the model and is considered a part of the training of the model.



Reference:

https://www.geeksforgeeks.org/training-vs-testing-vs-validation-sets/
https://medium.com/@jainvidip/understanding-train-test-and-validation-data-in-machine-learning-f8
https://kili-technology.com/training-data/training-validation-and-test-sets-how-to-split-machine-learn
https://en.wikipedia.org/wiki/Training,_validation,_and_test_data_sets

Thank you - for listening and participating

- ❑**Questions / Queries**
- ❑**Suggestions/Recommendation**
- ❑**Ideas…..?**

Shahzad Sarwar
Cognitive Convergence
https://cognitiveconvergence.com
shahzad@cognitiveconvergence.com
voice: +1 4242530744 (USA) +92-3004762901 (Pak)