Terms in machine learning

|  |  |  |  |
| --- | --- | --- | --- |
| Leaf colour | Stem colour | Leaf size | Type of tree |
| Light green | Brown | Small and tiny | Grapes |
| Dark green | Brown | Round and large | Papaya |
| Yellowish green | Brown | Narrow and small | Mango |

.Feature:

In the context of machine learning, a feature (also known as a variable or attribute) is an individual measurable property or characteristic of a data point that is used as input for a machine learning algorithm. E.g(leaf colour,stem colour,leaf size).

.Label:

Data labeling is the process of adding valuable information to raw data like images, text, audio, and videos. E.g(type of tree).

.Prediction:

The practise of using data to create predictions or foresee future events is known as machine learning prediction. E.g(with the help of this tree we can find the type of the tree).

.Outlier:

An outlier is a data point that significantly deviates from the rest of the data. It can be either much higher or much lower than the other data points, and its presence can have a significant impact on the results of machine learning algorithms.

.Test Data:

You will need unknown information to test your machine learning model after it was created (using your training data). This data is known as testing data.

.Training Data:

Training data is the power that supplies the model in machine learning, it is larger than testing data. Because more data helps to more effective predictive models.

.Model:

Machine Learning models are very powerful resources that automate multiple tasks and make them more accurate and efficient.

.Validation data:

A validation dataset is a sample of data held back from training your model that is used to give an estimate of model skill while tuning model’s hyperparameters.

.Hyperparameter:

A Machine Learning model is defined as a mathematical model with several parameters that need to be learned from the data. By training a model with existing data, we can fit the model parameters.   
However, there is another kind of parameter, known as ***Hyperparameters.***

.Epoch:

An epoch in machine learning is one complete pass through the entire training dataset. One pass means a complete **forward** and **backward** pass through the entire training dataset.

.Loss Function:

The loss function estimates how well a particular algorithm models the provided data.

.Learning Rate:

**Learning Rate** :The learning rate is a hyperparameter in [machine learning](https://www.geeksforgeeks.org/machine-learning/) that controls the step size at which the weights of a [neural network](https://www.geeksforgeeks.org/deep-learning-tutorial/) are updated during training. It specifies the amount by which the model’s parameters are altered in the direction opposite to the gradient of the loss function.

.Overfitting:

Overfitting in machine learning occurs when a model learns the training data too well.

.Underfitting:

This assessment looks back at how well the model learns from the data. It analyzes how the model generates predictions and compares the training data to the hypothetical forecasts it would create using the same data.

.Regularization:

Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function, discouraging the model from assigning too much importance to individual features or coefficients.

.Cross-Validation:

Cross validation is a technique used in machine learning to evaluate the performance of a model on unseen data. It involves dividing the available data into multiple folds or subsets, using one of these folds as a validation set, and training the model on the remaining folds.

.Feature Engineering:

Feature Engineering is the process of creating new features or transforming existing features to improve the performance of a machine-learning model.

.Dimensionality Reduction:

Dimensionality reduction is a technique used to reduce the number of features in a dataset while retaining as much of the important information as possible.

.Bias:

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.

.Variance:

The variability of model prediction for a given data point which tells us the spread of our data is called the variance of the model.