**What is Machine Learning?**

* Machine Learning is a subset of artificial intelligence which gives a machine the ability to learn without being explicitly programmed.
* Data is the key and the learning algorithm.

**Types of Machine Learning**

* Supervised Learning
* Unsupervised Learning
* Reinforcement Learning

# Supervised Learning

## Use case: Spam classifier

Filter used – Text filter, client filter

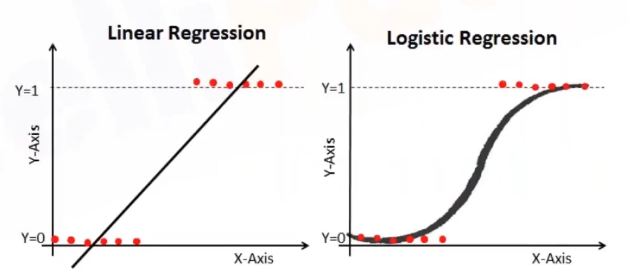
**Regression**:

A technique of finding the relationship between two or more variables.

Change in dependent variable is associated with a change in one or more independent variables.

There are various types of Regression, but we will focus on:

* Linear Regression
* Logistic Regression

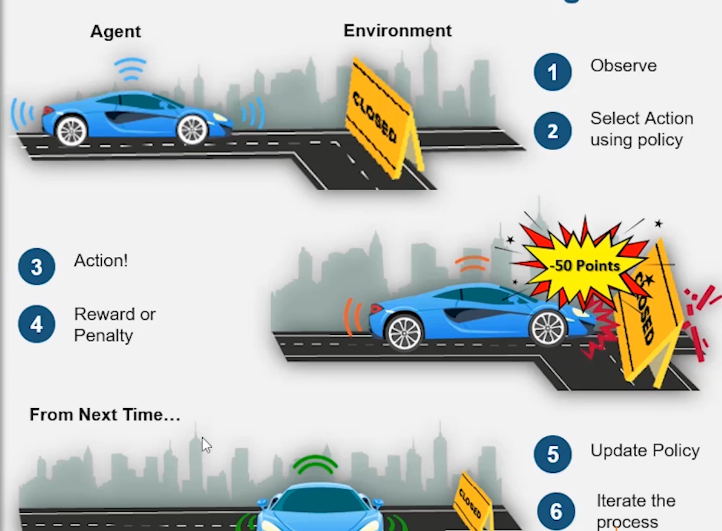


|  |  |
| --- | --- |
| Linear Regression | Logistic Regression |
| Continuous Variables | Categorical Variables |
| Solves Regression Issue | Solves Classification Issue |
| Straight Line | S-Curve |

**Linear Regression:**

* Simple linear regression is useful for finding relationship between two continuous variables
* One is predictor or independent variable and other is response or dependent variable

Reinforcement Learning:



**1. Accuracy:**

The proportion of correctly classified instances among all instances in a dataset. It's calculated as

* **Use**: Indicates the overall effectiveness of a model but can be misleading if the data is imbalanced.

**2. Confusion Matrix:**

* A table used to describe the performance of a classification model. It shows the counts of true positives, true negatives, false positives, and false negatives.
* **True Positives (TP)**: Correctly predicted positive instances.
* **True Negatives (TN)**: Correctly predicted negative instances.
* **False Positives (FP)**: Incorrectly predicted positive instances.
* **False Negatives (FN)**: Incorrectly predicted negative instances.

**3. Precision:**

* The proportion of true positive predictions among all positive predictions. It’s calculated as
* **Use**: Measures the accuracy of the positive predictions, useful when the cost of false positives is high.

**4. Recall:**

* The proportion of true positive predictions among all actual positives. It’s calculated as
* **Use**: Measures the ability of a model to find all relevant instances, important when missing positive cases is costly.

**5. F1 Score:**

* The harmonic mean of precision and recall. It’s calculated as
* **Use**: Provides a single metric that balances precision and recall, useful for imbalanced datasets.

**6. K-Fold Cross Validation:**

* A technique to evaluate the performance of a model by splitting the data into subsets (folds), training the model on folds, and testing it on the remaining fold. This process is repeated times, with each fold used as the test set once.
* **Use**: Helps in assessing the model's performance more reliably by reducing overfitting.

**7. Mean Squared Error (MSE):**

* **Definition**: The average squared difference between the actual values and the predicted values. It’s calculated as
* **Use**: A measure of the quality of an estimator; it’s always non-negative and values closer to zero are better.

**8. Root Mean Squared Error (RMSE):**

* **Definition**: The square root of the Mean Squared Error. It’s calculated as
* **Use**: Provides an interpretable error metric in the same units as the target variable, often preferred for regression problems.

**9. Hyperparameters:**

* **Definition**: Parameters that are set before the learning process begins and control the behavior of the training process (e.g., learning rate, number of trees in a random forest).
* **Use**: Tuned to optimize the performance of the model.

**10. Parameters:**

* The internal variables of the model that are learned from the training data (e.g., weights in a neural network).
* **Use**: Directly influence the predictions of the model.

**11. Tuning Model:**

* The process of optimizing hyperparameters to improve the performance of a model.
* **Use**: Ensures the model is well-suited for the specific problem and dataset.

**12. Learning Rate:**

* A hyperparameter that controls the step size at each iteration while moving toward a minimum of the loss function.
* **Use**: Affects the speed and quality of the model's learning process.

**13. Epochs:**

* One complete pass through the entire training dataset. Multiple epochs are often used to improve model performance.
* **Use**: Helps the model to learn from the data by iterating multiple times over the training set.

# Annotate: