Machine Learning Revision

Tags	
Due Date	
Formula	
Group	Studies
Project	
Seleção	
Status	Not started
Type	Draft
Version	

Definition

- subset of Artificial Intelligence
- enables systems to learn and improve from experience without being explicitly programmed.
- it involves feeding a computer data and allowing it to learn and improve on its own
- making predictions or decisions based on that data.

Usage of Machine Learning

- Recommender systems (e.g. Netflix, Amazon)
- Image recognition (e.g. self-driving cars, medical image analysis)
- Natural language processing (e.g. chatbots, speech recognition)
- Fraud detection (e.g. credit card fraud detection)
- Predictive maintenance (e.g. monitoring machines in factories)

Popularity of Machine Learning

- increasingly popular in recent years

- due to its ability to improve the accuracy and efficiency of automated systems
- availability of open source machine learning libraries has made it more accessible for developers

Differences between Machine Learning and Traditional Programming

Traditional programming:

- involves explicitly defining a set of rules and instructions for a computer to follow in order to complete a task.
- good as the rules and instructions provided by the programmer.
- deterministic, meaning that given the same input, they will always produce the same output.
- used for tasks that can be explicitly defined and solved

Machine learning:

- involves feeding a computer data and allowing it to learn and improve on its own.
- can adapt to new data and improve over time.
- probabilistic, meaning that they can produce different outputs for the same input based on the current state of the model and the randomness introduced during training.
- used for tasks that involve patterns and relationships in data that are difficult to explicitly define.

Learning Types

Supervised learning

- involves training a model on a labeled dataset where the correct outputs are known.
- learns to make predictions based on patterns in the data and is then used to make predictions on new, unlabeled data.

Unsupervised learning

- involves training a model on an unlabeled dataset where the correct outputs are unknown.

- learns to identify patterns and relationships in the data and is then used to find similar patterns in new, unlabeled data.

Reinforcement learning

- involves training a model to make decisions based on feedback from the environment.
- learns to take actions that maximize a reward signal, and is then used to make decisions in new, unseen environments.

Supervised learning

Classification:

- used when the output variable is categorical.
- learns to map input variables to a discrete output variable.

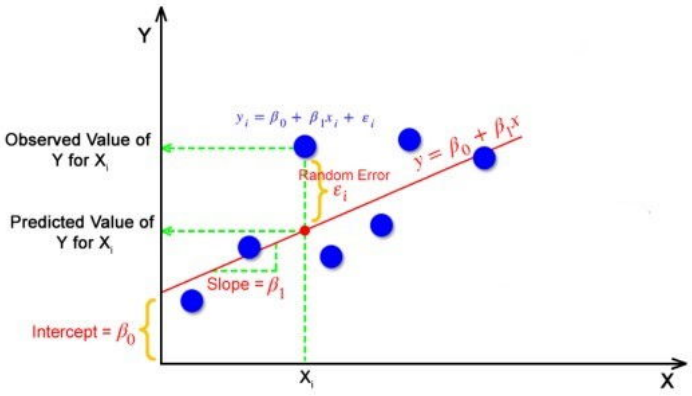
Regression:

- used when the output variable is continuous.
- learns to map input variables to a continuous output variable.

Regression

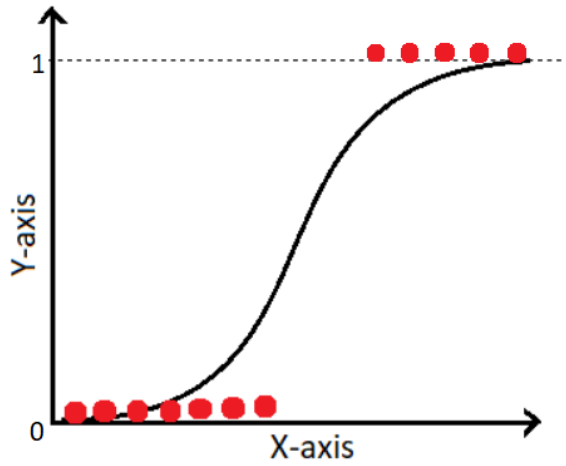
Linear Regression:

- the relationship between the input variables and the output variable is assumed to be linear.
- the goal is to find the best-fit line
- The line is represented by the equation `y = mx + b`, where `y` is the output variable, `x` is the input variable, `m` is the slope of the line, and `b` is the y-intercept.
- The slope `m` and the y-intercept `b` are estimated from the training data using an optimization algorithm.
- once the line is fitted to the training data, it can be used to make predictions on new, unseen data.



Logistic Regression:

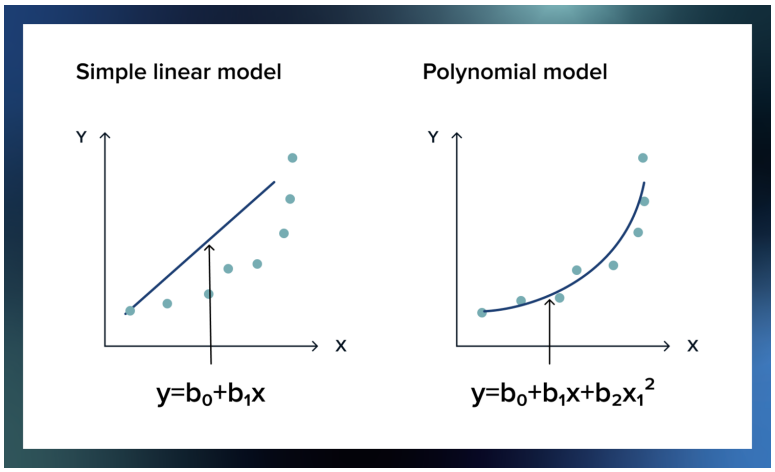
- the output variable is categorical and the relationship between the input variables and the output variable is assumed to be linear.
- used to predict the probability of an event occurring.
- the output is a value between 0 and 1, which can be interpreted as the probability of the event occurring.
- if the probability is greater than a certain threshold, the event is predicted to occur; otherwise, it is predicted not to occur.
- commonly used for binary classification problems, where the output variable has two possible values.



Polynomial Regression:

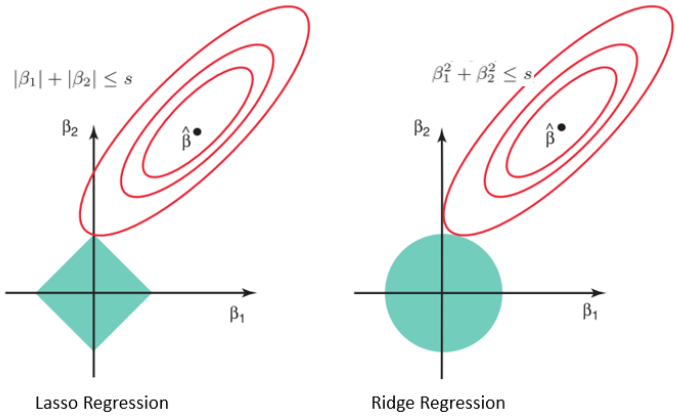
- the relationship between the input variables and the output variable is assumed to be a polynomial function.
- instead of fitting a straight line to the data, we fit a curve to the data.

- means that the equation that describes the relationship between the input variables and the output variable has higher-order terms, such as x^2, x^3, and so on.
- useful when the relationship between the input variables and the output variable is non-linear.
- can also be prone to overfitting if the degree of the polynomial is too high.



Ridge Regression:

- adds a penalty term to the regression equation to prevent overfitting of the data.
- the penalty term is a regularization term that penalizes large coefficients
- can help prevent the model from becoming too complex and overfitting the data.
- controlled by a hyperparameter called `alpha`, which is typically chosen using cross-validation.
- often used when there are many input variables, and some of them are highly correlated with each other.

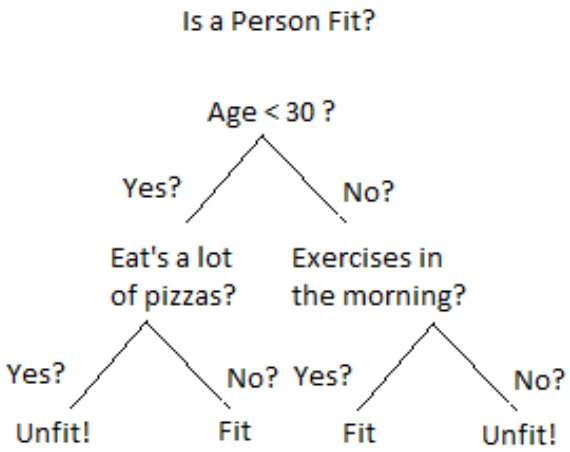


Lasso Regression:

- adds a penalty term to the regression equation to encourage sparsity in the coefficients.
- the goal is to fit a linear equation to a dataset by minimizing the sum of squared differences between the predicted and actual values.
- has the advantage of producing sparse models
- can force some coefficients to be exactly zero, while Ridge regression tends to shrink the coefficients towards zero
- makes lasso regression useful in situations where you want to explicitly identify the most important features.

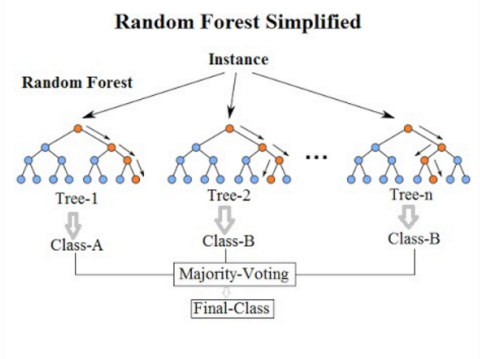
Decision Trees

- used to model decisions and their possible consequences.
- work by recursively partitioning the input space into smaller and smaller regions, based on the values of the input variables.
- chooses the input variable that best splits the data, based on some criterion, such as information gain or Gini impurity.



Random Forests

- ensemble method that combines multiple decision trees to improve the accuracy of the classification.
- Each decision tree in the random forest is trained on a random subset of the input data and a random subset of the input variables.
- The final classification is based on the majority vote of the individual decision trees.



Support Vector Machines (SVM)

- works by finding the hyperplane that best separates the input data into two classes.
- hyperplane is chosen to maximize the margin between the two classes, which is the distance between the hyperplane and the closest points from each class.

- the points that are closest to the hyperplane are called support vectors, hence the name Support Vector Machines.
- can be used for both classification and regression problems.
- can be extended to handle non-linearly separable data by using a kernel function to transform the data into a higher-dimensional space where it is linearly separable.
- popular kernel functions include the polynomial kernel and the radial basis function (RBF) kernel.
- can also be used for multi-class classification problems, by using a one-vs-all or one-vs-one approach.