Machine Learning

**Supervised Learning**:# uses labeled inputs (meaning the input has a corresponding output label) to train models and learn outputs

* # **Example**: Imagine a teacher supervising a class. The teacher already knows the correct answers but the learning process doesn't stop until the students learn the answers as well. This is the essence of Supervised Machine Learning Algorithms. Here, the algorithm learns from a training dataset and makes predictions that are compared with the actual output values. If the predictions are not correct, then the algorithm is modified until it is satisfactory. This learning process continues until the algorithm achieves the required level of performance. Then it can provide the desired output values for any new inputs

1. **Linear Regression Algorithm**:# The Linear Regression Algorithm provides teh relation between an independent and a dependent variable. It demonstrates the impact on the dependent variable when the independent variable is changed in any way. So the independent variable is called teh explanatory variable and the dependent variable called the factor of interest. An example of the Linear Regression Algorithm usage is to analyze the property prices in the area according to the size of the property, number of rooms, etc
2. **Logistic Regression Algorithm**:# The Logistic Regression Algorithm deals in discrete values wheras the Linear Regression Algorithm handles predictions in continuous values. This means that Logistic Regression is a better option for binary classification. An event in Logistic Regression is classified as 1 if it occurs and it is classified as 0 otherwise. Hence, the probability of a particular event occurrence is predicted based on the given predictor variables. An example of the Logistic Regression Algorithm usage is in medicine to predict if a person has malignant breast cancer tumors or not based on the size of the tumors
3. **Naive Bayes Classifier Algorithm**:# Naive Bayes Classifier Algorithm is used to classify data texts such as a web page, a document, an email, among other things. This algorithm is based on the Bayes Theorem of Probability and it allocates teh element value to a population from one of teh categories that are available. An example of teh Naive Bayes Classifier Algorithm usage is for Email Spam Filtering. Gmail uses this algorithm to classify an email as Spam or Not Spam.

* Bunch of inputs going into a model and single output
* A black square with white text

  Description automatically generatedInputs are called a Feature Vector
* **Features:**
  + **Qualitative:** Categorical data (finite number of categories or groups)
    - **Example:** Gender (M or F) 2 categories
    - **Example:** Nationality (US, France, Japan, Etc.)
  + Features can be Nominal Data (No inherent Order)
  + ONE-HOT Encoding: if it matches some category 🡪 make it a one, if not matches 🡪 make it a zero
    - A screen shot of a computer

      Description automatically generated
    - A screen shot of a black and white screen

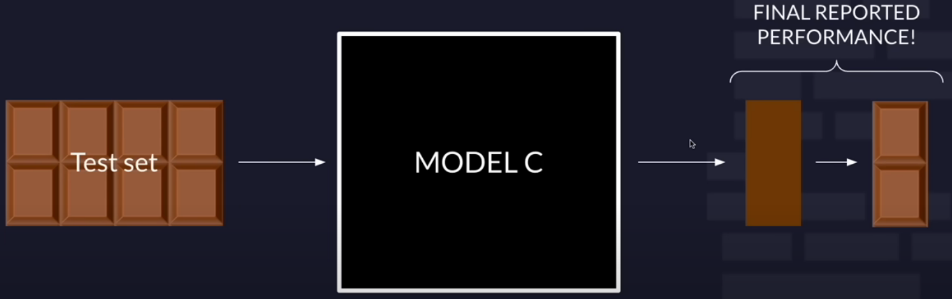
      Description automatically generated
  + Features can be ORDINAL DATA (Inherent Order) 🡪 baby | toddler | kid | teen | adult
    - 
  + Features can be Quantitative Data
    - Quantitative Data: numerical valued data (could be discrete or continuous)
      * Discrete: 1,2,3,4,etc. a solid number
      * Continuous: could be decimal
      * A blue flame on a stove

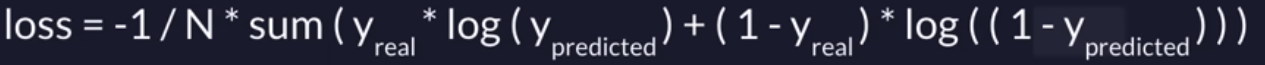
        Description automatically generated

**Supervised Learning Tasks:**

1. **Classification:** Predict Discrete Classes
   * 
   * 
   * **Other Examples of Classification:**
     1. **Binary Classification:** 
        1. Positive / Negative
        2. Cat / Dog
        3. Spam / Not Spam
     2. **Multiclass Classification:**
        1. Cat / Dog / Lizard / Dolphin
        2. Orange / Apple / Pear
        3. Plant species
2. **Regression: Predict Continuous Values**
   * 
   * Usually a number that is on a scale 🡪 basically floats
3. Note:
   * Each row in a dataset is a sample of the dataset
   * Each column is a feature of the dataset
     1. Except the output label, usually the determinant or what we trying to figure out
   * Feature vector is basically a row with all columns except the output label column
   * Target is the value of the output label on that specified row
   * Features Matrix is X: all of the rows with the columns included except output column
   * Labels / Target Vectors / Y: the entire output column
   * A screenshot of a computer

     Description automatically generated
   * What we do is we compare the prediction to the actual value of y that we have in our label data set, because that’s the whole point of supervised learning is we can compare what our model is outputting to. Then we can go back and we can adjust some things, so that the next iteration can eventually get closer to the true value
   * It’s not necessary a good thing to use the entire data set for training as it would only be proficient on that data and would be poor at new data or generalized data in whole
     1. Split Dataset into: Training Dataset | Validation Dataset | Testing Dataset
     2. Could be 60/20/20 or 80/10/10
     3. **Loss**: The difference between the our prediction and the true values
        1. Training is making adjustments to lower this “Loss” number/value
     4. Validation set used as a reality check during/after training to ensure model can handle unseen data. (Loss data is not feed back into the model)
   * A screenshot of a video game

     Description automatically generated
   * Test set used as to check how generalizable the final chosen model is
   * 
4. **Metrics of Performance**
   * **Loss:** Difference between prediction and true value
     1. **L1 Loss:** Take absolute value of (real value – predicted value)
     2. **L2 Loss:** Takes squared value of (real value – predicted value)
     3. A screenshot of a computer screen

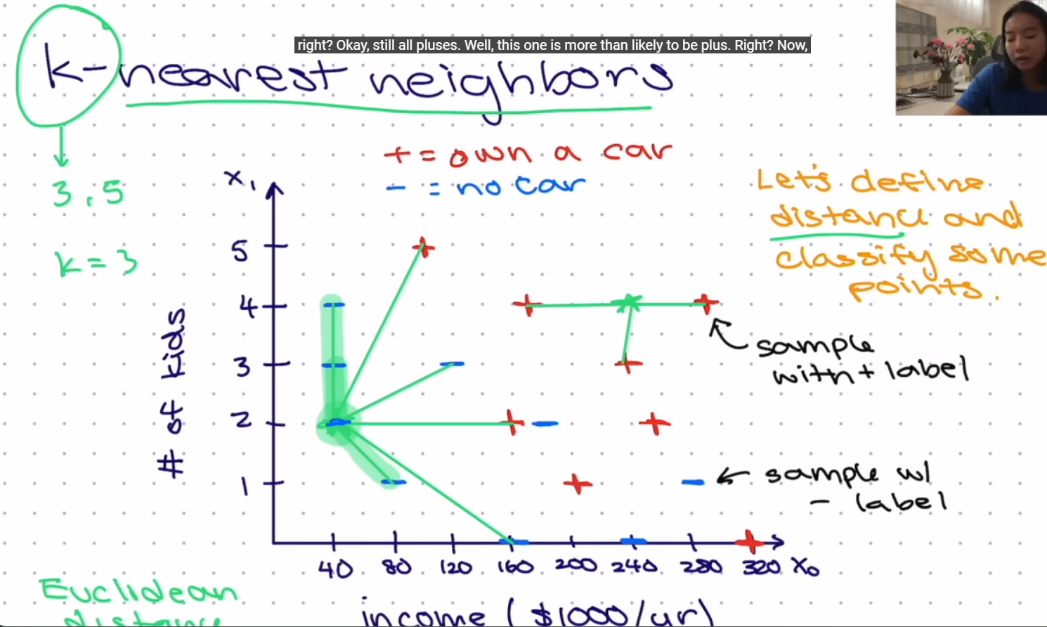
        Description automatically generated
     4. **Binary Cross-Entropy Loss**
        1. Loss decreases as the performance gets better
        2. 
   * **Accuracy:**
     1. A screenshot of a computer

        Description automatically generated

**Unsupervised Learning**:# uses unlabeled data to learn about patterns in data

* **# Example**: In this case, there is no teacher for the class and the students are left to learn for themselves. So for Unsupervised Machine Learning Algorithms, there is no specific answer to be learned and there is no teacher. In this way, the algorithm doesn't figure out any output for input but it explores the data. The algorithm is left unsupervised to find the underlying structure in the data in order to learn more and more about the data itself (Grouping/Clustering)

1. **K Means Clustering Algorithm**: # Let's imagine that you want to search the name "Harry" on Wikipedia. Now, "Harry" can refer to Harry Potter, Prince Harry of England, or any other popular Harry. So Wikipedia groups the web pages that talk about the same ideas using K Means Clustering Algorithm (since it is a popular algorithm for cluster analysis). K Means Clustering Algorithm in general uses K number of clusters to operate on a given data set. In this manner, the output contains K clusters with the input data partitioned among the clusters
   1. Basically takes the label of whatever group your more aligned with
   2. Euclidean distance: length of straight line
      1. A white board with green writing on it

         Description automatically generated
   3. K: how many neighbors do we look at to determine the aligned cluster for a element to be predicted
   4. 
2. **Apriori Algorthm**: # The Apriori Algorithm uses teh if-then format to create association rules. This means that if a certain event 1 occurs, then there is a high probability that a certain event 2 also occurs. For example: IF someone buys a car, THEN there is a high chance they buy car insurance as well. The Apriori Algorithm generates this association rule by observing the number of people who bought car insurance after buying a car. For example, Google auto-complete uses the Apriori Algorithm. When a word is typed in Google, the Apriori Algorithm looks for the associated words that are usually typed after that word and displays the possibilities

**Reinforcement learning**:# Agent learning in interactive environment based on rewards and penalties

* # **Example**: Here the hypothetical students who learn from their own mistakes over time (similar to life). So the Reinforcement Machine Learning Algorithms learn optimal actions through trial and error. This means that the algorithm decides the next action by learning behaviors that are based on its current state and that will maximize the reward in the future. This is done using forward feedback that allows the Reinforcement Algorithm to learn which are the best behaviors that lead to maximum reward. This reward feedback is known as a reinforcement signal

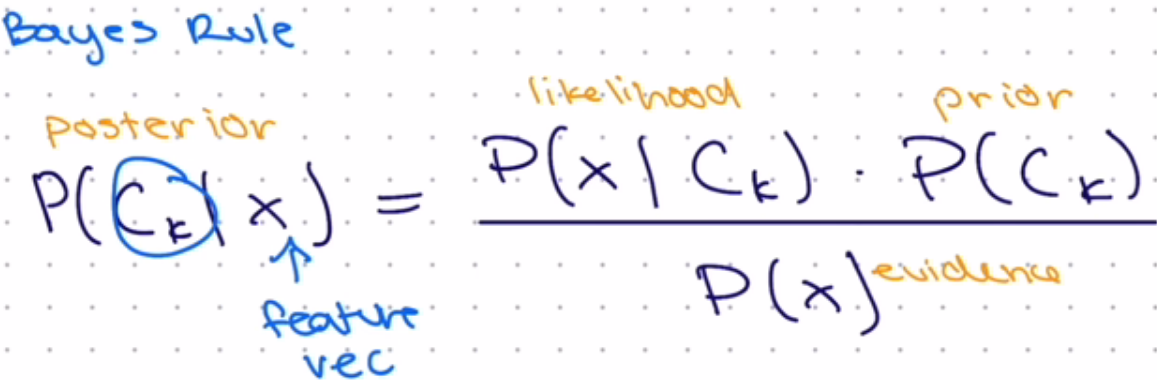
**Naïve Bayes**

**A screenshot of a test results

Description automatically generated**

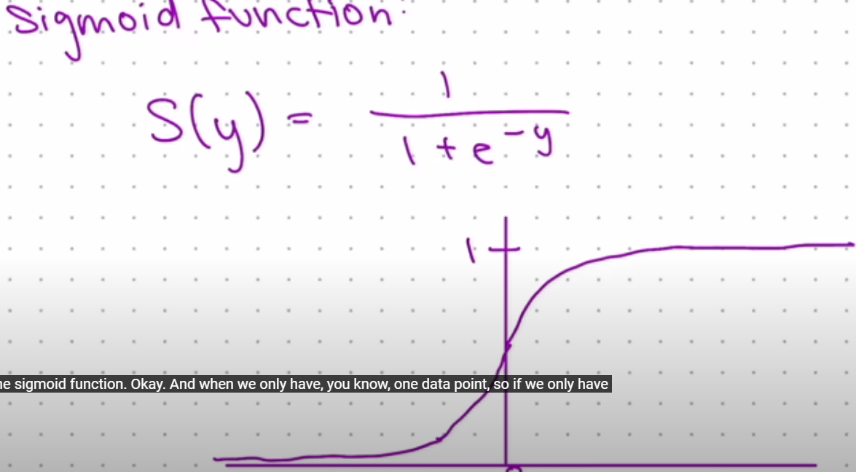
* What is probability of having covid given a positive test:
  + P( covid / +test ) = 531/551 🡪 96.4%
* **Bayes’ Rule**
  + Probability A given B
  + **P( A|B ) = [ P(B|A) \* P(A) ] / P(B)**
  + P(false positive) = 0.05
  + P(false negative) = 0.01
  + P(disease) = 0.1
  + P(disease|(+)test) = ?
  + **A math test with numbers and symbols

    Description automatically generated with medium confidence**
  + **A close-up of a math test

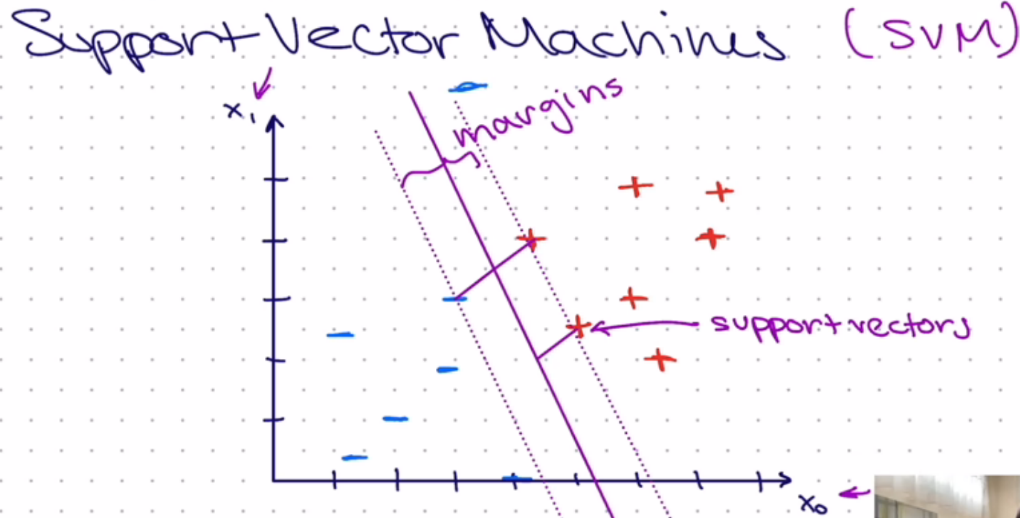
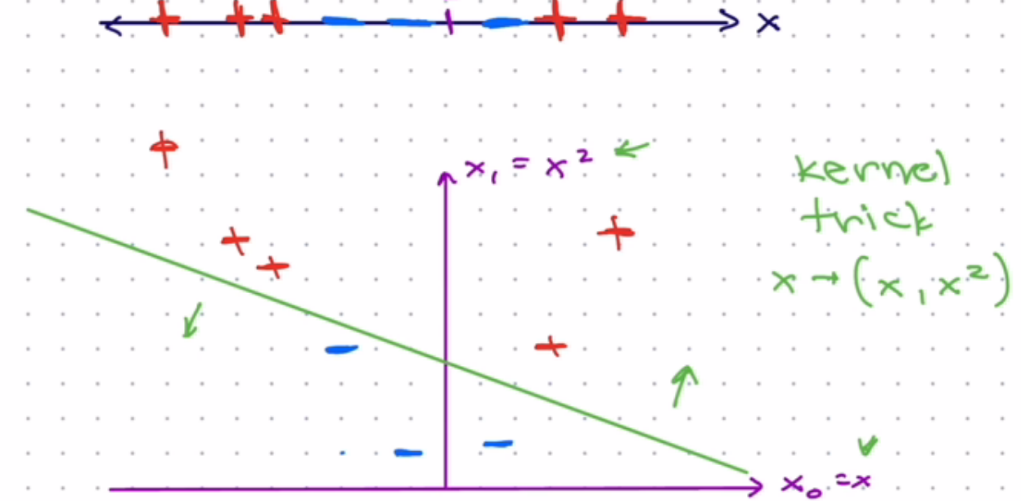
    Description automatically generated**
  + **C = categories / clusters**
  + ****

**Logistic Regression**

* **A graph on a white sheet

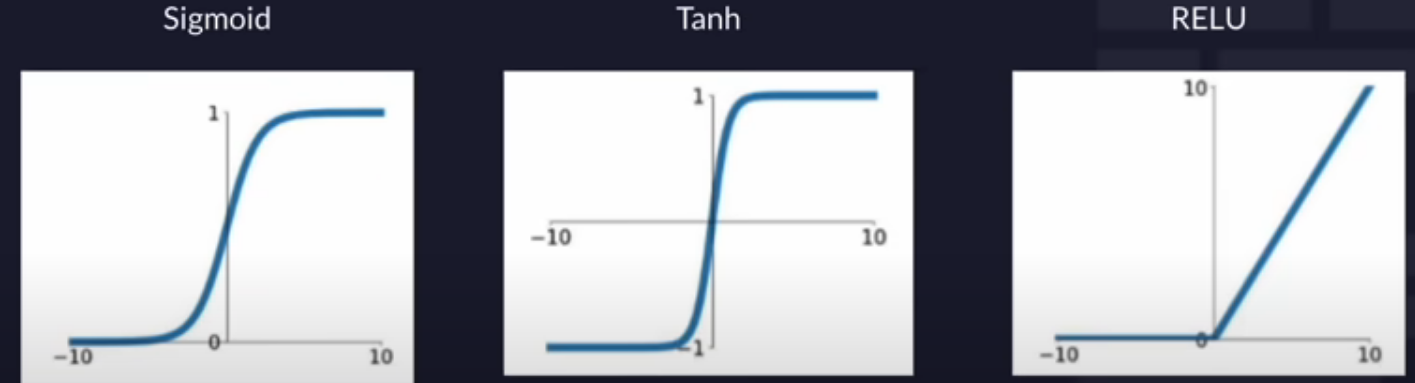
  Description automatically generated**
* ****
* **Simple Logistic Regression:** only 1 feature
* **Multiple Logistic Regression:** multiple features

**Support Vector Machines (SVM)**

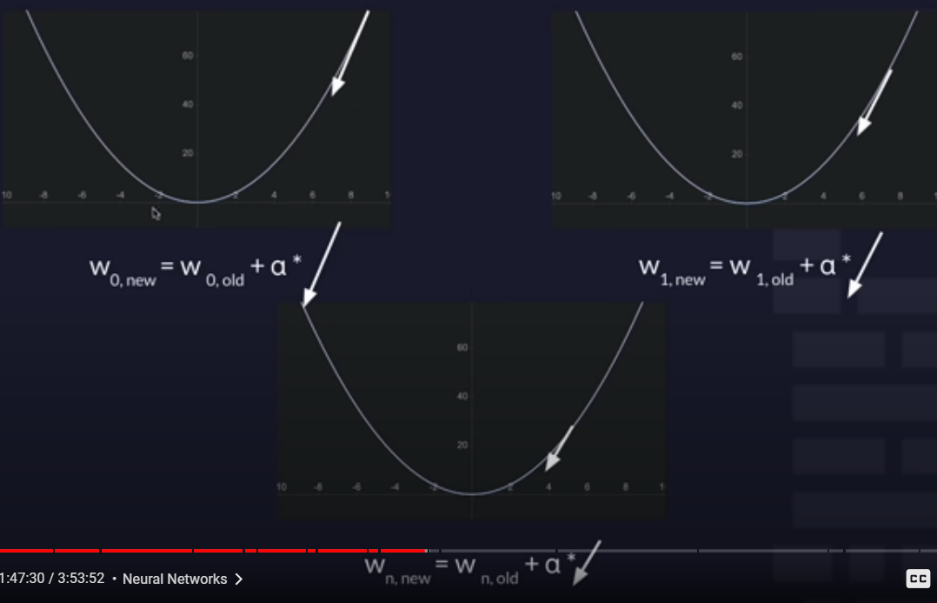
* Best Divider
* ****
* ****

**Neural Network:**

* **Artificial Neural Networks** are modeled after the neurons in the human brain. They contain artificial neurons which are called units. These units are arranged in a series of layers that together constitute the whole Artificial Neural Networks in a system. A layer can have only a dozen units or millions of units as this depends on the complexity of the system. Commonly, Artificial Neural Networks have an input layer, output layer as well as hidden layers. The input layer receives data from the outside world which the neural network needs to analyze or learn about. Then this data passes through one or multiple hidden layers that transform the input data that is valuable for the output layer. Finally, the output layer provides an output in the form of a response of the Artificial Neural Network to input data provided.
* **A diagram of a network

  Description automatically generated**
* **Activation Function Necessary or else it becomes a linear model**
* ****
* To what extent is this value(weight) contributing to our loss function
* A screenshot of a graph

  Description automatically generated
* Take old weight and decrease some more by some multitude alpha
* A graph of a function

  Description automatically generated
* Arrow is negative gradient
* 

**TensorFlow**

* Sequential: go from one layer to the next
* Dense: all nodes are interconnected