Machine learining for every body

Source

one hot encoding

[USA,India,Canda,France]

Country	Encoding
USA	[1,0,0,0]
Canada	[0,0,1,0]
India	[0,1,0,0]
France	[0,0,0,1]

Qualitative

- categorical data (finite number of categories or groups).
- Ordinal data (Inherent order)

Quantative

- numberical valued data (could be discrete of continious)

What are the predictions that our model can ouput ?

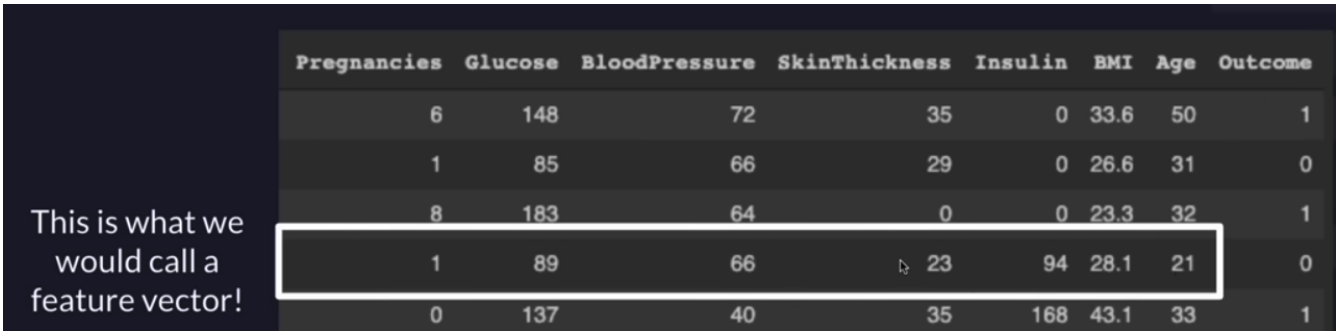
Supervised Learning tasks

1. Classification - Predict discrete classes

Binary classification	Multiclass classification
Positive/negative	cat/dog/lizard/dolpin
cat/dog	orange/apple/pear
Spam/notspam	plant species

2. Regression - predict continious values

- price of etherium
- temperation
- price of realestate
  predict future based on historic values of the features (continious)



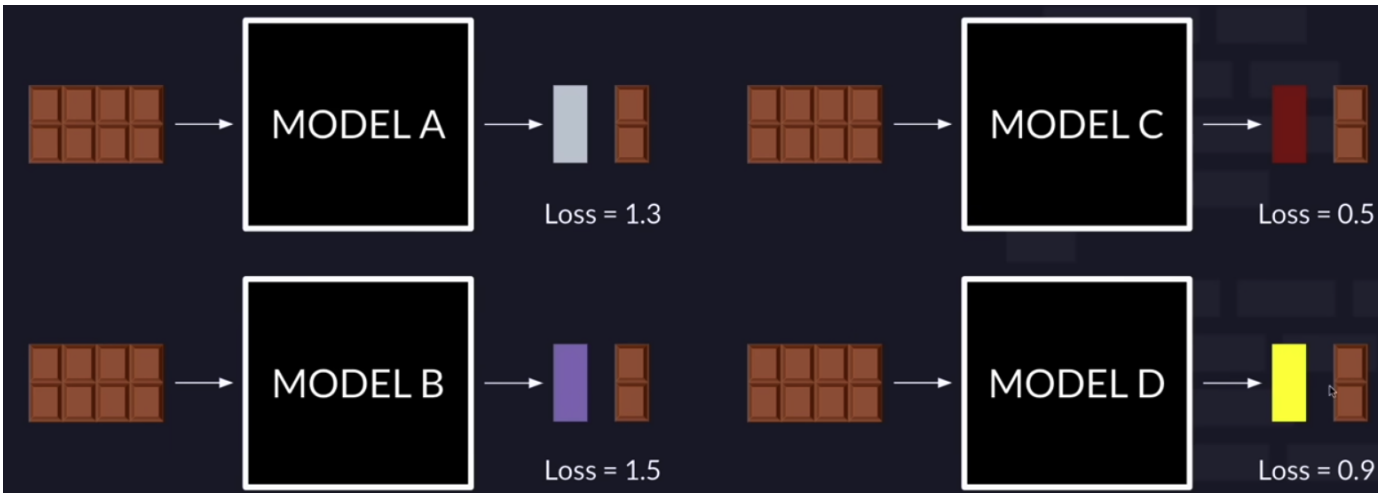
> Above image shows, All features of one sample/data set

Divide data into three

- Training dataset
- Validation dataset (used as reality check during/after training to ensure model can handle unseen data)
- Testing dataset

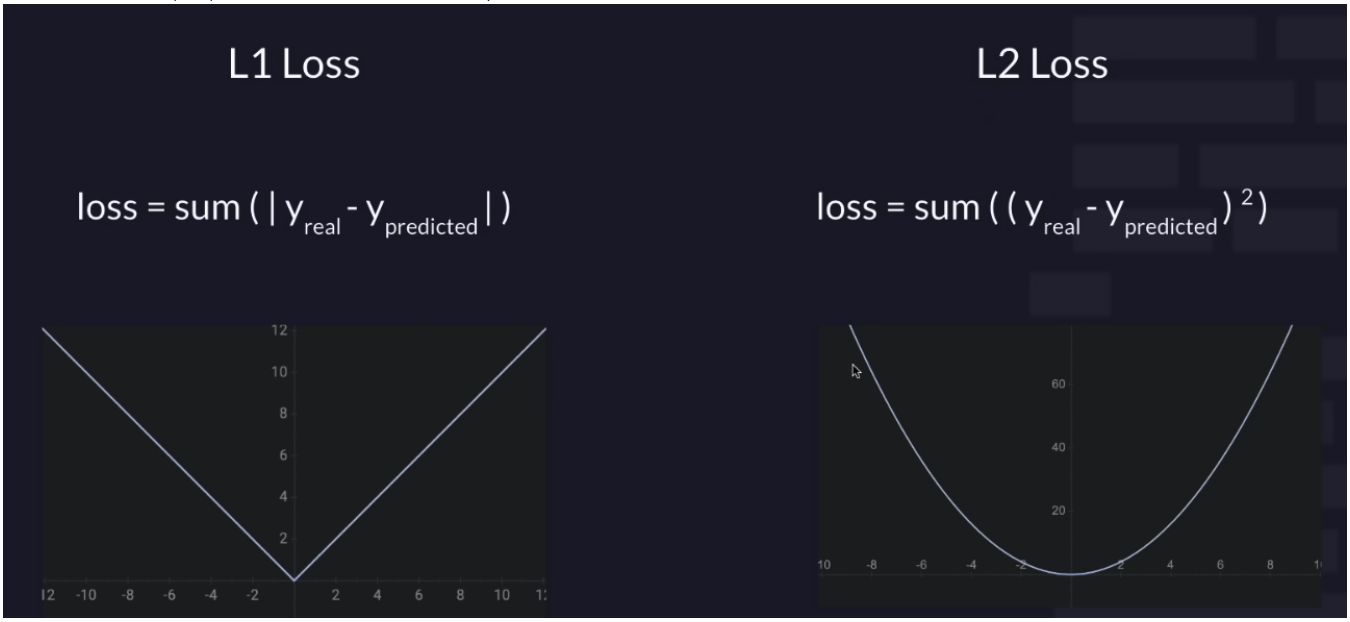
Loss :

Model which has less loss is better performed



Loss funtions

Far the value is from 0(zero) over x-axis the more the loss is the poor the model is



Loss decreases as performance increasees

Binary Crossentropy is the loss function used when there is a classification problem between 2 categories only.

It is self-explanatory from the name *Binary,* It means 2 quantities, which is why it is constructed in a way that fits the problem of classification of 2 quantities.

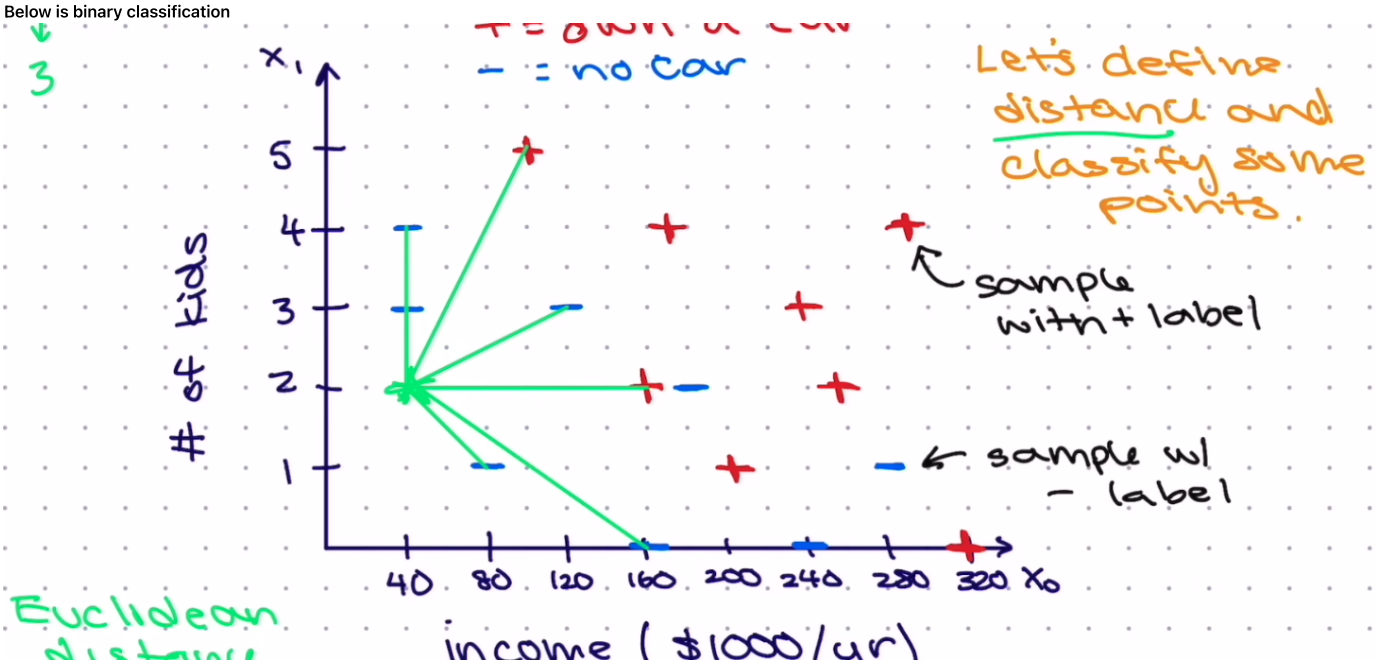
Classification models

- knn
- naive
- Logistic regression
- svm

K-nearest neighbors Algo

In this example given below we predict if given a person will have car or no. '+' indicates that person has car, '-' indicates that person doesnot have car.(mapping is from data set). Now we have to predict given a person with income range and children how likely is he can have a car.

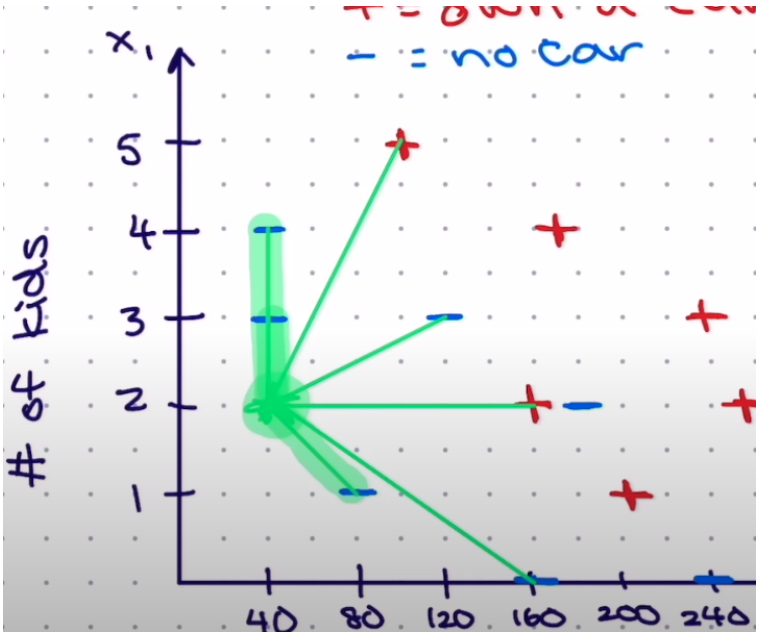
Below is binary classification



K- nearest neighbour algo

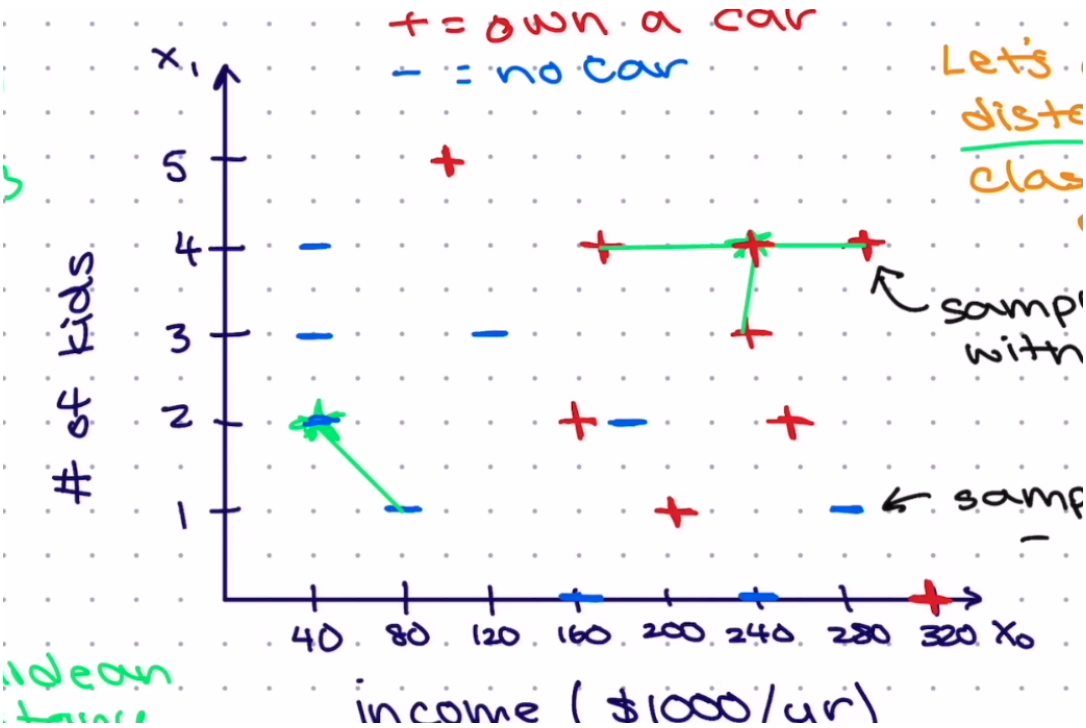
k -> Tell us how many neighbours we are use in order to judge what the label is (usually 3,5)

If k=3, Below image shows what three neighbours we can take. All those points are blue. Chances are prediction is no car

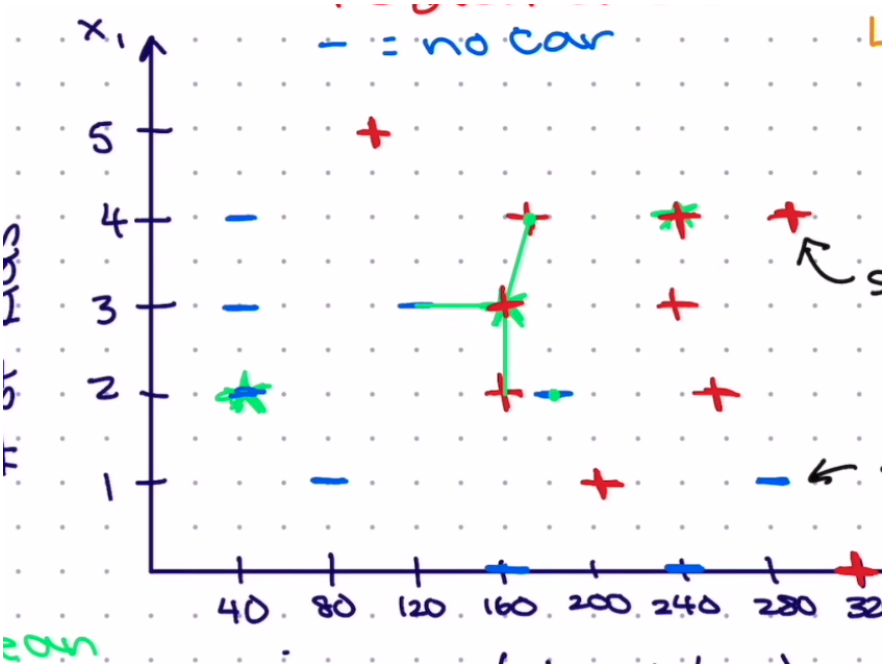


Prediction for -> What if a person has 4 kids and earns around 240k per year.

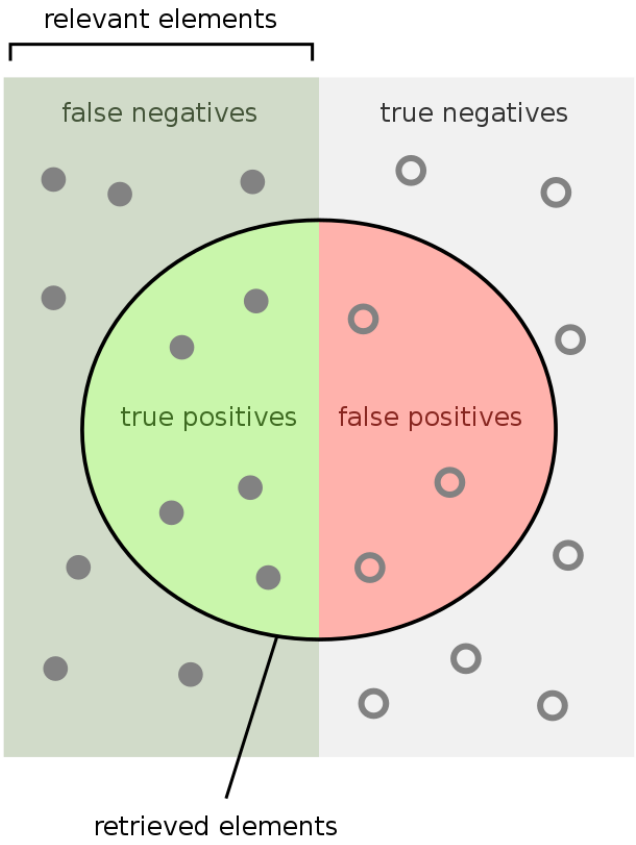
From below image, its more likely that a person will have car



Below image is sarrounded by ''+'' and ''-'' (red and blue), But two ''+'' are close to the point, hence conclusion is that the personn might have car



Presision and recall



How many retrieved items are relevant?

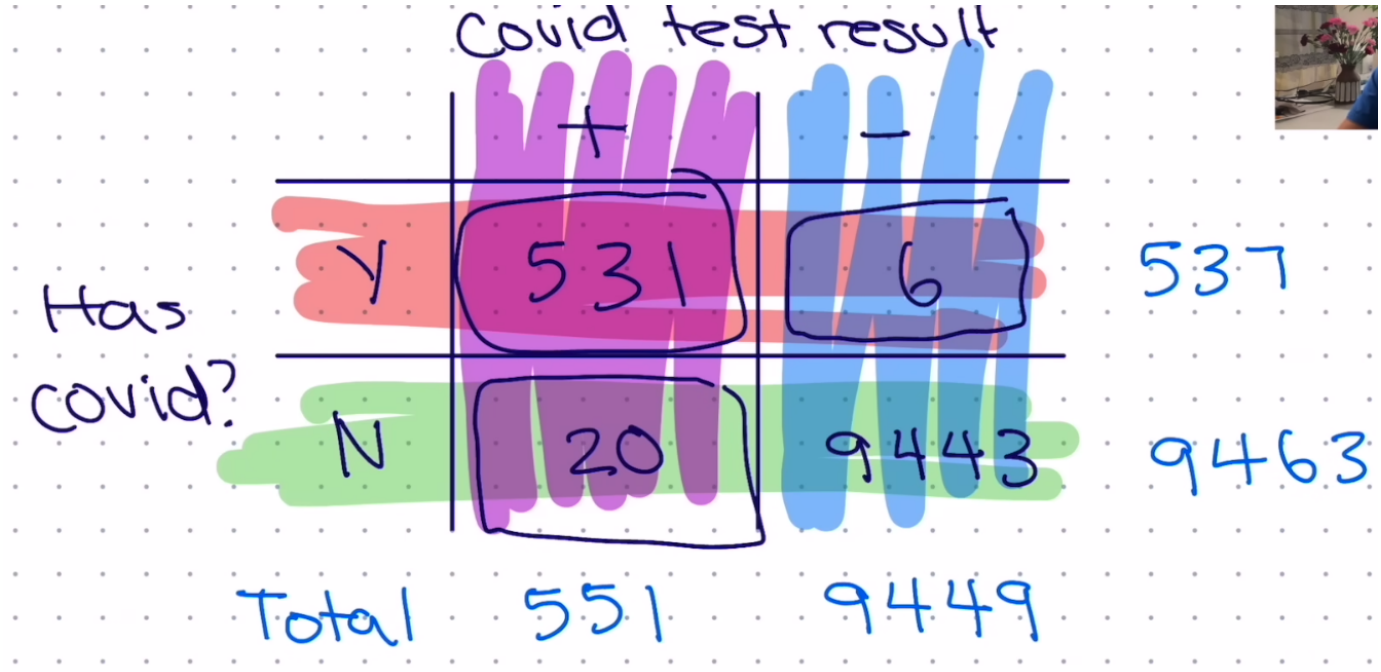
$$\text{Precision} = \frac{\text{true positives}}{\text{retrieved elements}}$$

How many relevant items are retrieved?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

Presision -> Out of all the ones we labeled positives how many are true positives
Recall -> Out of all the ones we know truly positive how many do we actually get right

Bayes formula



What is the probability of having covid given a positive test?

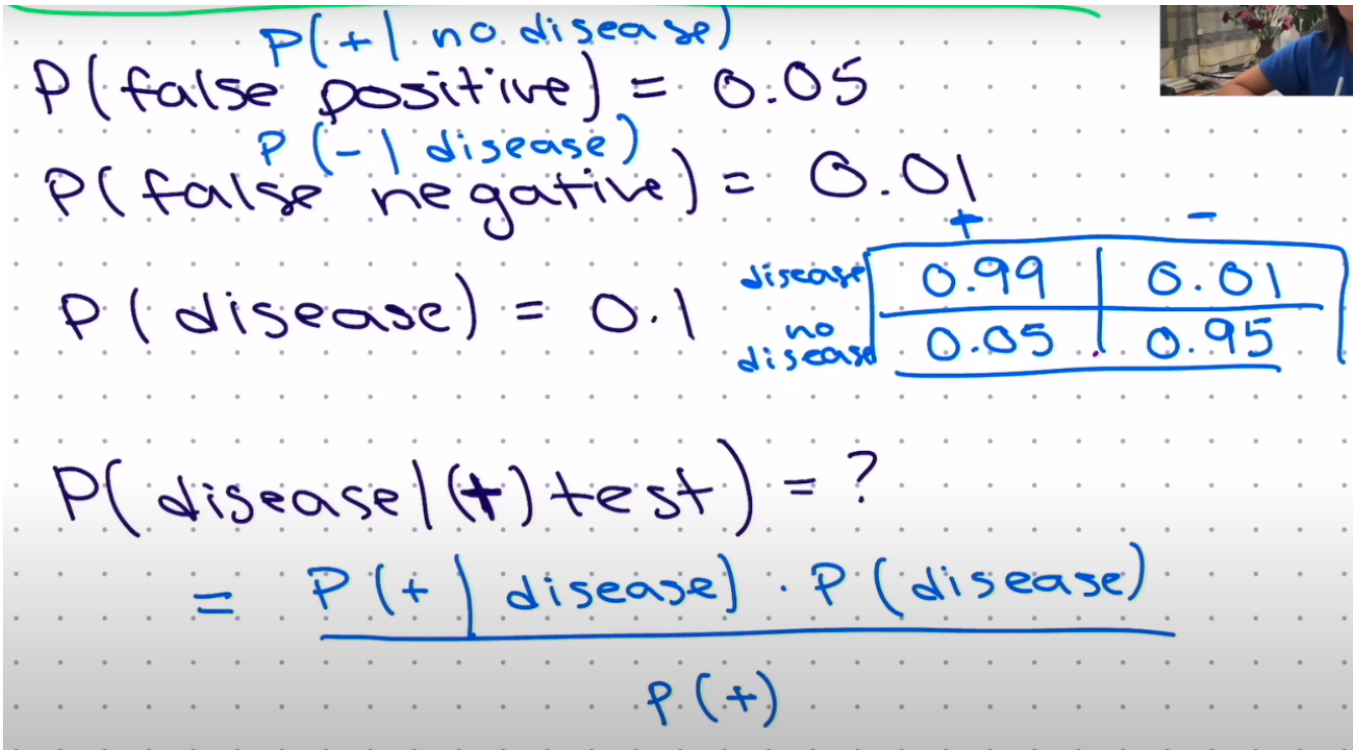
P(Covid | + test) =531/551 => 96.4%

> | -> given that

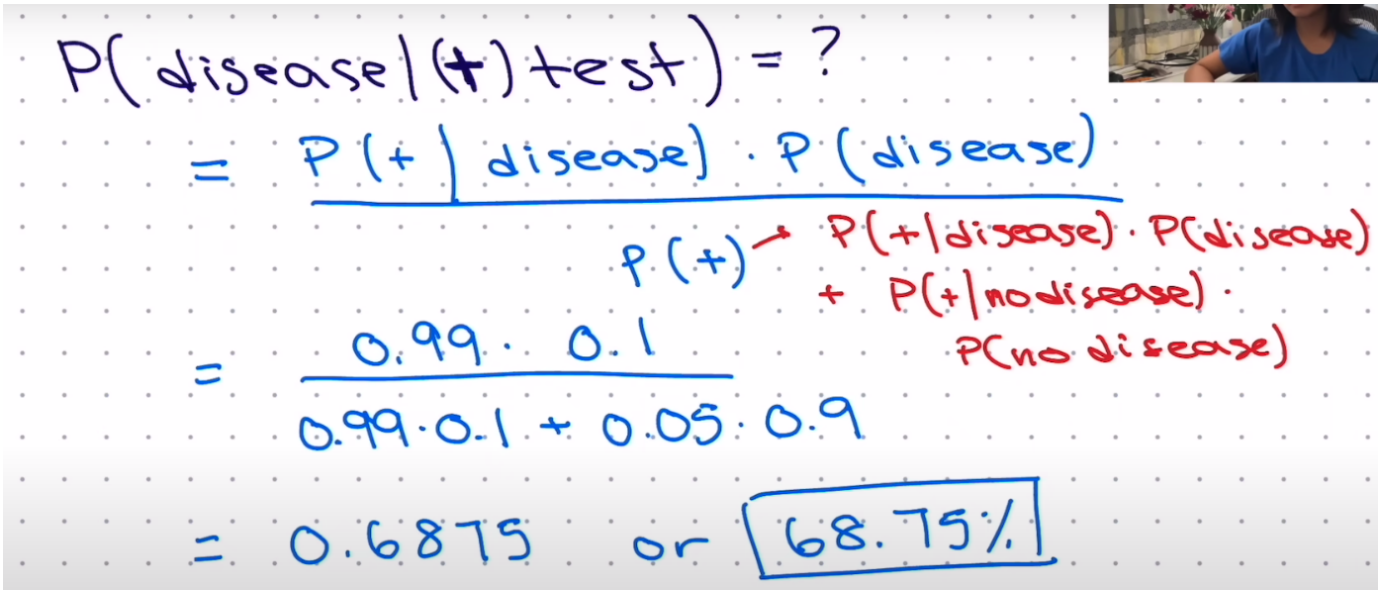
Bayes Formula

```
P(A|B)=P(A∩B)/P(B)=P(A)·P(B|A)/P(B)
where:
P(A)= The probability of A occurring
P(B)= The probability of B occurring
P(A|B)=The probability of A given B
P(B|A)= The probability of B given A
P(A∩B))= The probability of both A and B occurring
```

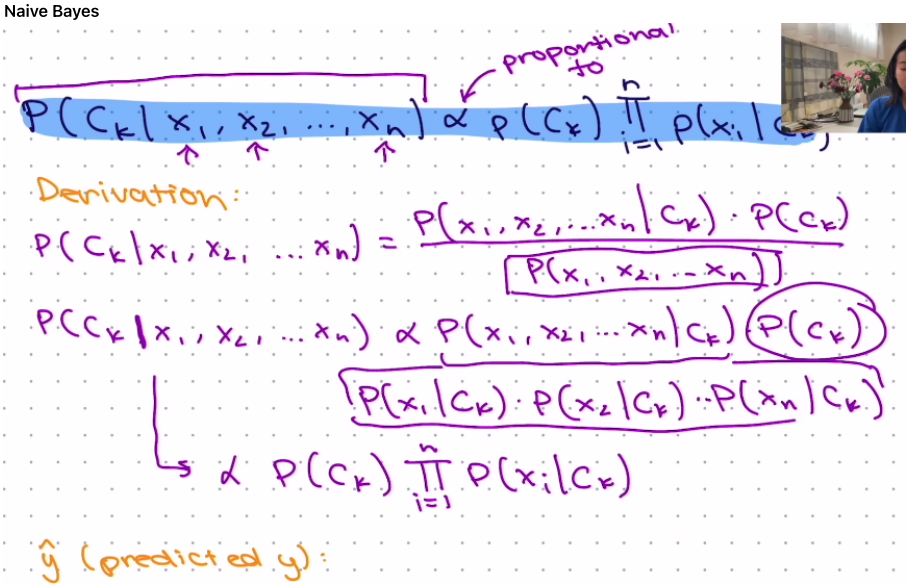
From above example : P(Covid | + test) == P(A|B)
Probability will sum up to 1



$$
\begin{aligned}
P(\text{User} \mid +) &= \frac{P(+ \mid \text{User})P(\text{User})}{P(+)} \\
&= \frac{P(+ \mid \text{User})P(\text{User})}{P(+ \mid \text{User})P(\text{User}) + P(+ \mid \text{Non-user})P(\text{Non-user})} \\
&= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\
&\approx 33.2\%
\end{aligned}
$$

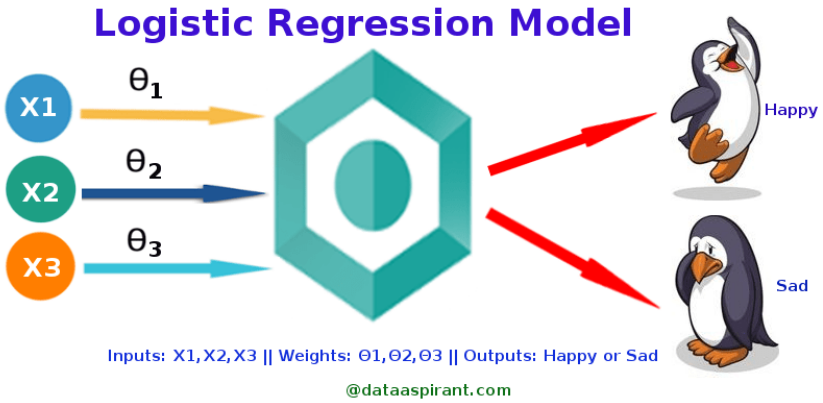


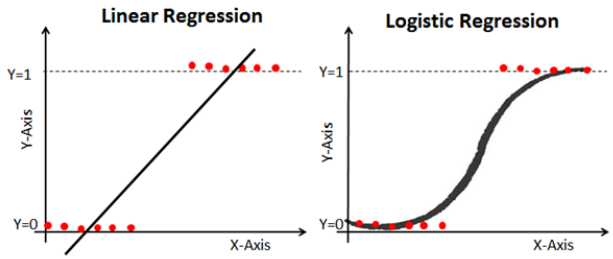
Naive Bayes



Logistic Regression:

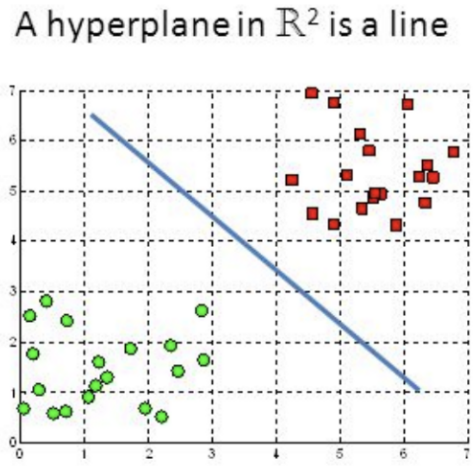
- Classify using regression
- we will try to fit data in to sinmoid function so the probability is between 0 and 1

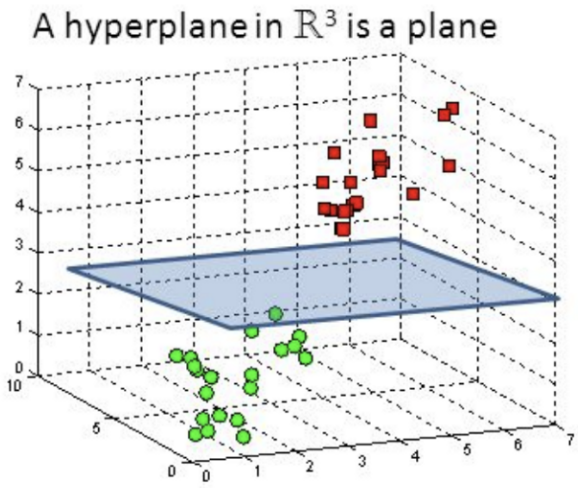




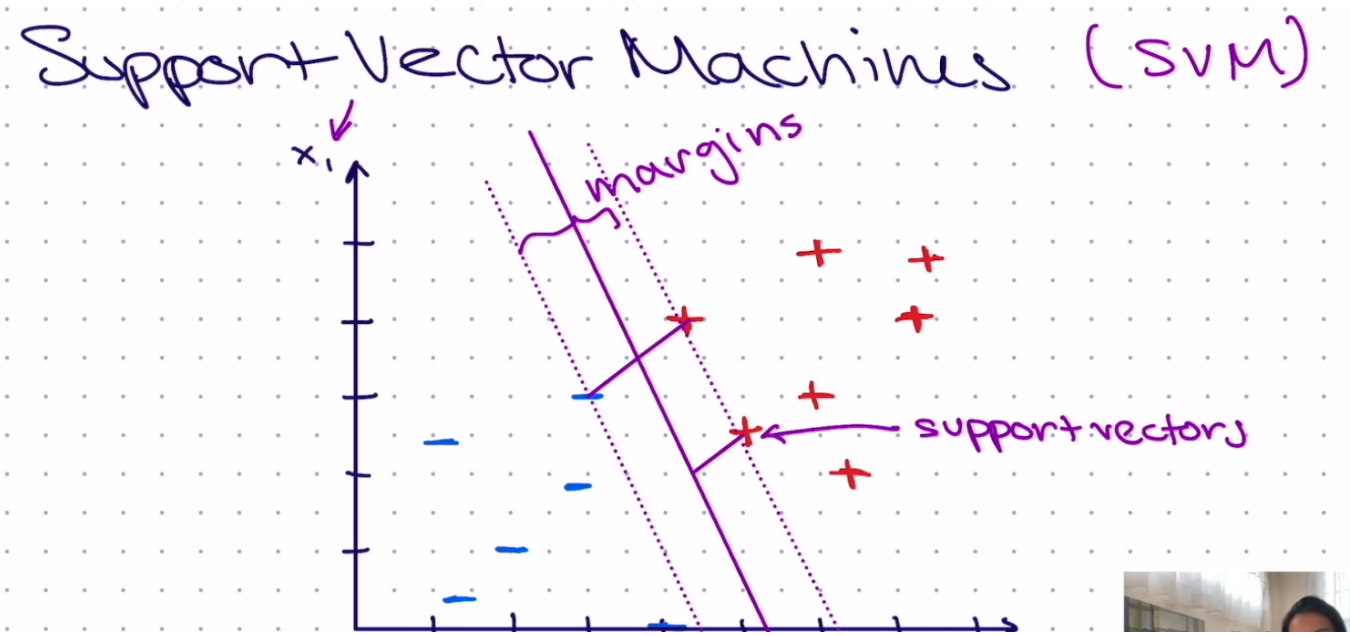
Support Vector Machine - SVM

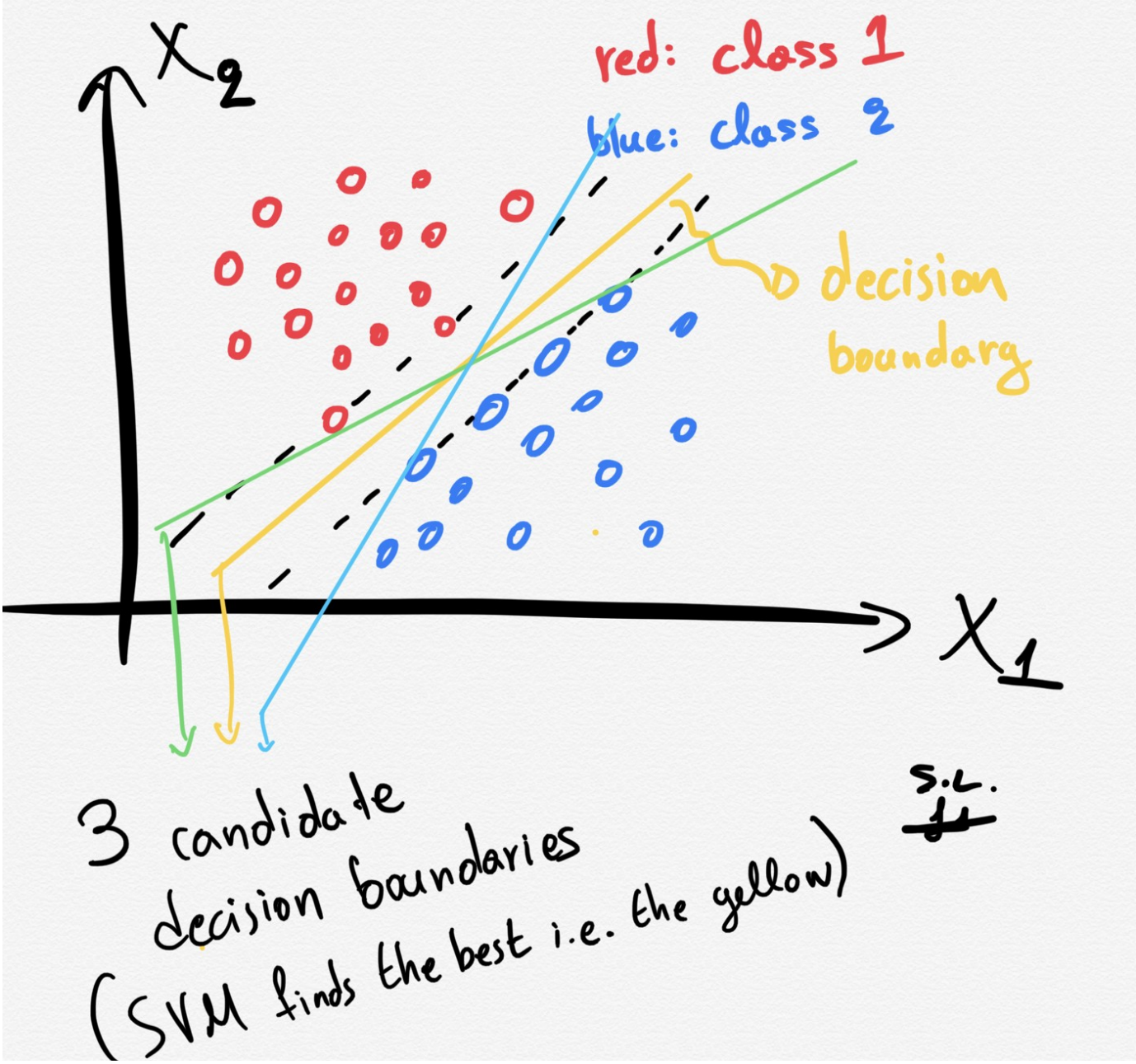
- In 2d it looks like a line dividing but in 3d its a plane that divides data





In SVM its not just classifing data, we also look at the boundry(margin) of the classifying line. (It should not be close to one type of data)

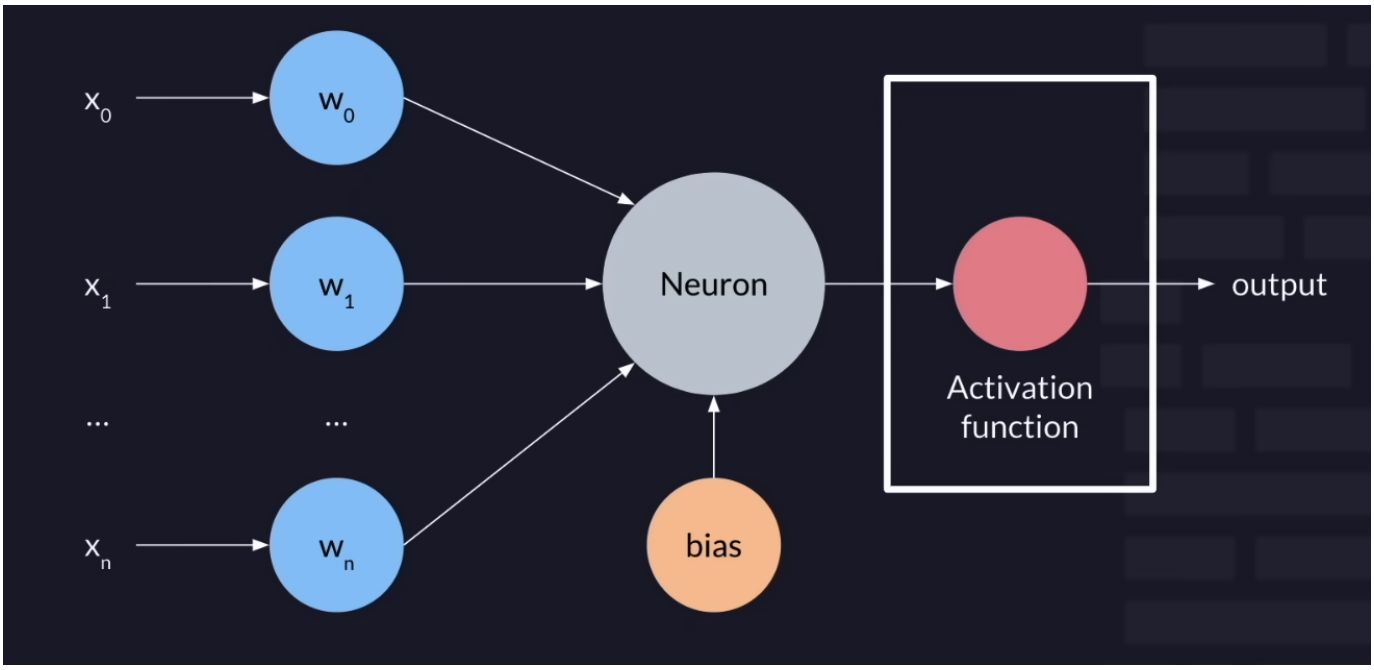




Kernel trick -> x to (x,x^2)

Neural Network

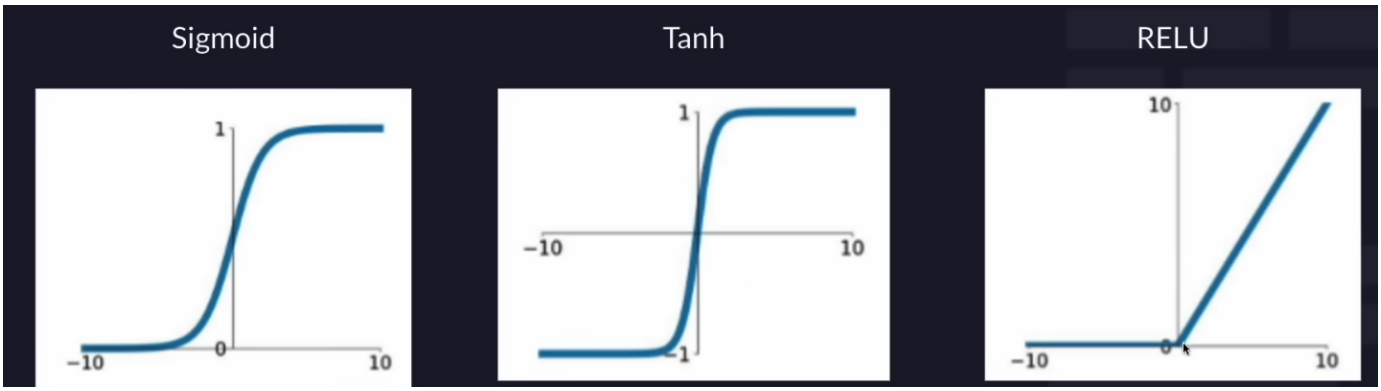
Classification using Neural network
A neural network is **a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates**



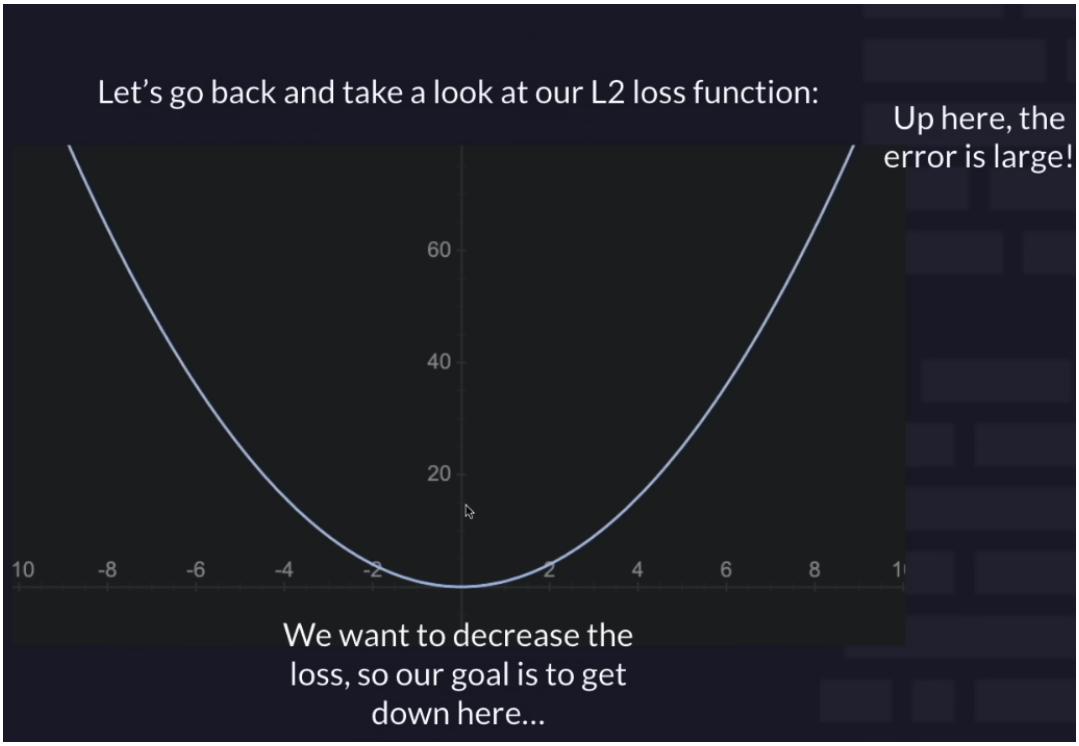
> With out activation funtion neuralnetwork is like linear combination of models

Activation Funtion

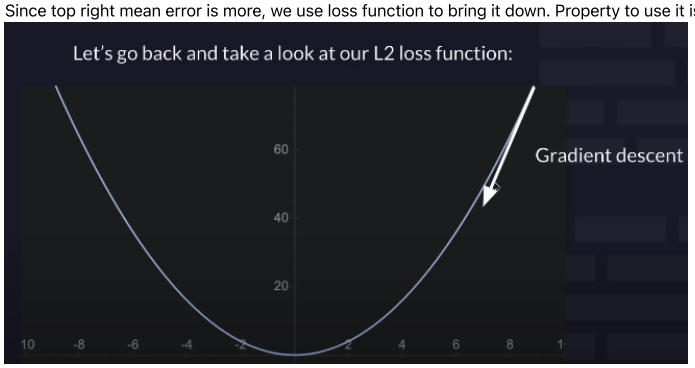
Below are activation function which is not linear functions



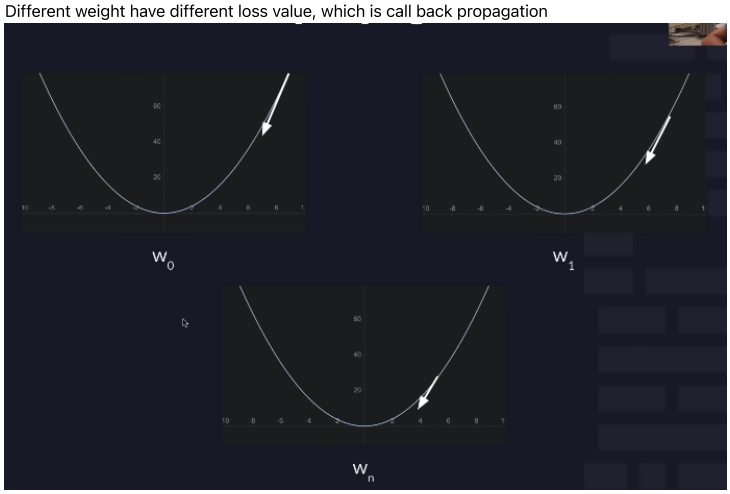
Lossfunction

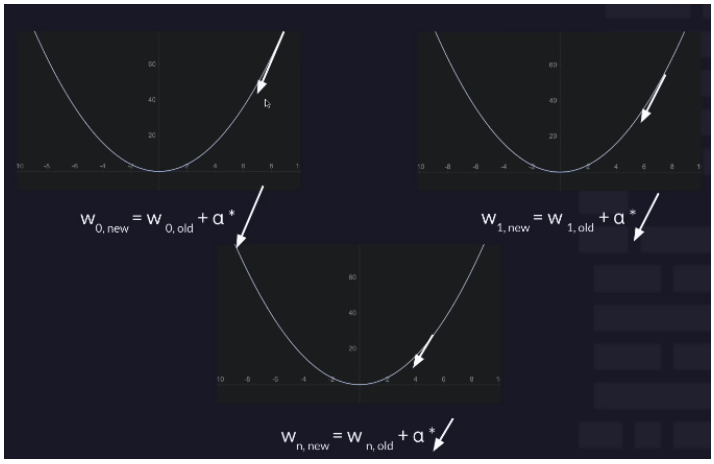


Since top right mean error is more, we use loss function to bring it down. Property to use it is call gradient decent



Different weight have different loss value, which is call back propagation



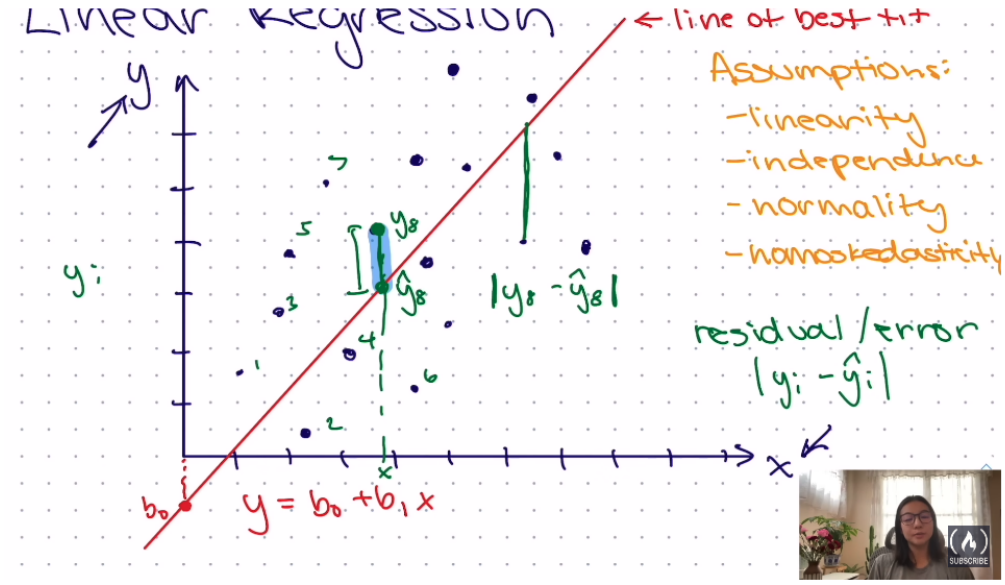


Ternsor flow Is predefined opensource code for neural network. Its a open source library to build models

Linear Regression

Simple linear regression

y= b0+b1X



Minimisising error means minimising residual values. So we need to sum up all residual value and get the lowest sum.
Multiple linear regression
y= b0+b1X1+b2X2+b3X3

Evaluating linear regression model

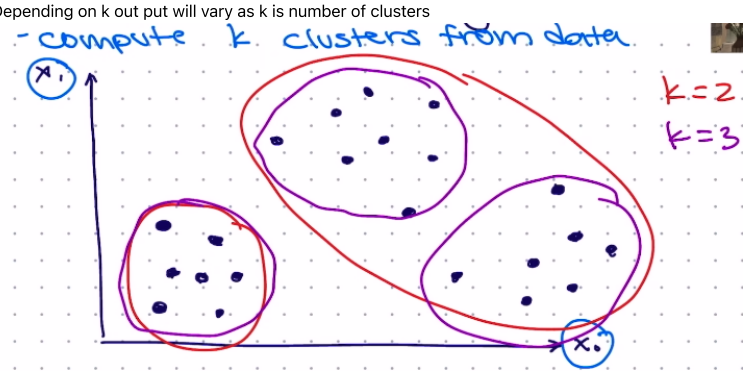
1. Mean Absolute Error (MAE)
2. Mean Squared Error (MSE)
3. Root mean square error (RMSE)
4. Co efficient of Determination (R^2)

We can use neural net (tensor flow keras) and calculate linear regression. Difference is in NN we are using back propogation, in regular linear regression we dont have back propogation

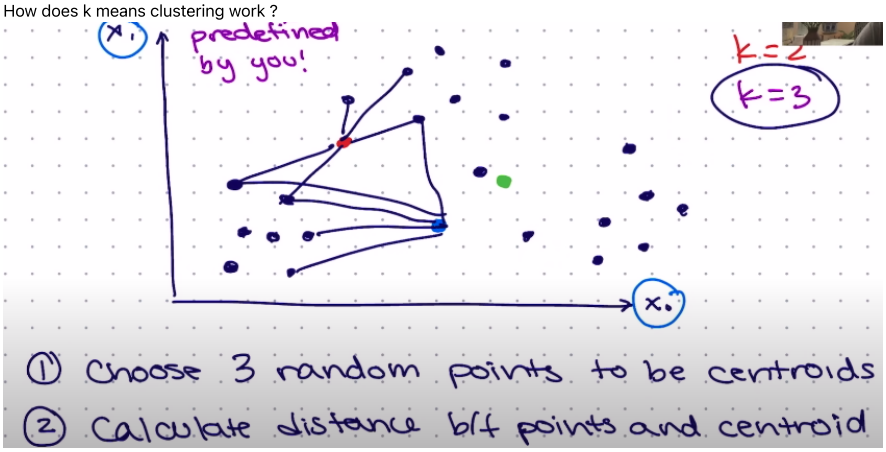
Unsupervised Learning

K-Means clustering

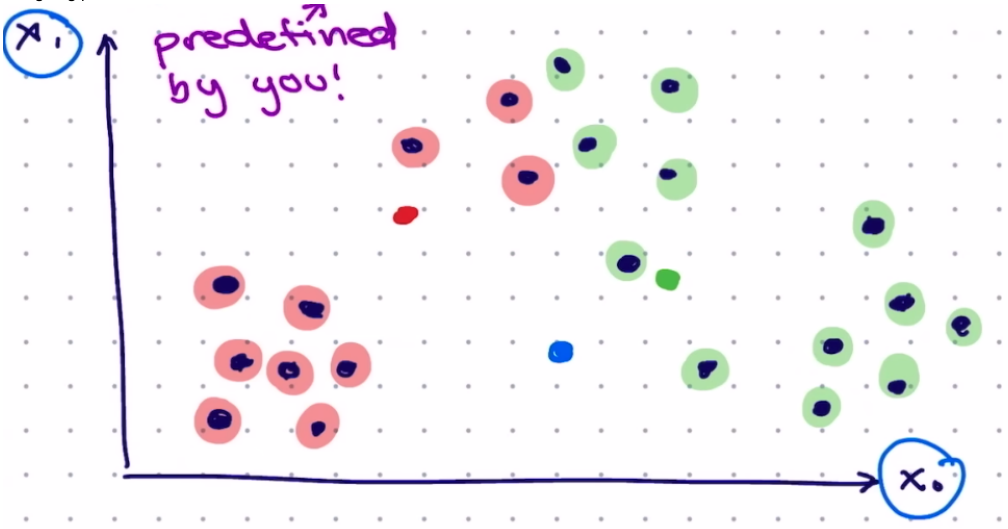
- unsupervised learning
- Compute k clusters from data
- Depending on k out put will vary as k is number of clusters



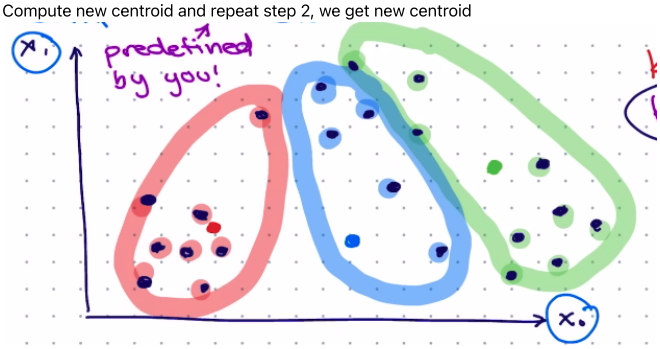
How does k means clustering work ?



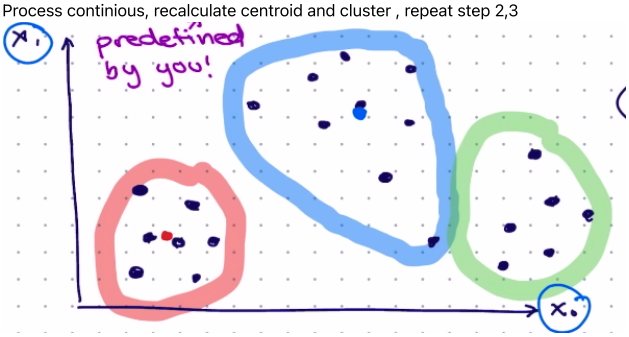
3. Assigning points to closest centroid



4. Compute new centroid and repeat step 2, we get new centroid



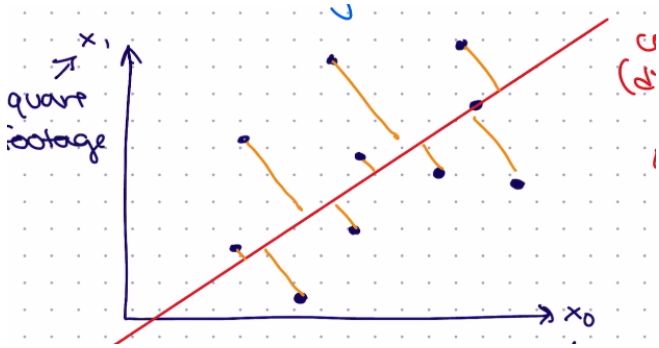
5. Process continious, recalculate centroid and cluster , repeat step 2,3



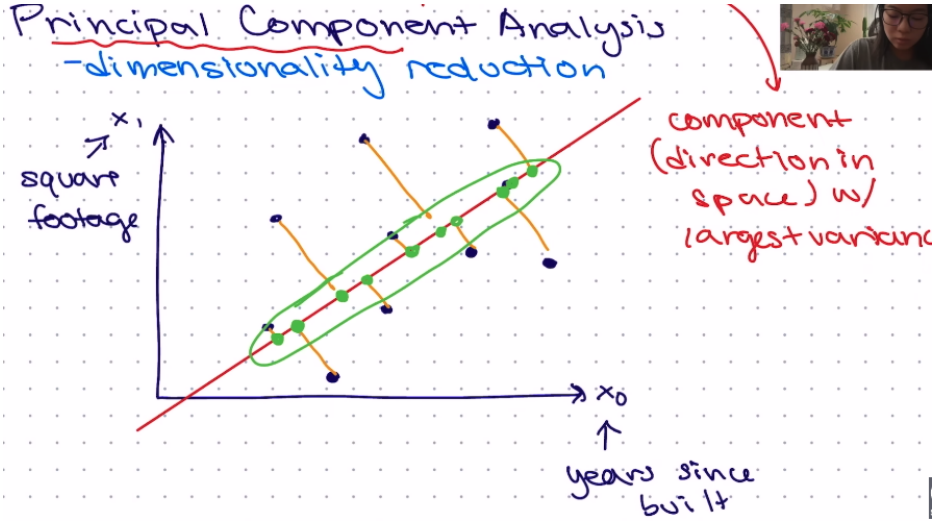
We stop when the points are stable and nothing is changing (centroid and cluster are same). This process is call expectation maximisation

Principal component Analysis(PCA)

- Used for dimentsionality reduction
- Technique used in Unsupervised learning
- Definition: Component direction in space with largest variance
- Looks similar to liniear regression but not, In this we don't have y component

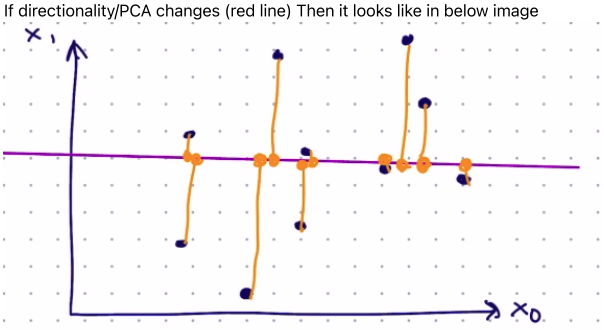


- We started with 2d data set and end with 1d

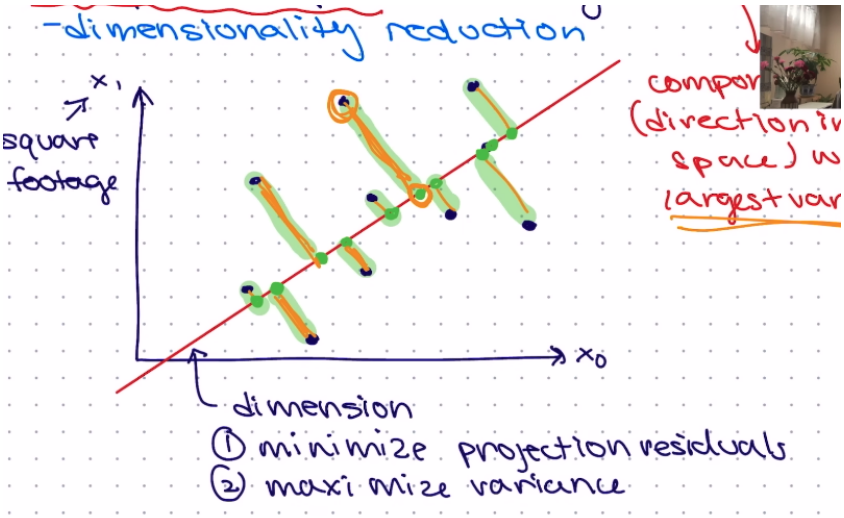


From above picture, we poloted blue ones(data set) in two dimensional graph. We calculated green dots which is one dimensional data set, Red line is the dimensional (directionality). Green dots are the one number to represent two dimensional points (blue

If directionality/PCA changes (red line) Then it looks like in below image



> In Linear regression we find the distance between y and x components, But in PCA we find the distance between x and pojected point (dot on red line)



If we have 100 different features and then we need to take 5 features , we can calculate PCA and take top ones. Example if we have sphere and we have 2d area to plot , what is the best way to represent sphere ?? its circle. (Extract max info from many features/columns)