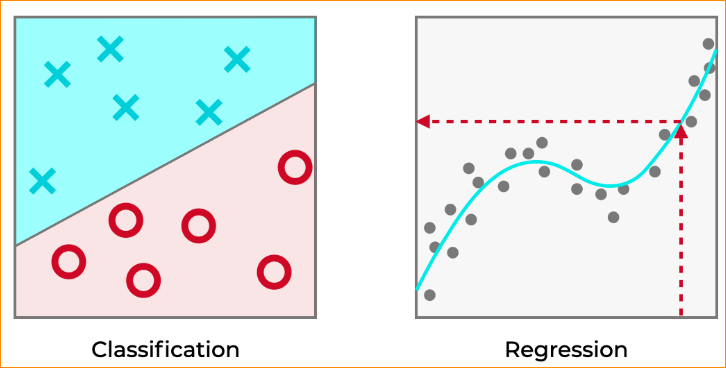
* If you have a data set with labeled observation – can train data using already labeled data. We call service learning.
* If you have continues variable to predict for that we are using **Regression**
* If you have to predict the categorical variable (like gender, or difficulty of exam) using **Classification**.

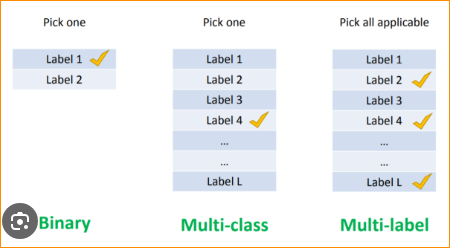


**Example of Classification-**

***Cancer patients in malignant or not that consider size of the Tumor***

In here that variables that we have to predict has two labels. (Class labels)

The cancer is malignant or not

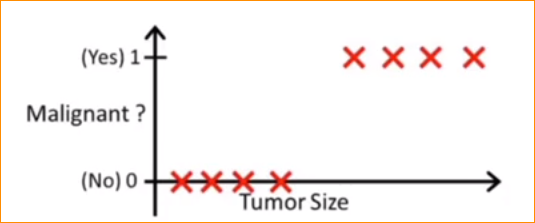


Why Classification predict the categorical variable? Why not suitable Regression for the this Example

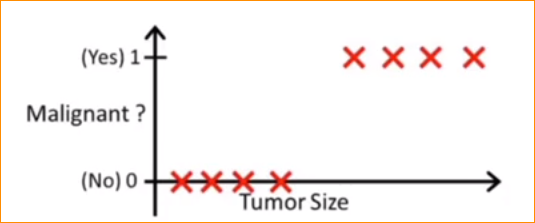
Data set –

|  |  |
| --- | --- |
| Tumor | Malignant or not |
| 2 | NO (0) |
| 3 | NO(0) |
| 1 | NO(0) |
| 4 | NO(0) |
| 5 | YES(1) |
| 6 | YES(1) |
| 7 | YES(1) |
| 8 | YES(1) |

Graph -



Why Regression is not Appropriate –



No

Yes

0.5

In here , we draw regression line and give malignant middle as 0.5

And just assume equation Y = 10 + 2X

And just right side Yes and Left side No

Just think X = 2.5

Y = 10 + 2 x 2.5

Y = 15

That mean 15>0.5 So Consider as Yes – Correct

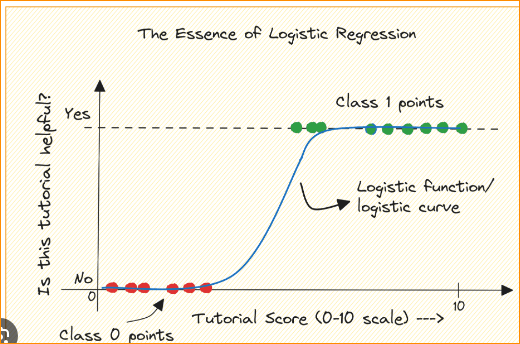
If Y = 50 -Yellow dot – Consider Yes – Correct

Green Dot – It should be Yes – But this Regression line give NO – its Wrong

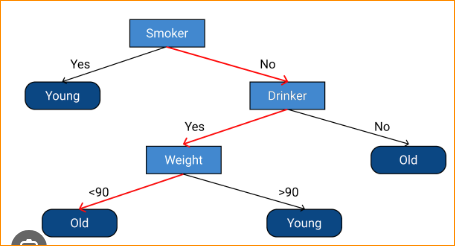
That why Regression can not be use to the categorical variable.

For that We are using Classification models-

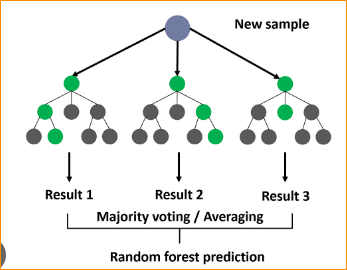
* **Logistic Regression**: Despite its name, logistic regression is a linear model used for binary classification tasks, where the target variable has only two possible outcomes.



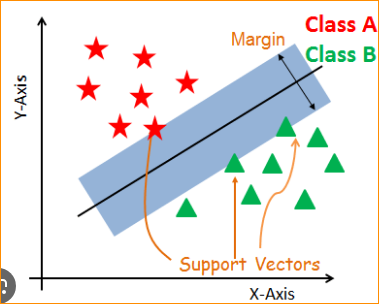
* **Decision Trees**: Decision trees partition the feature space into a hierarchy of binary decisions, forming a tree-like structure. They are intuitive and easy to interpret.



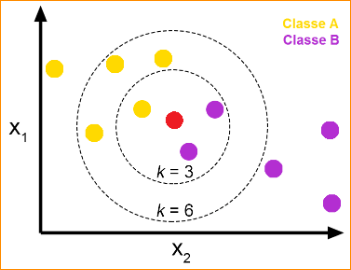
* **Random Forest:** Random forests are an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and robustness.



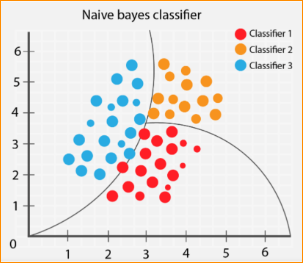
* **Support Vector Machines (SVM):** SVMs find the hyperplane that best separates the classes in the feature space. They are effective for both linear and non-linear classification tasks.



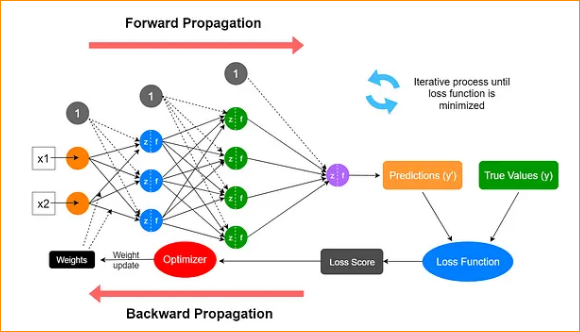
* **k-Nearest Neighbors (k-NN):** k-NN classifies new data points based on the majority class among their k nearest neighbors in the feature space.



* **Naive Bayes:** Naive Bayes classifiers are based on Bayes' theorem and assume that features are conditionally independent given the class. They are simple and efficient, often used for text classification tasks.



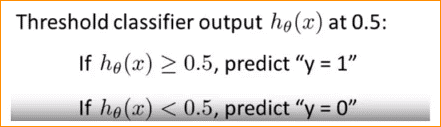
* **Neural Networks:** Neural networks, particularly deep learning models, have gained popularity for classification tasks due to their ability to learn complex patterns from large amounts of data.

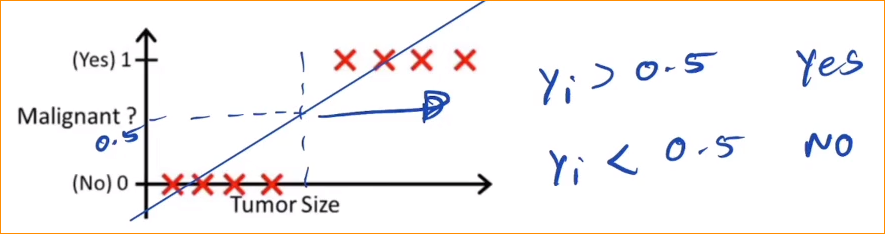


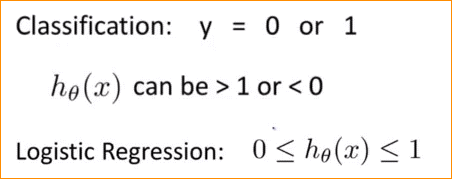
**Can be use the Logistic regression to Predict the categorical variable**

1. Set the threshold value

H(X) - Probability of having malignant cancer







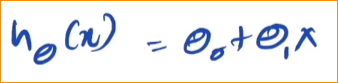
In here if you take example that Y is large.

Y = 10 + 2.5 X

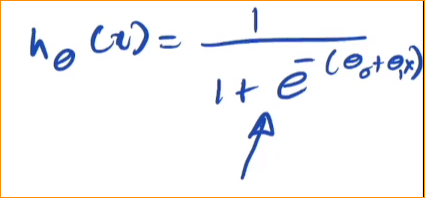
So X= 2 , Y is 15, So h(X) to be between 0 and 1

So now 15??

It is very hard to predict.

Simple linear Regression = 

So in here We use Equation like this- For the categorical variable equation =

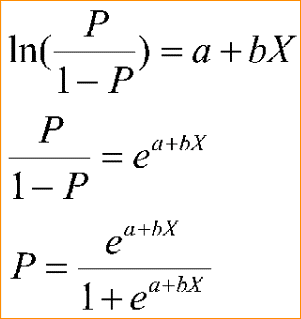


This call Logistic Regression module

This call sigmoid function , logistic function

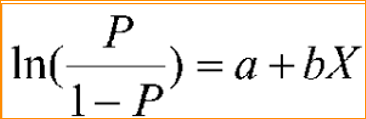
So in here all prediction is between 0 and 1.

Other logistic regression equation is considered probability

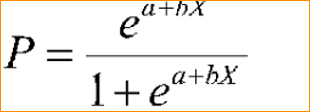


P – Probability of having malignant cancer [h(X)]

If you solve this



Can get this



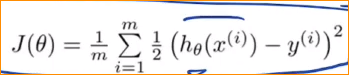
* Again we have to get the values of theta 0 and theta 1.

How to get theta 0 and theta 1 (There are two methods)

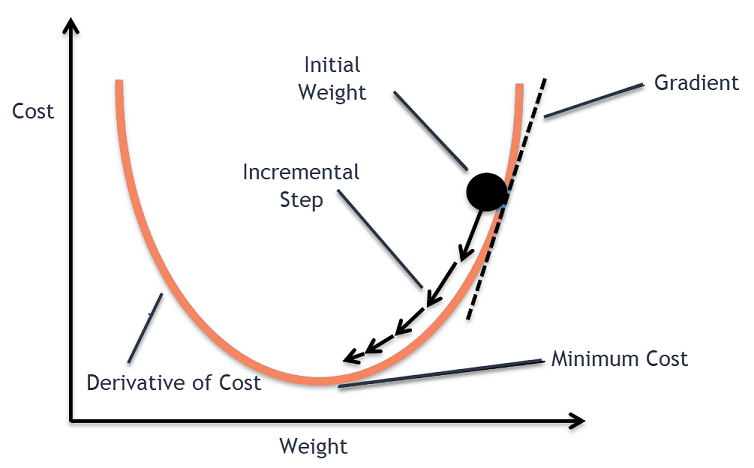
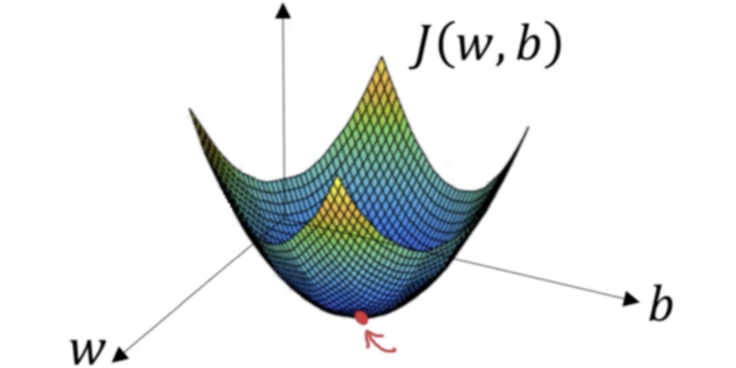
1. Leas square method
2. Gradient decent Algorithm

Can we use cost function to find theta 0 and theta 1.

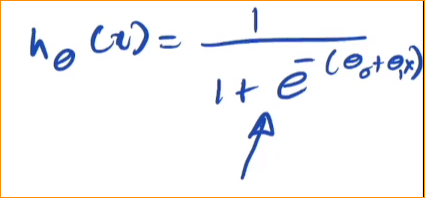
Single Linear regression –

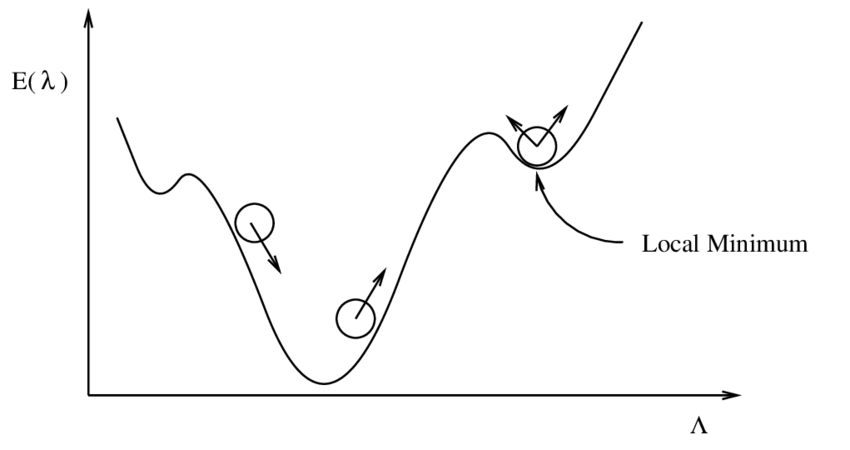
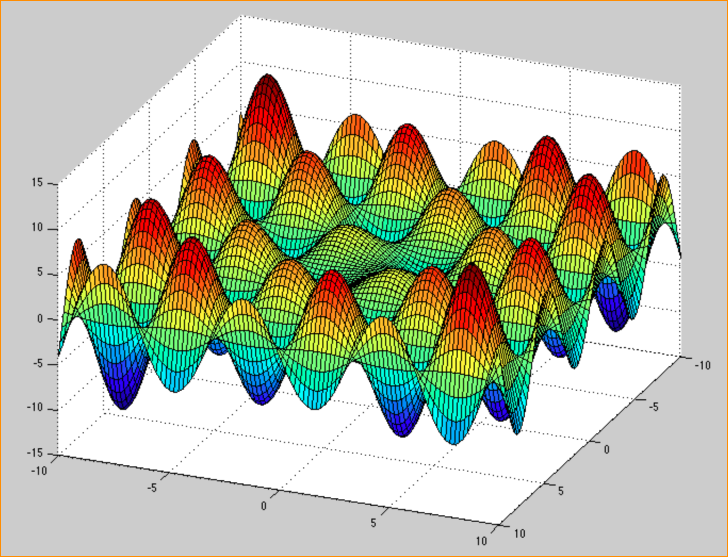


You have this king of graph – one minimum point



But considering Logistic Regression module –



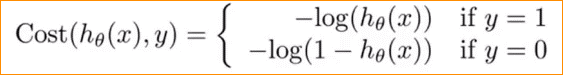


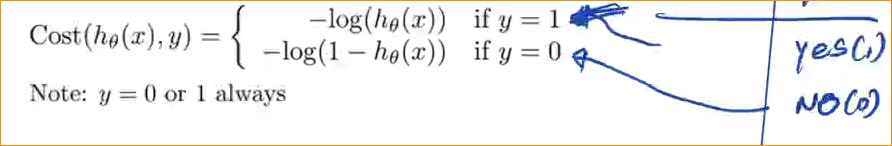
In here we have multiple local minimum points.

This mean we have to change the cost function also-

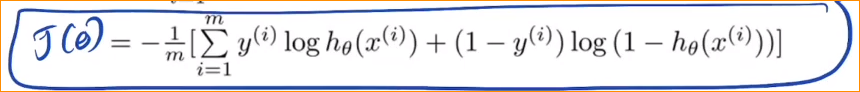


In here we are taking if y = 1 and y = 0 values of cost function

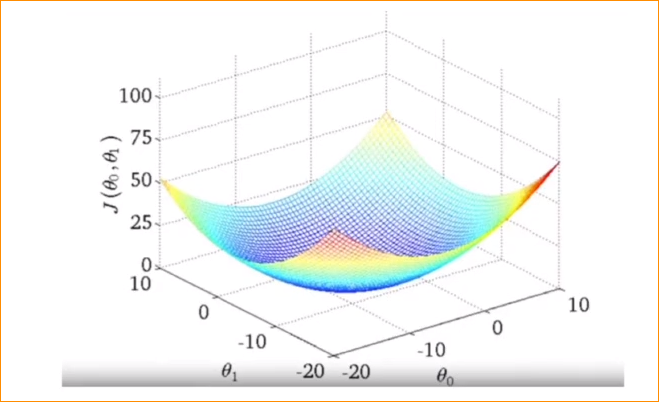




Now we can re write the cost function like this –



In this particular regression behave like Simple Linear Regression –

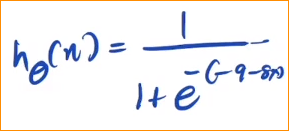


Now we are going to follow the same steps –

Now follow the gradient descent algorithm

Lets assume theta 0 = -9 and theta 1 = - 8

Probability of having malignant cancer [h(X)]



Or

