MACHINE LEARNING

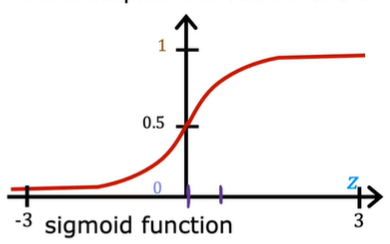
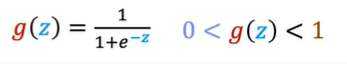
LOGISTIC

1. **Logistic Regression**

In contrast, logistic regression fit the curve like S shape to dataset. The curve logistic regression

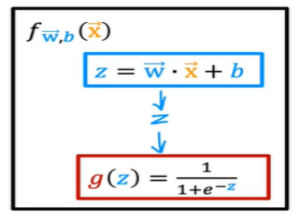
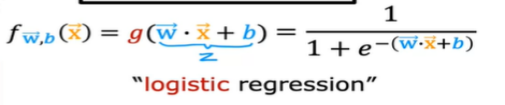
1. **Sigmoid Function**

The sigmoid function have ouput value between 0 and 1. this function that can we are predict classification.



When z is very large the value sigmoid can be very close to 1, and

When z is very small the value sigmoid can be very close to 0.



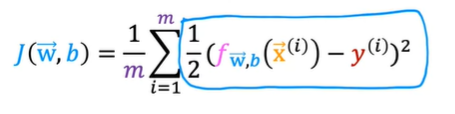
We know linear regression same with linear equatoin ( z = W.x + b).

And in Logistic Regression we can add the some **equation** to **Sigmoid Function.** So if some equation subtitute/add to Sigmoid Function is to be **Logistic Regression**

1. **Cost Function**

With cost function we can check/comparasion between predict data and actually data.

In linear regression we have this

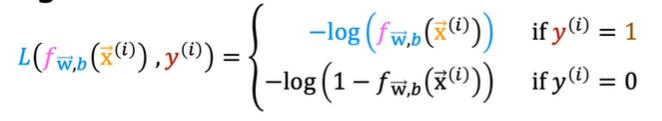


Actually the cost function for logistic regression just a little a bit different with cost function for linear regression. In logistic regression there a some function a call is **loss cost function**.

We can change the value in square to be loss cost function.

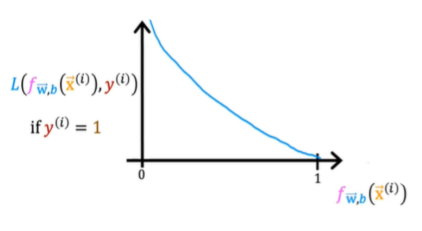


This is loss cost function



So for loss cost function we have 2 condition, the first when y = 1, and second when y = 0 .

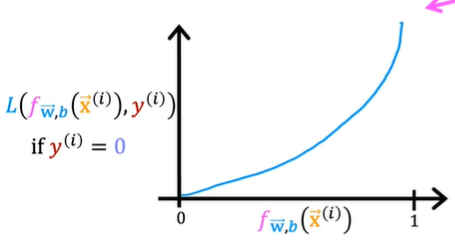
So we can look the curve when y = 1 , this curve use when y = 0 or when the actual data = 0 .



For example

When the value y = 1   
 if we have **predict** value is **1** , the loss to be **0**.

And if we have **predict** value is **0**, the loss to be **infinite.**



And than for y = 0, the curve can like grafik on abouve

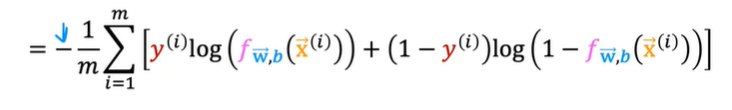
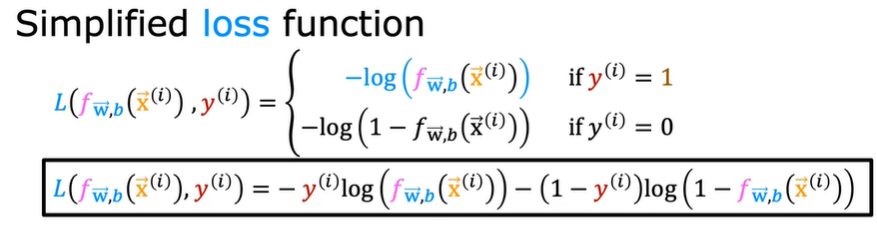
So this for tagret/actual data is 0,

If we have **predict** is **0**, the loss to be **0** , and

If we have **predict** is **1**, the loss to be **infinite**

So after we know the loss function, we can make the loss function simpefied.

The equation can help to implement in gradient descent

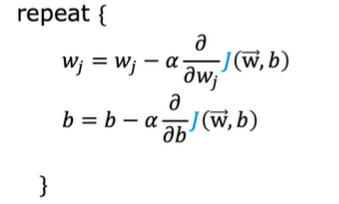


And the this is final equation for loss funtion after we combine 2 condition above.

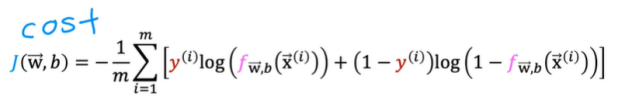
1. **Gradient Descent**

after that we can calculate how to comparation actual and predict data with lost function, for have good result modeling we must **find best of parameter** from model, in this case the parameter is W1, W2, … and B.

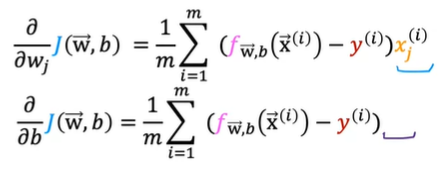
This is Gradient Descent Equation



We know the loss function can be like this

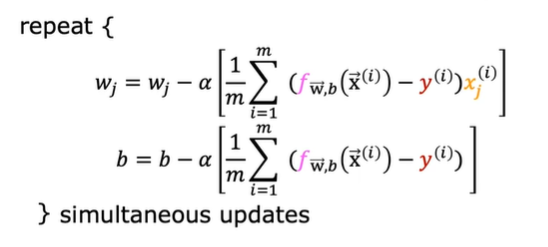


And we can try to diffentiation the equation



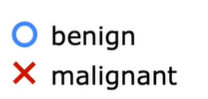
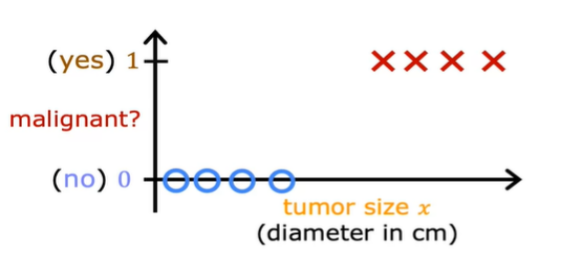
After we have value after derivative some equation.

We have gradient descent like this.



1. Example Calculation
2. Example Coding in python
3. **Logistic Regression Reason**
4. **Why we dont use Linear Regression for Classification Algoritma**

In this course we can learn classification algoritm



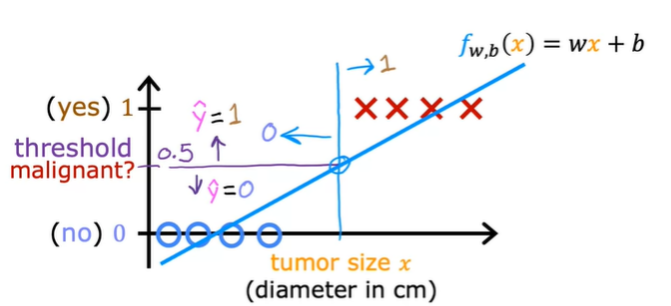
For example case classification is tumor detection, in chart at above we can define

1 = malignant

0 = benign

We can try to predict tumor malignant (1) or benign (0)

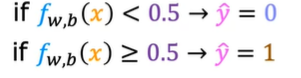
For the experience we can try user linear regression to predict them, and the result like this



Linear regression not just predict 0 and 1 , but number between 0 and 1 (ex : 0.2 , 0.1, 0.55).

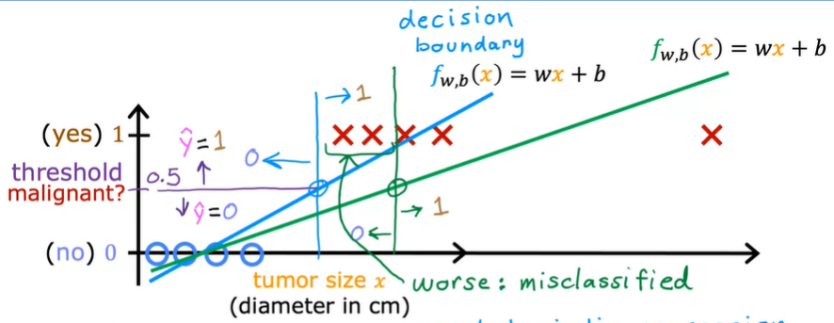
but for now we can predict category, so we can take **Threshold**  so separate predict.

With threshold we can predict .



This about can work, if data simple.

But how if the data some like this



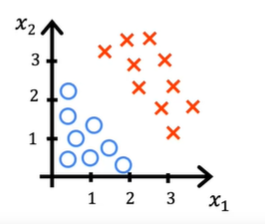
If we still use linear regression the prediction can be worse (misclassified)

1. **Decision Boundary**

The Decision Boundary is the equation that can separate between classification. The great algoritm have the great Decision Boundary. And why we use cost funtion and gradient descent, because to find the best of Decision Boundary.

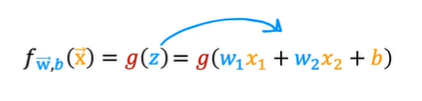
So in this case I have example data like this.

**Example 1**



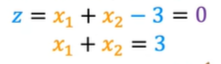
We have data like this, the data have 2 feature x1 and x2, we can define red color is 1 and blue color is 0 .

So we have logistic algortim like this

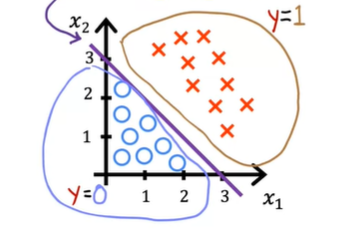


For example we take W1 = 1 , W2 = 1 , b = -3

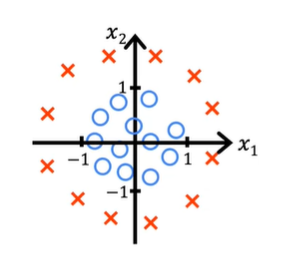
Decesion boundary is **z = 0 ,**  in this case if we subtitute the value equal like this.



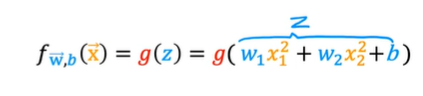
So the Decision boundary like this. Z equation the purple line in curve.

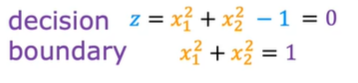


**Example 2**

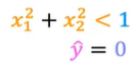
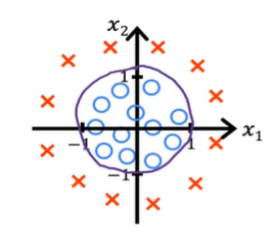


So we have logistic algoritm like this



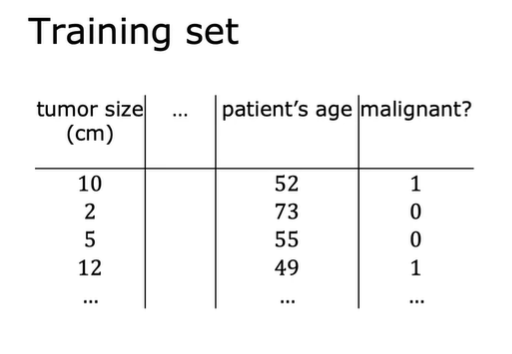


we have know the equation is circle equation so the decision boundaries like this.

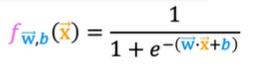


1. **Why use loss Cost Function in Logistic Regression**

For example, we have data like this



The target for predict y is 0 or 1, we know the logistic regression equation like this.

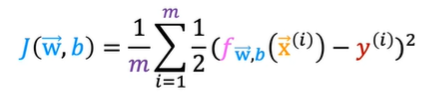


So to check find best classification and best of boundary decision you must choice W and B .

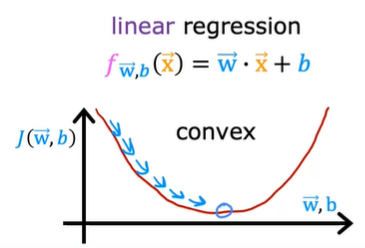


In linear regression we can use Square Error Cost / Cost Function for linear regression for comparatin between actual and predict.

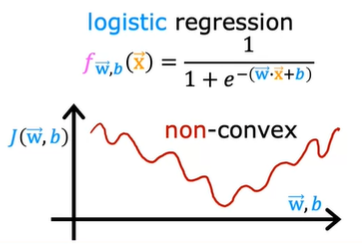
The equation like this



And we can have curve like this .

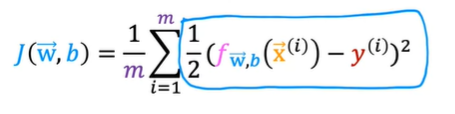


And if we use logistict regression function, the curve to be like this :



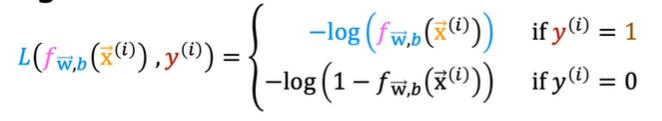
But if we use logistic regresiion the curve to be non convex , that mean if we are try to find gradient descent it to be local minimum and the curve have many local minumum . so in this case **squared error cost** function is **not choice**.

So we must have another cost function :



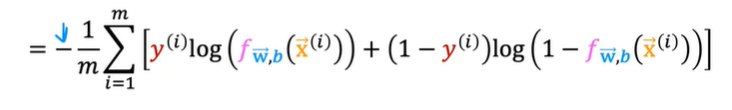
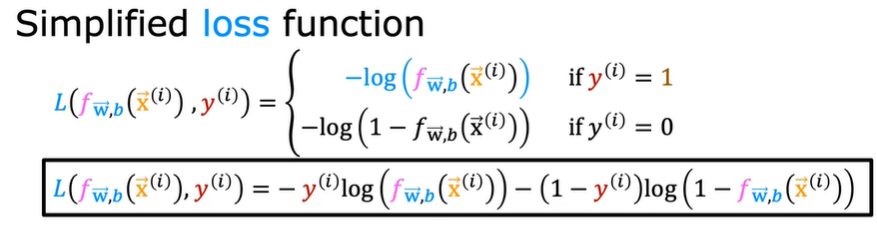
We can change the value in square to be **loss cost function**.





So after we know the loss function, we can make the loss function simpefied.

The equation can help to implement in gradient descent

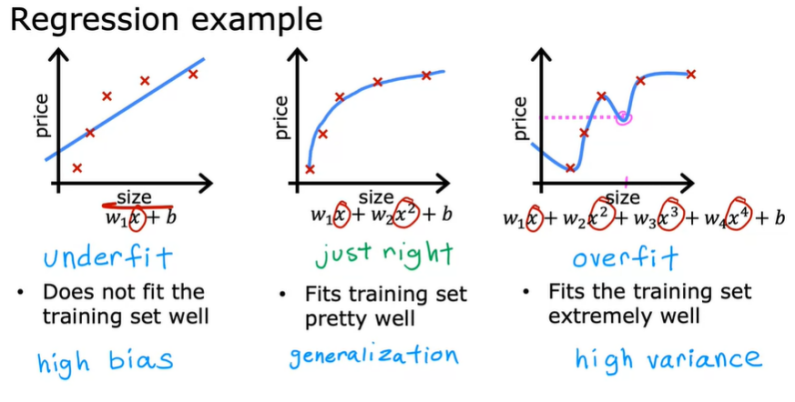


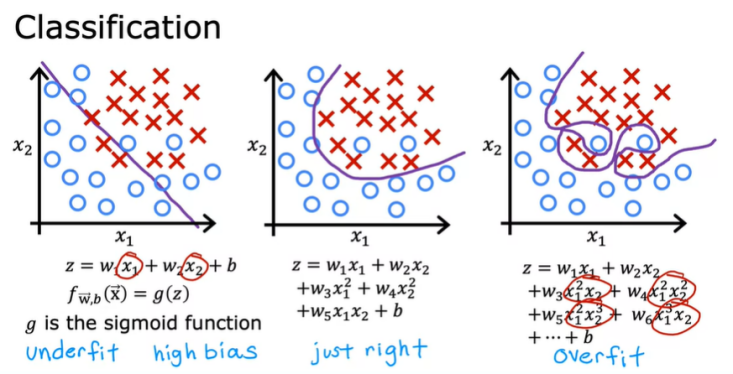
And the this is final equation for loss function

1. Overfitting
2. **Introduction**

What is overfitting? In below this is example of underfitting, overfitting and generalization.

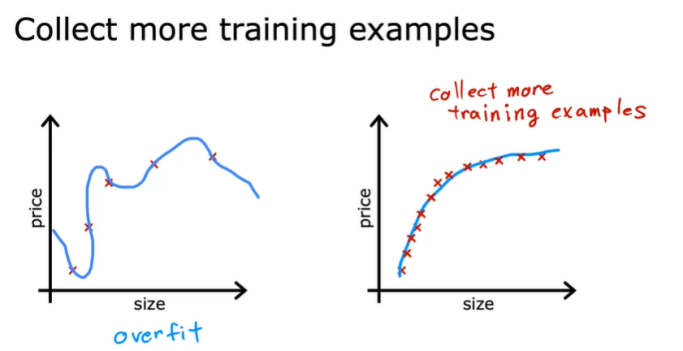
The best model must to be generalization (just right).



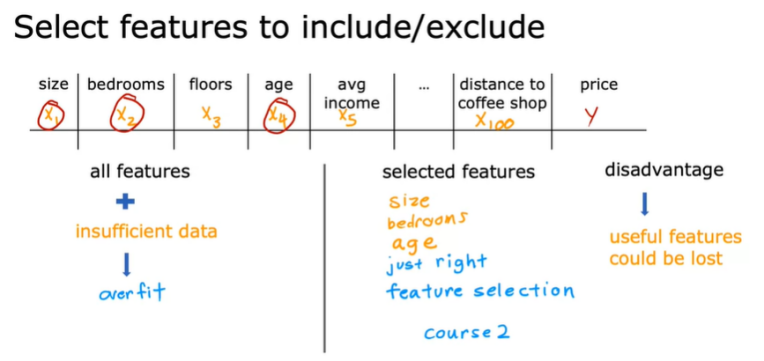


For the handle the overfitting we have 3 way

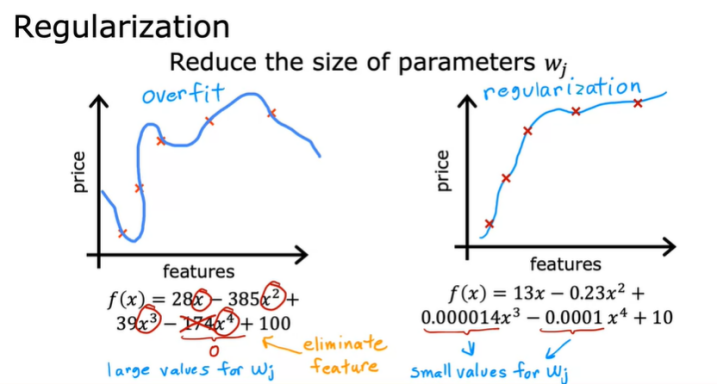
1. Collect more data



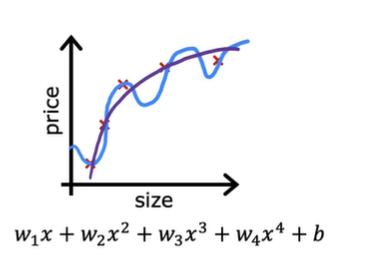
1. Select Features (explaination detail in course 2)



1. Regulisatoin



1. **Regularization**

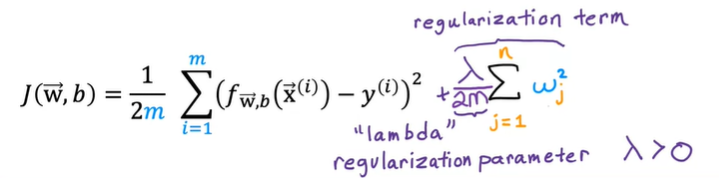


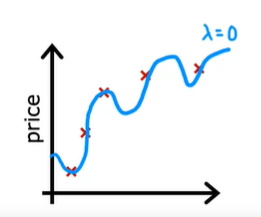
We can look, the curve at above is overfitting the idea of regularization is make w3, w4 is really small. When w3 and w4 very small in can decrease impact for function.

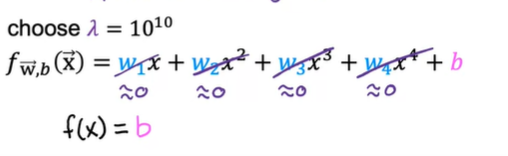


1. **Regularization Linear Regression**
2. **Cost Function**

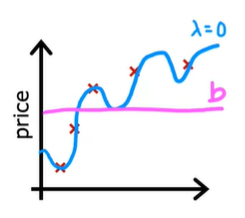
For do that we can modifie the cost function.





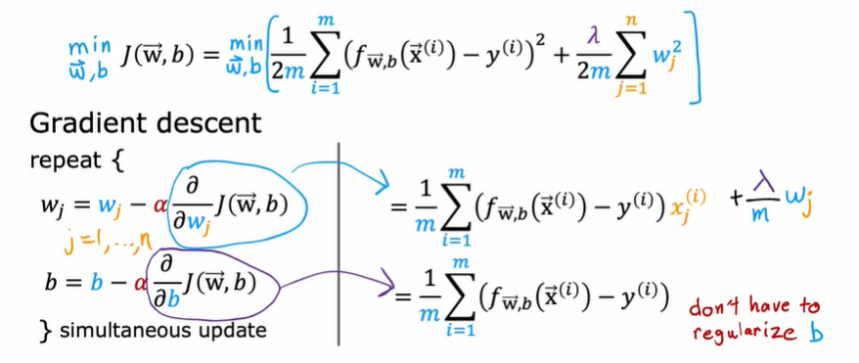


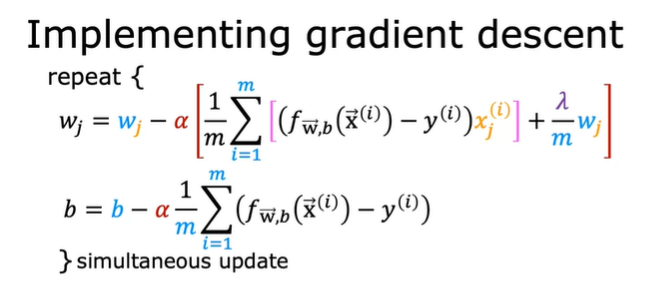
We must find lambda not small and not large

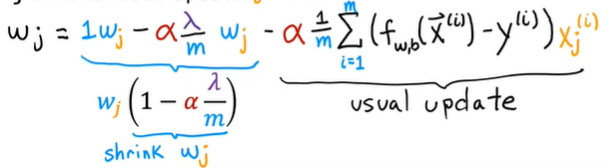


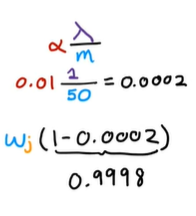
1. **Gradient Descent**

Gradient Descent with Regularized Linear Regression









1. **Regularization Logistic Regression**

Regularized Logistic Regression

