**Part 2**  
Perform two iterations of the gradient algorithm to find the minima of

The starting point is

Draw the contours and show your learning path graphically.

So we begin with the original function and starting vector:

(Equation 1)



The gradient is just a set of partial derivatives of the vector. So we can define our gradient as:

(Equation 2)



With this, we can plug in the functions above and differentiate with respect to :



(Equation 3)



and :



(Equation 4)



And our gradient is then defined as:

(Equation 5)



Gradient descent is an assignment function that updates the value based on the gradient and learning rate. The default function:

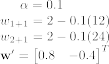
(Equation 6)



where alpha is the learning rate and the gradient is defined above.

Now we will do two iterations of the function and gather the coordinates of two points in the direction of the gradient:

(Equation 7)



So we’ve moved from to .



Now starting at we will go down another step:



(Equation 8)



**Part 3**

Show that logistic regression is a nonlinear regression problem. Is it possible to treat logistic discrimination in terms of an equivalent linear regression problem? Justify your answer.

Let’s say we have two classes. There exists some probability of each glass given y and some training examples . For argument’s sake, both and are normally distributed:



(Equation 9)

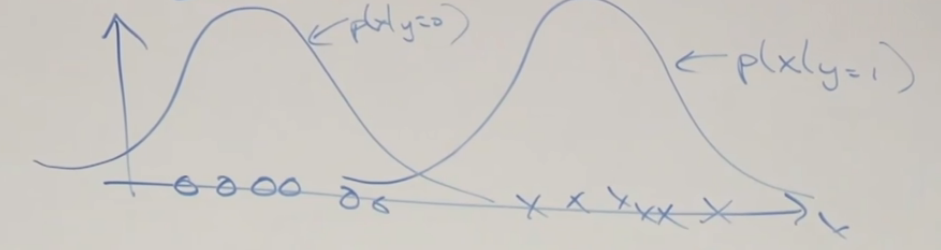


Say we had some data that was either 0 or 1 (yes/no, benign/malignant, etc) and say we had just one feature



We could then draw Gaussian curves over each space to see the distribution. As can be seen, if we have a point, say then we can see there is a high probability of it being from the left hand distribution. Same for a point that would be to the right.

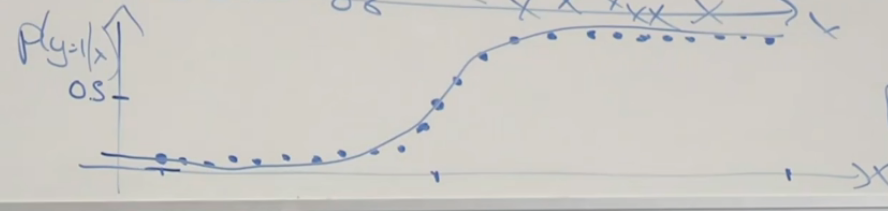




If we were to choose one class and then draw these probabilities out on a graph…

(Equation 10)





And we can see that as the data progresses in x, the probability of y being 1 increases, creating a sigmoid function.