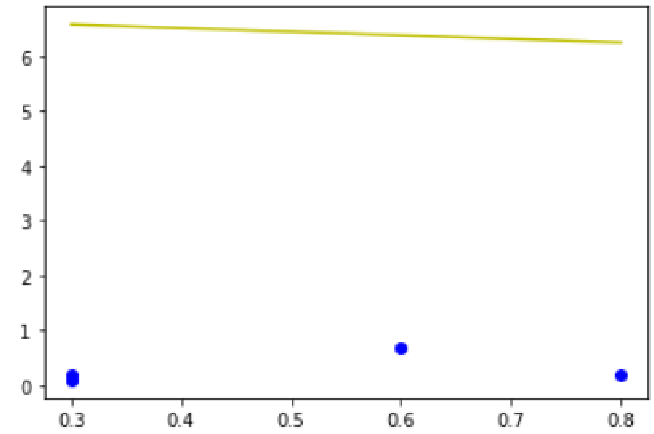
Qestion1:

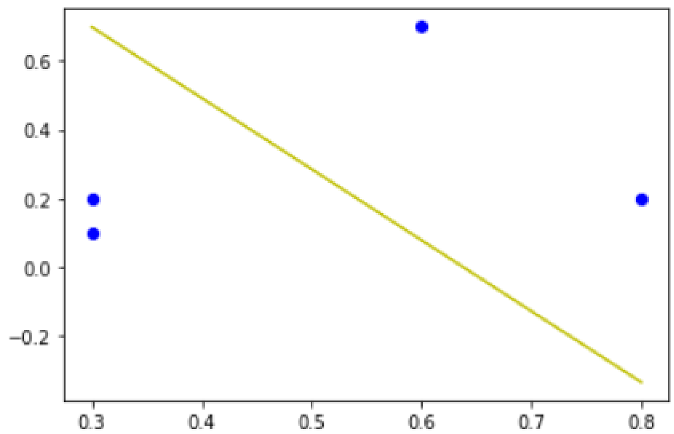
a) I don't think the resulting model is appropriate for the task being learned. From the given code we can get:

array([-6.69773652, 0.64952447, 0.98823519])

which the input variables are: gradientDescent([[1,0.3,0.1],[1,0.3,0.2],[1,0.6,0.7], [1,0.8,0.2]], [0,0,1,1], 5, 0.1, 1000)

And we can see have lambda 0.1. For using lambda as 0.1, it may be too big to train the small w, which means may cause underfitting.



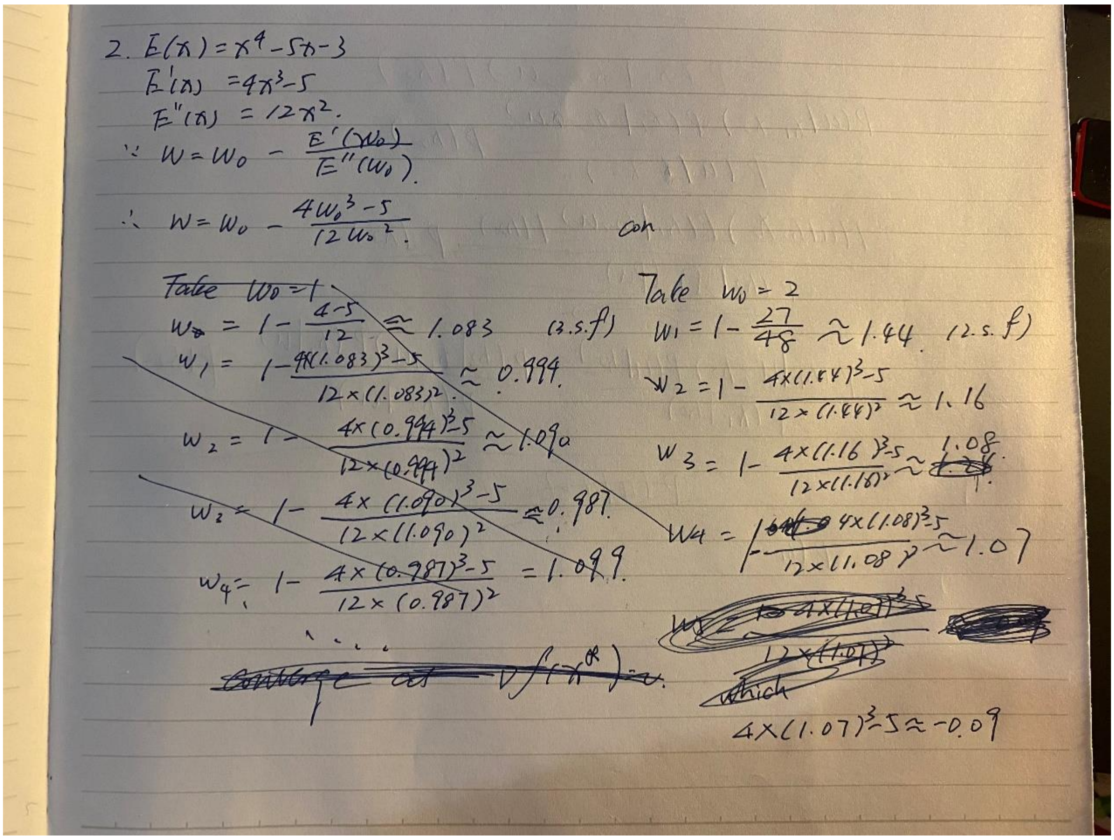


[ ]:

These two figures show when lambda with value 0.1 and 0.001, which we can see 0,001 is much better。

For the positive point,, I think is the iterations number which is 1000, it's enough for the data here.

Question2:



2. $E(x) = x^4 - 5x - 3$

$E'(x) = 4x^3 - 5$

$E''(x) = 12x^2.$

$\because W = W_0 - \frac{E'(W_0)}{E''(W_0)}.$

$\therefore W = W_0 - \frac{4W_0^3 - 5}{12W_0^2}.$

con

~~Take $W_0 = 1$~~

$W_0 = 1 - \frac{4-5}{12} \approx 1.083$ (3.s.f)

$W_1 = 1 - \frac{4\times(1.083)^3 - 5}{12\times(1.083)^2} \approx 0.994.$

$W_2 = 1 - \frac{4\times(0.994)^3 - 5}{12\times(0.994)^2} \approx 1.090$

$W_3 = 1 - \frac{4\times(1.090)^3 - 5}{12\times(1.090)^2} \approx 0.987.$

$W_4 = 1 - \frac{4\times(0.987)^3 - 5}{12\times(0.987)^2} = 1.099.$

...

~~Converge at $f(x^*)$~~

Take $W_0 = 2$

$W_1 = 1 - \frac{27}{48} \approx 1.44$ (2.s.f)

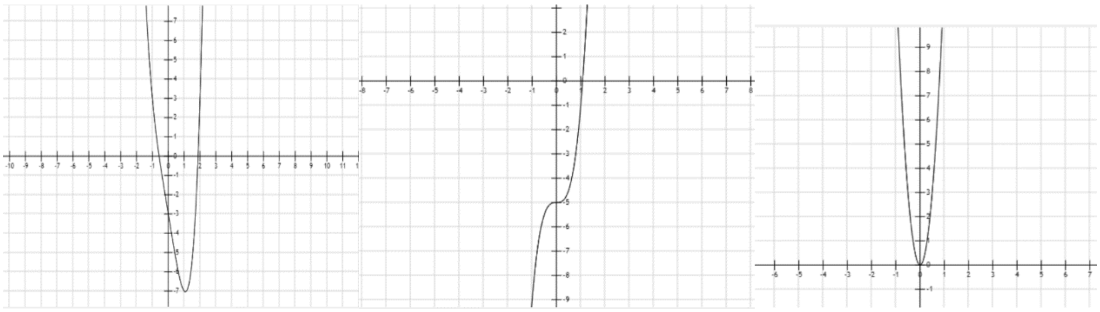
$W_2 = 1 - \frac{4\times(1.44)^3 - 5}{12\times(1.44)^2} \approx 1.16$

$W_3 = 1 - \frac{4\times(1.16)^3 - 5}{12\times(1.16)^2} \approx 1.08.$

$W_4 = 1 - \frac{4\times(1.08)^3 - 5}{12\times(1.08)^2} \approx 1.07$

~~which~~

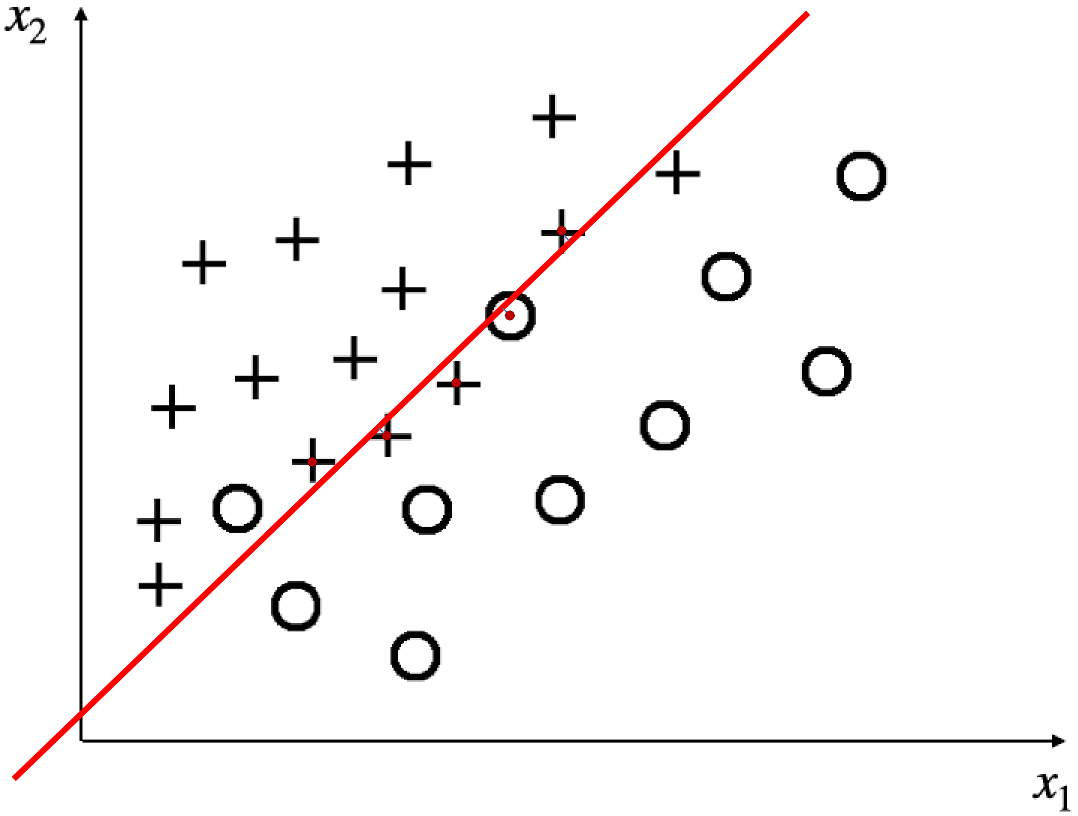
$4\times(1.07)^3 - 5 \approx -0.09$

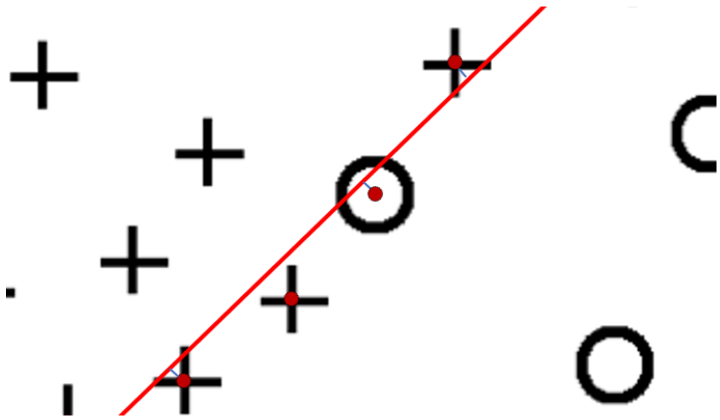


(image of E(x), E’(x), E’’(x))

Using the iterative Newton-Rapson rule we can find the optimal x(or w) value to find the minimum E(x). And from the graph of E’’(x) we can know that it must have minimum point. From above, we can see when w = 1.08 it will be the optimal, since E’(w_4) is less than 0. And at that point, we can have a minimum value since the graph is convex.

Question3:





Firstly, I find the closet samples of class 0 and class 1 and mark their center as some red circles, then choose a line to be the decision boundary, to make sure it gets the distance gamma (blue lines), which is the margin, so that the decision boundary can be the maximum one. And also, the decision boundary is depending on the level of model's overfitting.

Due to it is trained without regularization, and the decision boundary here may not classify these two classes very correctly, the points on the line or very close to the line still can be 0 or 1. The data here is linearly separate, then it could be overfitting, which means it have the probability for using the cross entropy, weight w is also with the probability of value 0 or 1. From the cross entropy to minimize the loss function, we can know that we have to maximize the $lnp(x^(i),w)$(since y^(i) is a constant number 1).Then if $lnp(x^(i),w)$ is more closer to 1,the maximum value it will reach., and with the negative sign "-" in the formula, we can now have a less value, which means we will have a minimum loss function value. Above all and the line of decision boundary, we can see that it still has overfitting. To fix this, we can use the model to train with regularization. Then reduce the probability of being 0 or 1 , and in case of a large w. Or we can create a penalty for the constraints, and the penalty should be high when the constraints are violated.