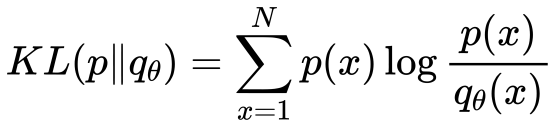
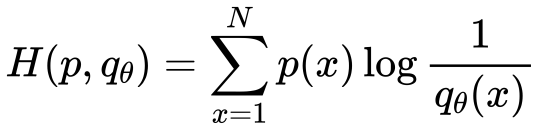
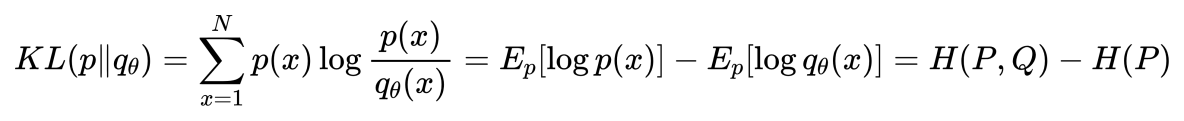
Q1:



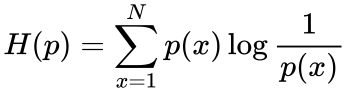
(a)



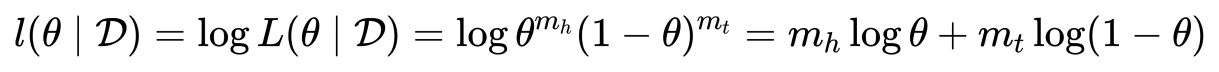
(b)



(c)

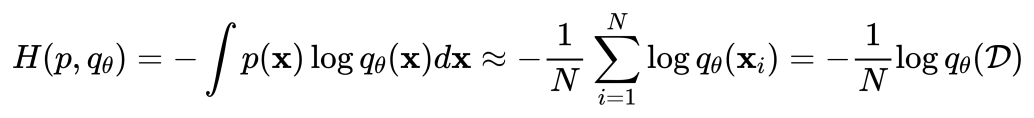


where

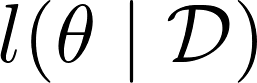
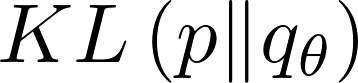
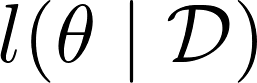
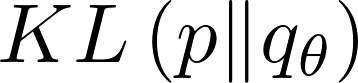


(d)

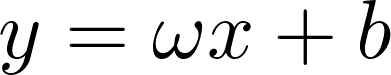
The pair /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.YZFgXswpsoffice here is a sufficient statistic.



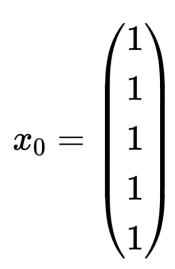
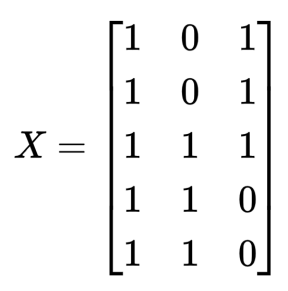
(e)

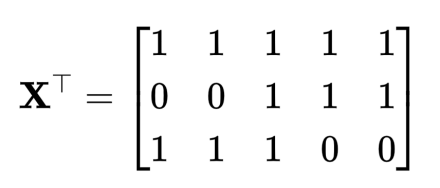
According to this formula,due to the negative wpsoffice parameter, in order to maximize  , we need to minimize  and minimize the . It means that for the parameter θ and distribution D, with bigger  has a smaller  and .

Q2:

1.  Assume that  , and the formula for the vector form is

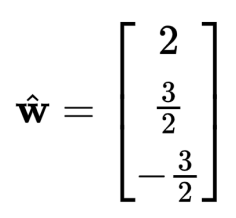
where and wpsoffice is the bias term.

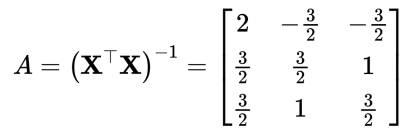
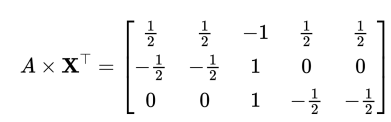


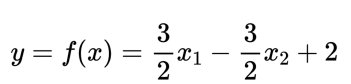


and therefore

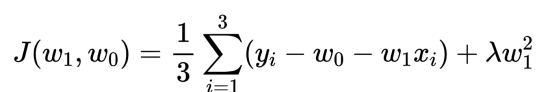
According to the lecture, to get compute the ordinary least squares solution,



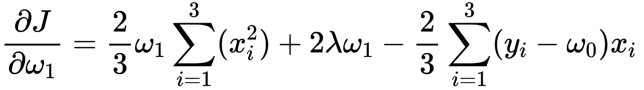


So, the result is

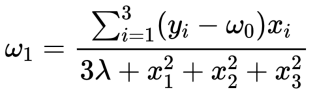
Q3: In my opinion, 1 matches (c), 2 matches (b), 3 matches (a), 4 matches (d)



The right penalized least squares is

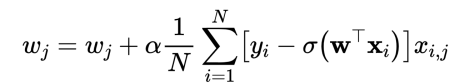


and

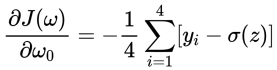
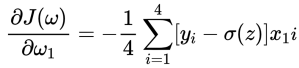


when

From the above equation we can conclude that as wpsoffice is larger, wpsoffice is smaller and the slope is smaller. so we can initially determine that (a) and (c) can correspond to 1, 3, while (b) and (d) can correspond to 2, 4. The variance of the 2 and 4 formulas is greater due to the presence of wpsoffice. This means that the variance of the distance between the point on the axis and the line is greater, so 2 matches (b),4 matches (d), and similarly 1 matches (c), 3 matches (a).

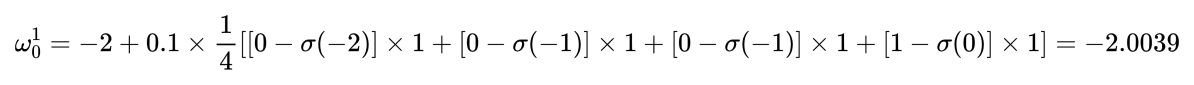
Q4:

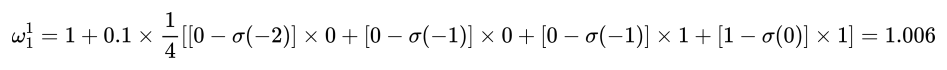
According to the lecture, in batch gradient descent algorithm

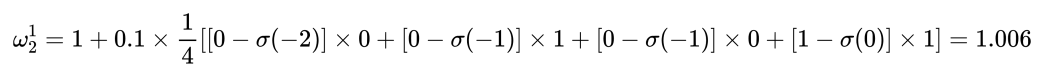
In this case, let;s turn this formula into vector form

Assume that , therefore



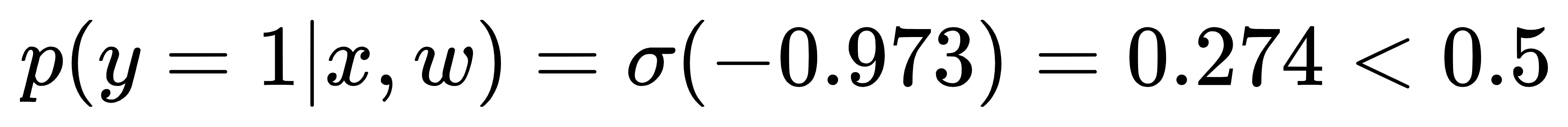


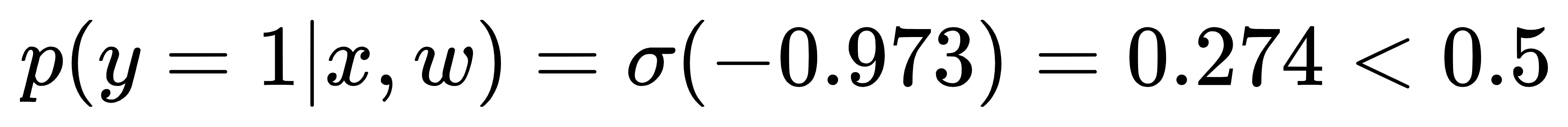
So

Similarly,

/Users/yuanfangxu/Desktop/Machine Learning/1.32.png1.32

/Users/yuanfangxu/Desktop/Machine Learning/1.34.png1.34The first point:  = (0,0) y=0, correct

The second point:  = (0,1) y=0, correct

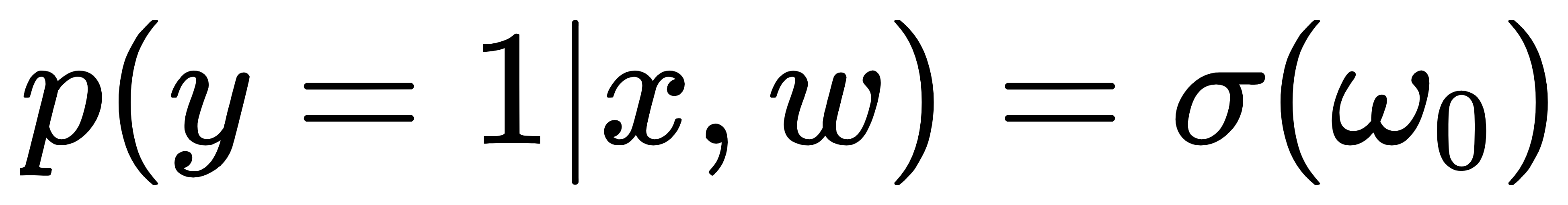
The third point:  = (1,0) y=0, correct

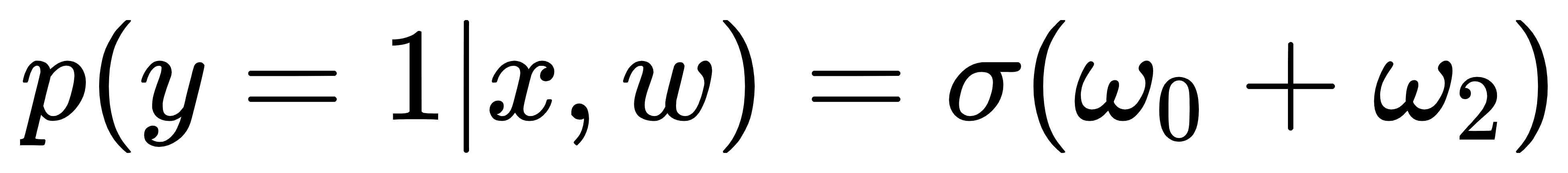
/Users/yuanfangxu/Desktop/Machine Learning/1.35.png1.35The last point:  = (1,1) y=1, correct

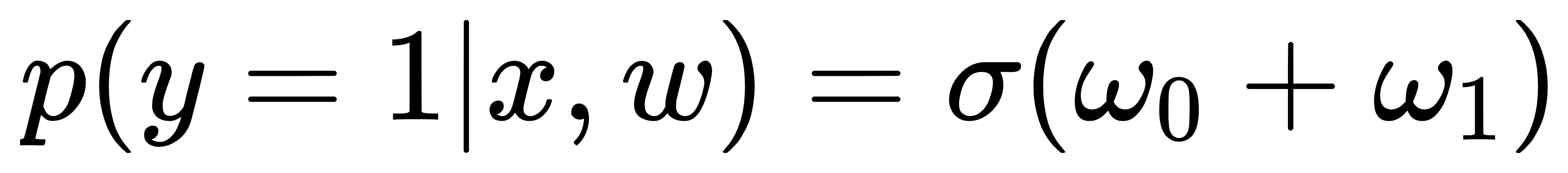
The training error is 0.

Q5:

1.

The first point: y=1,  = (0,0) if this training point is correct, then wpsoffice>0.

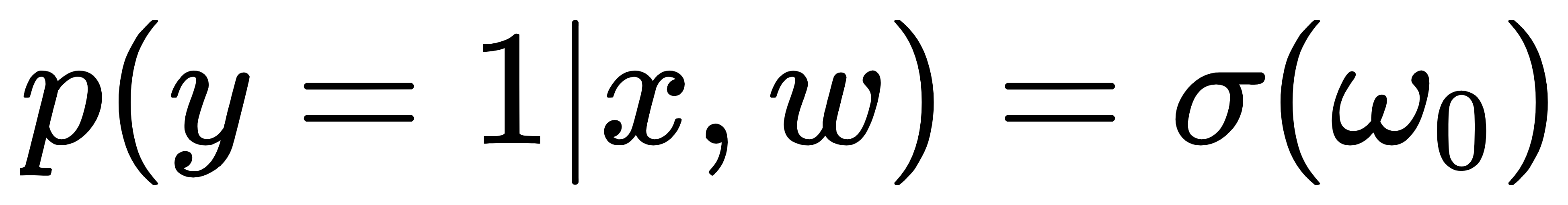
The second point: y=0,  = (0,1) if this training point is correct, then /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.PHIirYwpsoffice<0.

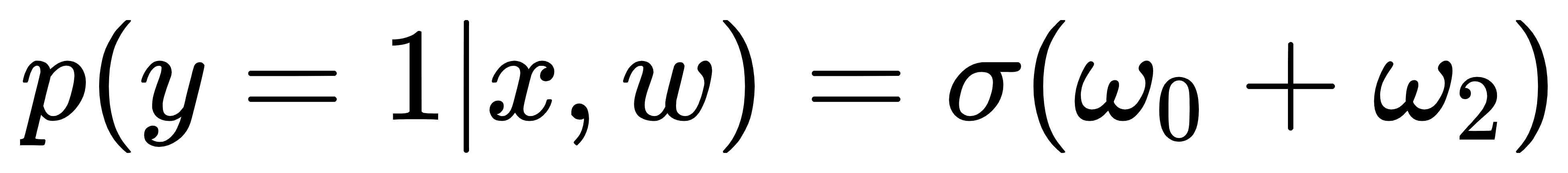
The third point: y=0,  = (1,0) if this training point is correct, then /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.ZTHDFPwpsoffice<0.

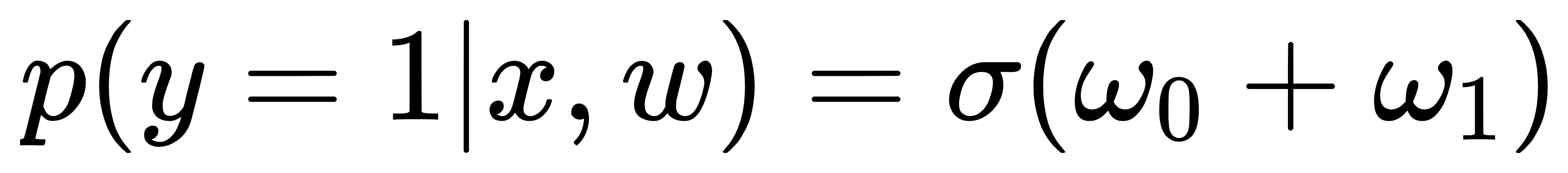
/Users/yuanfangxu/Desktop/Machine Learning/1.15.png1.15The last point: y=1,  = (1,1) if this training point is correct, then /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.PBEEmcwpsoffice>0.

There is no way to satisfy all four conditions at the same time, at most three conditions at the same time. So the minimum achievable training error in this case is 25%. An example that satisfies the minimum training error rate is /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.dTMjAEwpsoffice .

2.

The first point: y=1,  = (0,0) if this training point is correct, then wpsoffice>0.

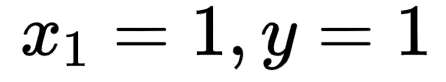
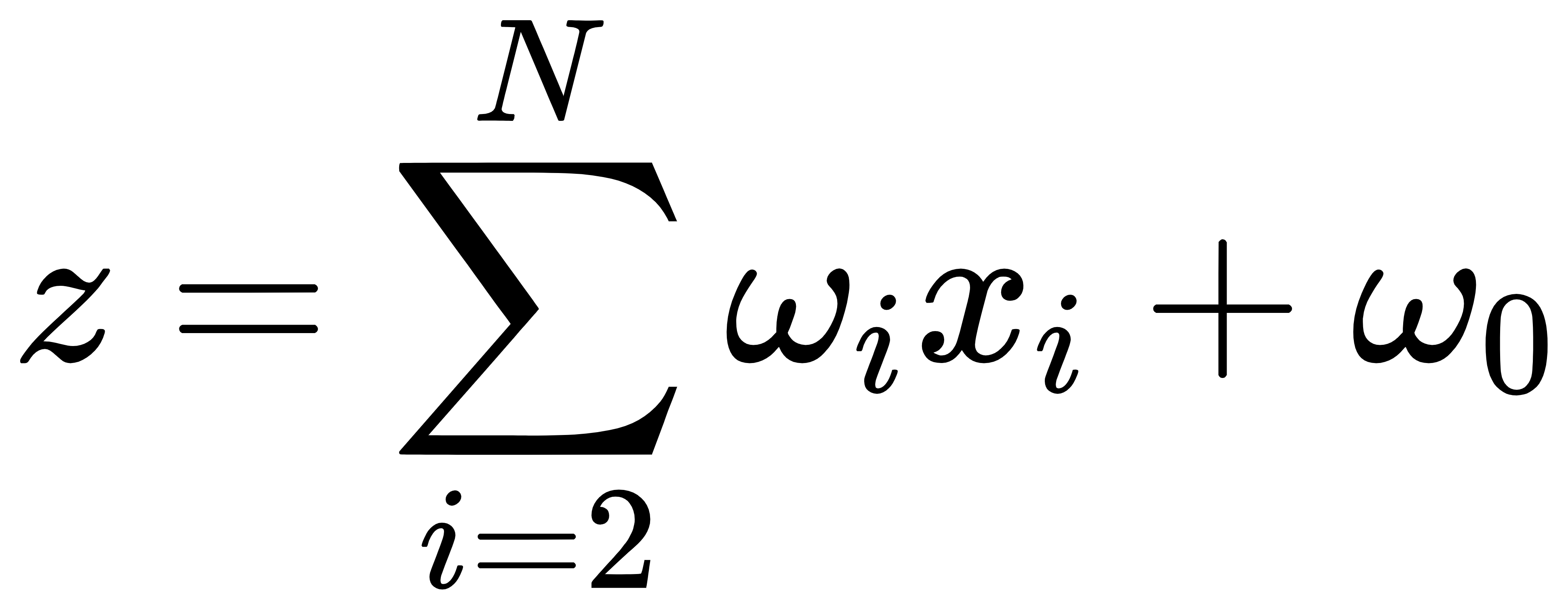
The second point: y=0,  = (0,1) if this training point is correct, then /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.PHIirYwpsoffice<0.

The third point: y=0,  = (1,0) if this training point is correct, then /private/var/folders/_0/3dvv6qpd4fnbq57bps0k1c6h0000gn/T/com.kingsoft.wpsoffice.mac/wpsoffice.ZTHDFPwpsoffice<0.

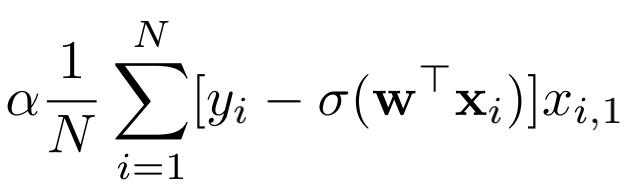
/Users/yuanfangxu/Desktop/Machine Learning/1.16.png1.16The last point: y=1,  = (1,1) if this training point is correct, then .

1.17It is easy to satisfy the above four conditions simultaneously, and the minimum training error is 0。An example that satisfies the minimum training error rate is

Q6

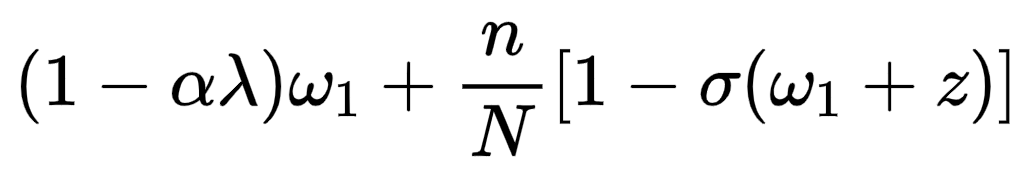
1. According to the question, when wpsoffice equals is 0, y equals 0, and when wpsoffice equals 1, y equals 1, And the data set with wpsoffice equal to 0 accounts for most of the data, which means that for most of the . And for the data that ,

1> > 0.5, where and invariant.

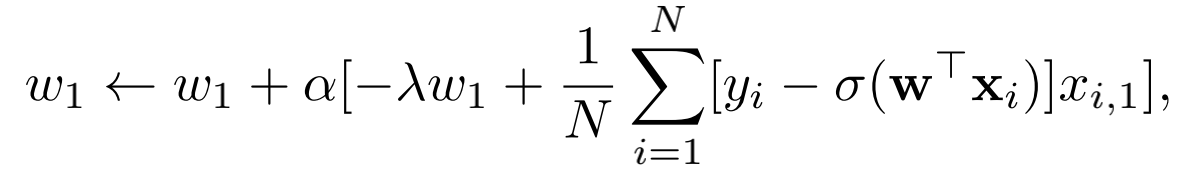


Therefore, >0(i.e. ) and equals to a small value a little larger

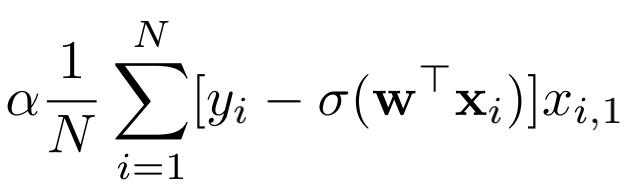
than 0, wpsoffice will increase little by little.

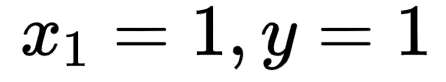


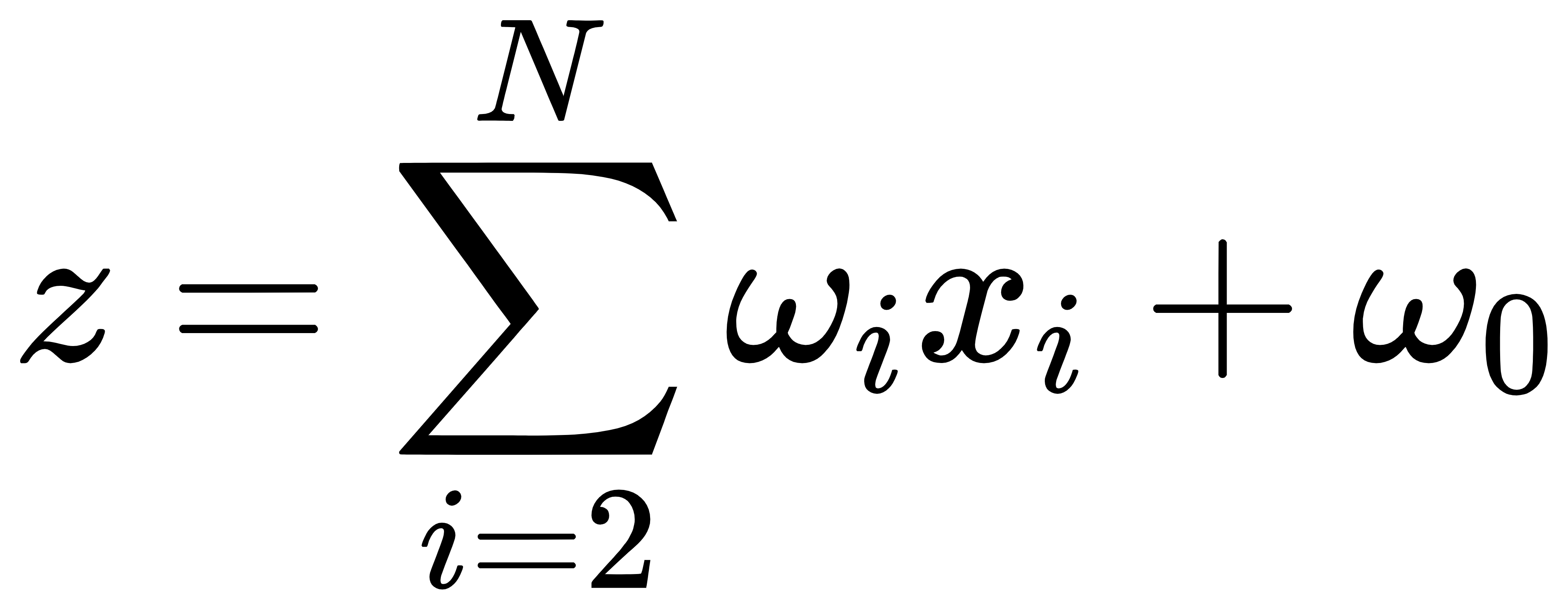
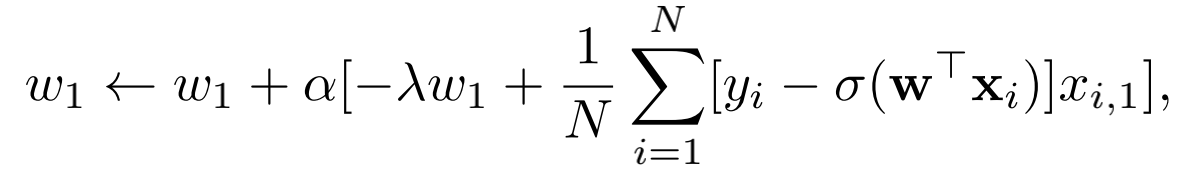
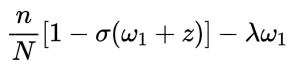
Conclusion： will not converge to a point, but increase all the time.



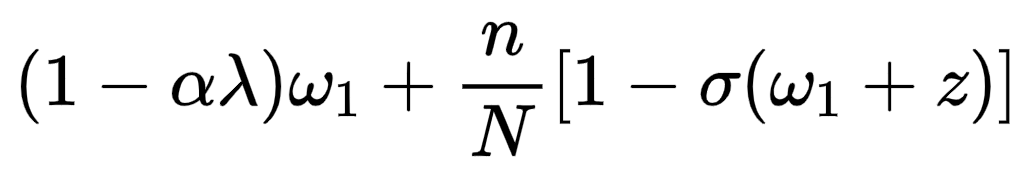
1. when we use the new update rule we assume that

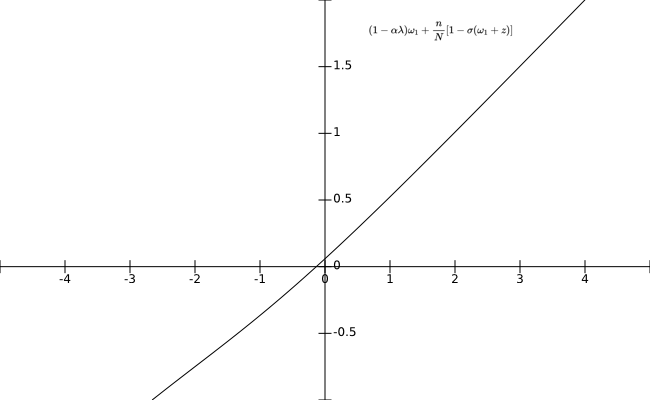
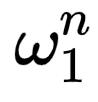
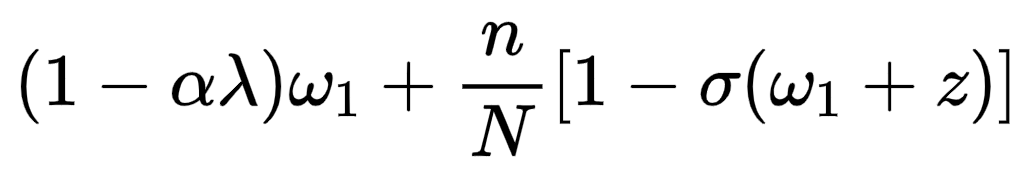


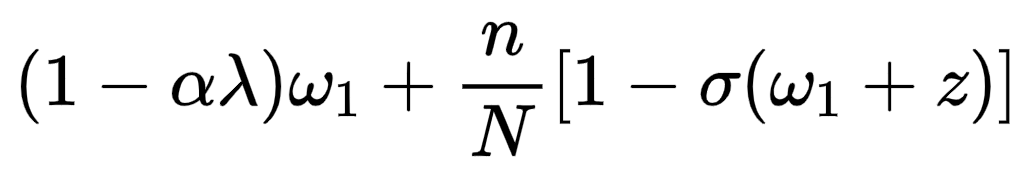
there are a total count of n that , so = where



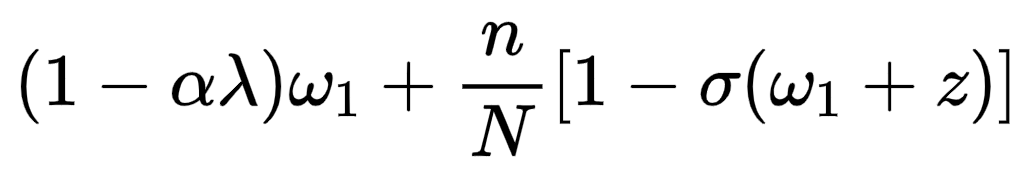
, so =

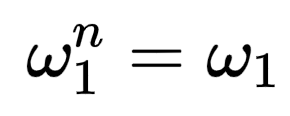
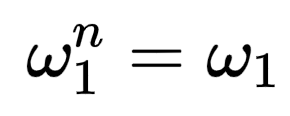


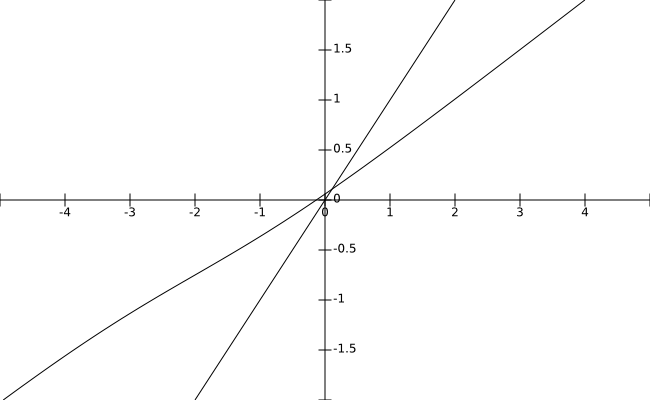
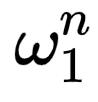
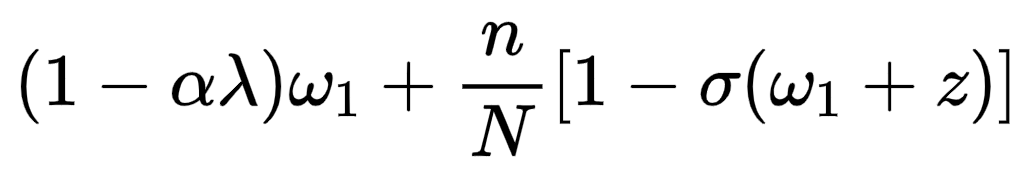
 =



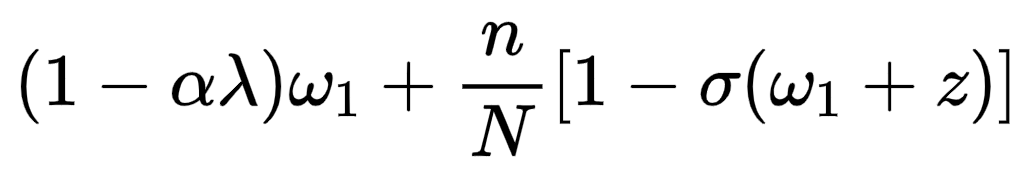
And we can get a figure for

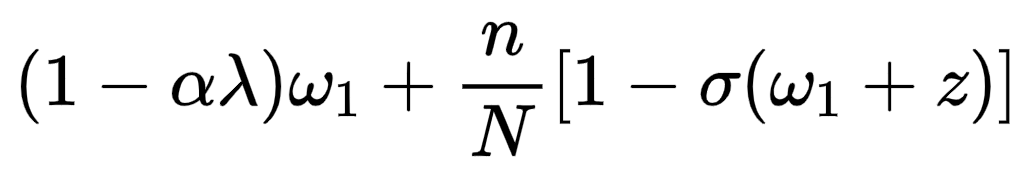
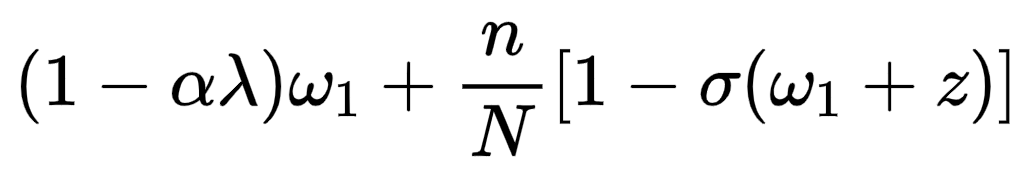
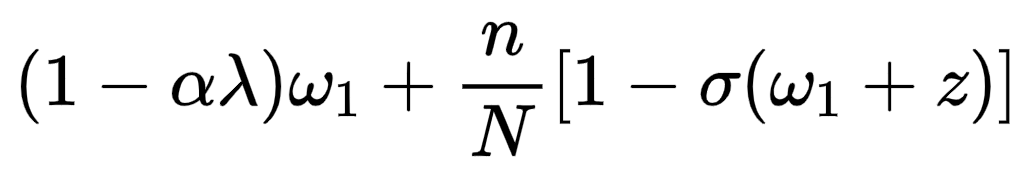


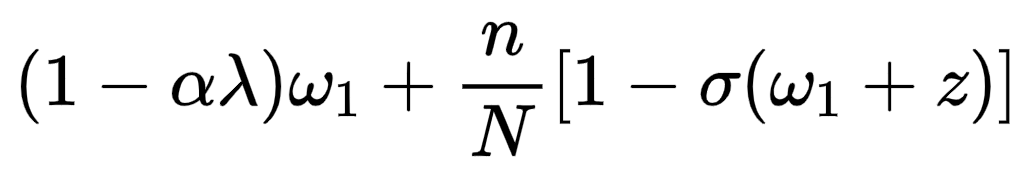
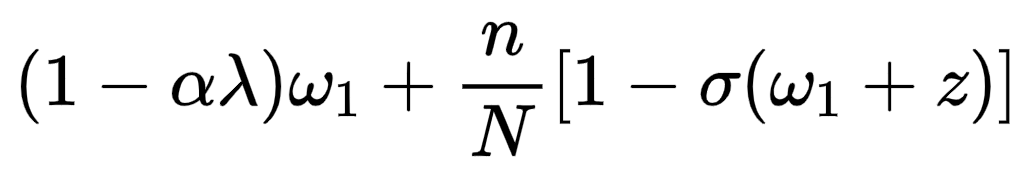
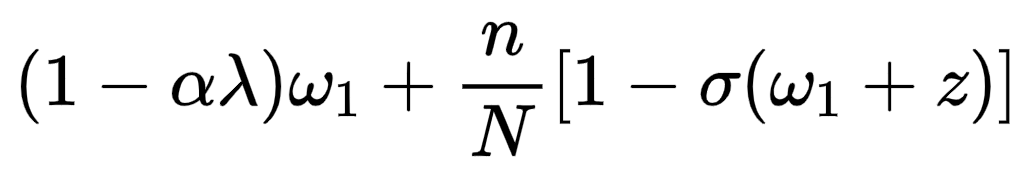
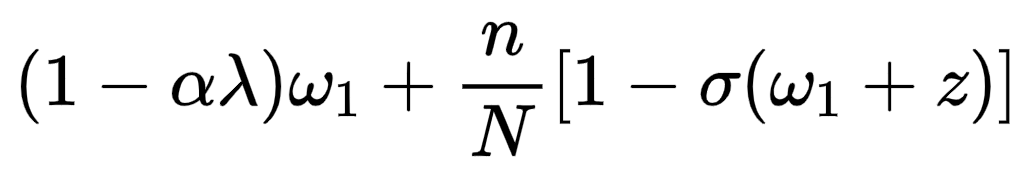
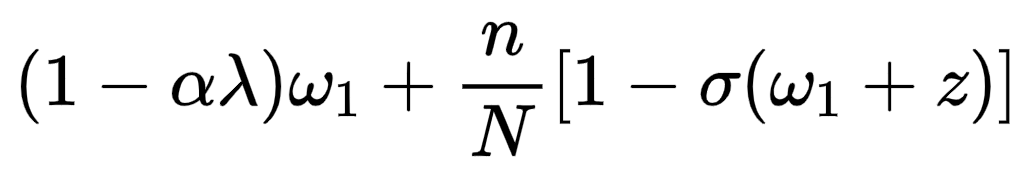
And than we compare it with

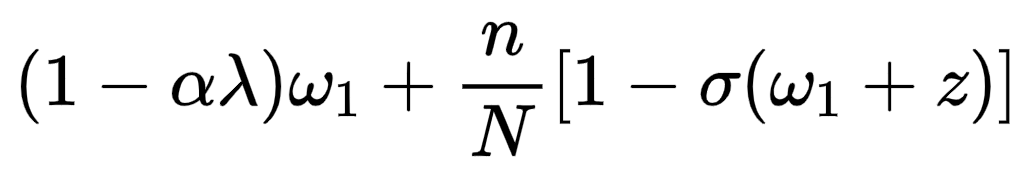
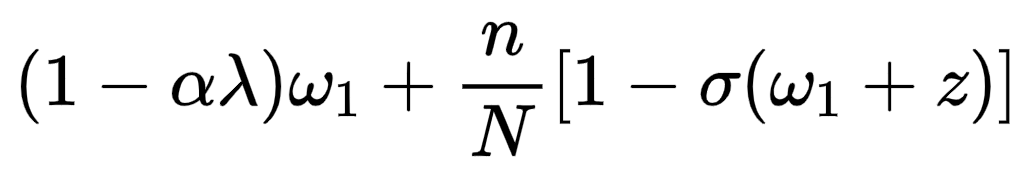


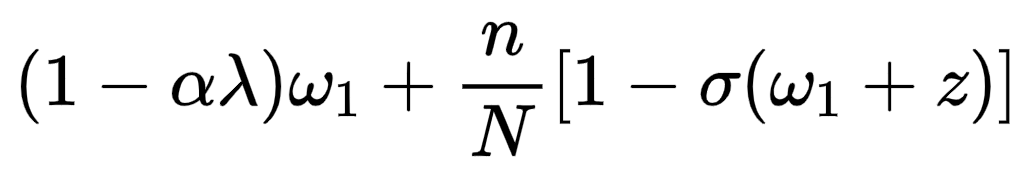
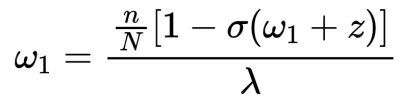






The two lines intersect at the point A(a,a), and we can find that f( )> when is initially less than a, which means that this f( ) will keep increasing until f( ) = , that is, to reach the location of point A.

Similarly, when is initially bigger than a, which means that this f( ) will keep decreasing until f( ) = , that is, to reach the location of point A. Actually, point A（a,a） is value that .



Conclusion: will converge to a point where