# ASSIGNMENT – 2

Saiteja Sirikonda, 1211246826  
[ssirikon@asu.edu](mailto:ssirikon@asu.edu)

1.

A. The number of independent model parameters are, (3n+1). Here for each Xi of a given label we have 2, the prior and since the variance is shared between both class labels.

B. P(Y = 1|X ) = P(Y=1) \* P(X|Y = 1) / (P(Y=1) \* P(X|Y = 1) + P(Y=0) \* P(X|Y = 0))

=

=

Since we are given P(Y=1) = θ,

=

Since we are given that, the P(X­i |Y = k) ∼ N (µik, σik), we have, P(Xi = x | Y = k) = substituting it in the above equation, we get:

So we have, = + and

C. If, = . So irrespective of X, it is equal to P(Y=1)

D. If the equivalence does not hold because we cannot account for the and Xi2

2. The concavity of the log likelihood function is assured if basically the l’’ (w) < 0.

L(w) =

Differentiating w.r.t to w

=>

Differentiating again,

=>

This quantity is always negative, hence it is concave.

3.

A. for L1 distance metric:

For the figure given, if (x,y) is an arbitrary point

If x> 0 and y< 0, it is classified as positive example:  
D+ = |x-1| + |y| = x-1 - y, D- = |x| + |y-1| = x –y + 1, here clearly D+ is clearly smaller than D- due to the given initial condition. Hence positive class

If x< 0 and y> 0, it is classified as negative example:  
D+ = |x-1| + |y| = 1-x + y, D- = |x| + |y-1| = -x + y - 1, here clearly D- is clearly smaller than D+ due to the given initial condition. Hence Negative class

If x < 0 and y < 0, then it is classified as decision boundary:  
D+ = |x-1| + |y| = 1-x-y, D- = |x| + |y-1| = 1-y-x , here clearly D- = D+ due to the given initial condition. Hence it is a decision boundary.

If x > 1 and y > 1, then it is classified as decision boundary:  
D+ = |x-1| + |y| = x-1+y, D- = |x| + |y-1| = x+y-1 , here clearly D- = D+ due to the given initial condition. Hence it is a decision boundary.

If 0<x<1 and 0<y<1, if x> y, then it is classified as positive example:  
D+ = |x-1| + |y| = 1-x + y, D- = |x| + |y-1| = x + 1-y, here clearly D+ is clearly smaller than D- due to the given initial condition. Hence positive class

If 0<x<1 and 0<y<1, if x < y, then it is classified as negative example:  
D+ = |x-1| + |y| = 1-x + y, D- = |x| + |y-1| = x + 1-y, here clearly D- is clearly smaller than D+ due to the given initial condition. Hence negative class

If 0<x<1 and y>1, then it is classified as negative example:  
D+ = |x-1| + |y| = 1-x + y, D- = |x| + |y-1| = x + y - 1, here clearly D- is clearly smaller than D+ due to the given initial condition. Hence Negative class

If 0<y<1 and x>1, then it is classified as positive example:  
D+ = |x-1| + |y| = x -1 + y, D- = |x| + |y-1| = x + 1- y, here clearly D+ is clearly smaller than D- due to the given initial condition. Hence positive class

B. For L2 distance metric:

For the figure given, if (x,y) is an arbitrary point

If x >0 and y > 0 and x < y:  
D+ = sqrt [(x-1)2 + y2] , D- ­= sqrt [(x)2 + (y-1)2], here clearly D- is clearly smaller than D+ due to the given initial condition. Hence Negative class

If x >0 and y > 0 and x > y:  
D+ = sqrt [(x-1)2 + y2] , D- ­= sqrt [(x)2 + (y-1)2], here clearly D+ is clearly smaller than D- due to the given initial condition. Hence positive class

If x< 0 and y> 0:  
D+ = sqrt [(x-1)2 + y2] , D- ­= sqrt [(x)2 + (y-1)2], here clearly D- is clearly smaller than D+ due to the given initial condition. Hence Negative class

If x> 0 and y< 0,:  
D+ = sqrt [(x-1)2 + y2] , D- ­= sqrt [(x)2 + (y-1)2], here clearly D+ is clearly smaller than D- due to the given initial condition. Hence positive class

If x < 0 and y < 0, and x < y:  
D+ = sqrt [(x-1)2 + y2] , D- ­= sqrt [(x)2 + (y-1)2], here clearly D- is clearly smaller than D+ due to the given initial condition. Hence Negative class

If x < 0 and y < 0, and x > y:  
D+ = sqrt [(x-1)2 + y2] , D- ­= sqrt [(x)2 + (y-1)2], here clearly D+ is clearly smaller than D- due to the given initial condition. Hence positive class

C.

For L∞ distance metric,

D+ = max (|x-1|, |y|), D- = max (|x|,|y-1|)

Consider the case if x > y, we can always show that D+ < D- , hence it is classified as the positive class.

And similarly if we consider, y> x, we get D+ > D-  and will hence be classified as negative class.

D.

For L0 distance metric, we have

D+ = NNZ(x-1,y), D­- = NNZ(x,y-1), where NNZ stands for number of non-zero terms in the vector.

If x = 0, y ≠ 0 or 1 - D+ = 2 and D- = 1, all points on the Y-axis are negative class

If y = 0, x ≠ 0 or 1, D+ = 1 and D- = 2, all points on the X-axis are positive class

If x = 0, y = 0 – is the decision boundary

If x = 0, y = 1 – D+ = 2, D- = 0, hence would be classified as negative

If y = 1, x ≠ 0 or 1 - D+ = 2, D- = 1, hence would be a negative boundary

If x = 1, y ≠ 0 or 1 – D+ = 1, D- = 2, would be classified as positive class

If x = 1 and y = 1 – D+ = 1, D- = 1, hence it would be decision boundary

If x = 1, y = 0 – D+ = 0, D- = 2, hence would be positive class.

For every x,y values – it will be a decision boundary.

4. Suppose if we consider the data points closest to the decision boundary, their equations would be,

w.x1 + b = 1, for the data point x1, belonging to the positive class label and   
 w.x2 + b = -1, for the data point x2, belonging to the negative class label

by subtracting the above equation, we get,

* w.(x1 – x2) = 2,
* Taking magnitude on both sides, we get,
* || w ||. m = 2, where m is the size of the margin
* m =

5.

A.

Given, X1 = (-1,-1), y1 = -1 ; X2 = (−1, +1), y2 = +1 ; X3 = (+1, −1), y3 = +1 and X4 = (+1,+1), y4 = -1

Φ(X1) = [1,1,, 1, -

Φ(X2) = [1,1,-, 1, -

Φ(X3) = [1,1,, 1,

Φ(X4) = [1,1,, 1,

K(Xi, Xj) =

|  |  |  |  |
| --- | --- | --- | --- |
| 9 | 1 | 1 | 1 |
| 1 | 9 | 1 | 1 |
| 1 | 1 | 9 | 1 |
| 1 | 1 | 1 | 9 |

B.

J(α) = α1 + α2+ α3 + α4  - ½(9 α12 + 9α12 + 9α12 + 9α12- 2 α1 α2 - 2 α1 α3 + 2α1 α4 + 2α2 α3 - 2α2 α4 - 2α3 α4)

S.T ∑i= 1m αiyi  = 0 and αi >= 0, i = 1,…..k

C.

After simplifying the J(α) given above, neglecting the constant terms it finally turns out to be,

J(α) = -(8 α32 - 2 α3 + 3/32 )

Since we need to maximize the J(α), we differentiate the polynomial on the left hand side, there by getting a value of α3  = 1/8, and therefore we also get, α2 = 1/8.

D.

We compute the values of w vector using the formula,

Wj = ∑ αiyixij

And therefore we finally only get, w = (w1 ,w2 , w3 , w4 , w5 , w6) = (0,0, -1/ , 0,0,0)

When we find the corresponding b values by substituting them the formula,

https://lh6.googleusercontent.com/LlDaxskktBXkiLMz79lAeDv_N57tGmuGwoBmQb30cH7tLmUEcnbQUTXRRX73HPTCGk57AQ16CaG43a9QeS_1kzOXBVXSTynp8SwT0dxoiH48dDOfJVmIDI2NCK6qmAgFsNpWG0ElGt_VYfyyaQ

and take the average of them, we get scalar b value to be 0. (b = 0)

The equation of the decision boundary is w.x + b = 0 => w.x = 0, since b = 0.

7. The problem with choosing k, is, if you pick k to be too small, it is sensitive to noise points. If K is too large, the neighborhood may include points from the wrong class into the decision boundary.

For K-nn constraints given in the problem, we rise from k = 1 to k =3, which has the highest accuracy of around 97.05% and then it decreases from then on. This may be because as the value of k increases we might be including the values from the neighboring classes.

The approximate runtime of the algorithm was around 90 mins, on a PC in cidse which runs on a 3.6Ghz i7 processor and 16GB RAM.