Question 1

We have three training observations , x1 , x2 and x3

X1 represents the x values of the data set while X2 represents the Y values of the data set

Since the data set is lineally separable , we know it has a hyperplane of the form

F(x) = W^T . X – *Y*

Y represents the bias while W presents the weight vector

We know that the function yields a value >= to zero when y takes the value 1 and a value <= zero when y takes the value -1

So were know that training set (1,1) yields a value less that 0 when substituted in f(x).

We also know that the training sets (2,2) and (3,1) yields a value greater that zero when substituted in F(x)

The training sets are separated by the function f(x)= w ^ x - *Y = 0*

We locate the hyperplane that maximizes the separation between the two classes of x

To do this we find the unit vector w that minimizes a cost function ½ (w^t) w = ½ ||w||^ 2 , subject to the constrains

Yi (w^t* xi) >= 1 ; where i= 1,2,3,,,,n

0.2 becomes the value of the unit vector that will maximize the cost fuction

The hyperplane {x : f(x) = W^t*X - y} divides the input space into two , and the sign of f(x)

The decision boundary will be the differentiation between the regions we have classified as positive and negative

Based on the classes of the inputs and the value of the weight vectors ,

The decision boundaries will be

F(x+) = 0.2^T .X –*Y = 1* ,

And F(X.) = 0.2 ^ T .X - Y = -1

Question Two

We want to prove that given x= [x1, x2] ^ *T and a* Feature map Φ (x) = [1, x1^2 √2x1x2 ,x2^2 , √2x1 , √2x2] ^ T when given the kernel function as k(x,z) = *(X^T.Z + 1)^2*

The general form of the feature vector is Φ(·)

We replace the dots with an equivalent kernel function

We know that

Φ(X) = Φ((x1 x2)) = (x1^2 √2x1x2 x2^2)

Such that k(x, x’) = Φ(x)^T Φ(x’)

From the feature map Φ (x) = [1, x1^2 √2x1x2 ,x2^2 , √2x1 , √2x2] ^ T

And the X1 and X2 as the classes of inputs ,

Therefore to find K(x,z),

We Write the general formula of the kernel function as

(Xi ^T.Y + Y1) ^ x

X should represent the number of observation which are two , x and y

The function Y1 takes the value 1

We know that based of the mapping feature , z provides the range of inputs ,

Therefore

K(x,z) = (x^T. the value of y + 1) ^2

= (x^t.z +1)

Part B

We have four observations and two dimensions give as X1 and X2,

The observations (-1, -1) and (1, 1) provide a value of F(x) less than or equal to zero

The observation (-1, 1) and (1,-1) provide a value of F(x) greater than or Zero

The training sets are separated by the function f(x)= w ^ x - *Y = 0*

We locate the hyperplane that maximizes the separation between the two classes of x

To do this we find the unit vector w that minimizes a cost function ½ (w^t) w = ½ ||w||^ 2 , subject to the constrains

Yi (w^t* xi) >= 1 ; where i= 1,2,3,,,,n

We use the kernel valuable as the slack variable

We get the function

$F(X.) = 0.2 ^\wedge T .X - Y + (X^\wedge T.Z + 1) ^\wedge 2$