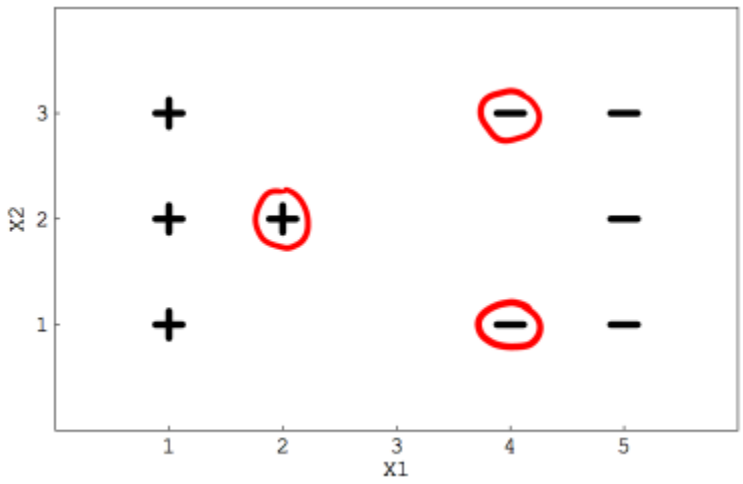
Assignment -5

1. Which of the following is false about support vectors?
   A) The support vectors are the subset of datapoints that determines the max-margin separator.
   B) The Lagrangian multipliers corresponding to the support vectors are non-zero.
   C) The support vectors are used to decide which side of the separator a test case is on.
   D) The max-margin separator is a non-linear combination of the support vectors.

   Answer: D (Please refer to the lecture slides)
2. Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



   1) If you remove the following any one red points from the data. Does the decision boundary will change?

   A) Yes

   B) No

   Solution: A (These three examples are positioned such that removing any one of them introduces slack in the constraints. So the decision boundary would completely change.)
3. Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to

my earlier training set and re-train, in which cases is the learnt decision boundary likely to change?

A) When my model is a perceptron.

B) When my model is logistic regression.

C) When my model is an SVM.

D) When my model is Gaussian discriminant analysis.

Answer: B and D (Please refer to lecture notes)

4. After training an SVM, we can discard all examples which do not support vectors and can still classify new examples?

   A) TRUE

   B) FALSE

   Answer: A (This is true because only support vectors affect the boundary.)

5. If g(z) is the sigmoid function, then its derivative with respect to z may be written in term of g(z) as

   A) g(z)(1-g(z))

   B) g(z)(1+g(z))

   C) -g(z)(1+g(z))

   D) g(z)(g(z)-1)

   Answer: A (Please refer to the lecture notes)

6. Which of the following are true when comparing ANNs and SVMs?

   A) ANN error surface has multiple local minima while SVM error surface has only one minima

   B) After training, an ANN might land on a di erent minimum each time, when initialized with random weights during each run.

   C) In training, ANN's error surface is navigated using a gradient descent technique while SVM's error surface is navigated using convex optimization solvers.

   D) As shown for Perceptron, there are some classes of functions that cannot be learnt by an ANN. An SVM can learn a hyperplane for any kind of distribution.

Answer: A, B, C (By universal approximate theorem, we can argue that option (d) is not true)

7. Which of the following is not a kernel function?

   A) $K(x_i,x_j)=x_i.x_j$

B) $K(x_i,x_j)=(1 - x_i.x_j)^3$

C) $K(x_i,x_j)=e^{(-\|x_i - x_j\|2 /(2\sigma 2))}$

D) $K(x_i,x_j)= \tanh(\beta_0 x_i.x_j+\beta_1)$

Answer: B (The kernel matrix K should be positive semi definite. B may result in negative value.)

8. Which of the following is true about SMO algorithm (multiple answers)?

A) The SMO can efficiently solve the primal problem.

B) The SMO can efficiently solve the dual problem

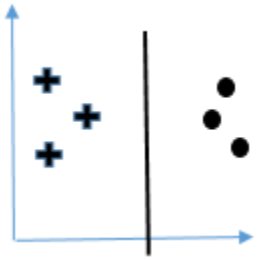
C) The SMO solves the optimization problem by co-ordinate ascent.

D) The SMO solves the optimization problem by co-ordinate descent.

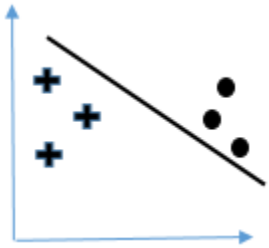
Answer: B, C (Please refer to the lecture notes)

9. For the given set of points which of the following lines is most suitable to be the decision boundary?

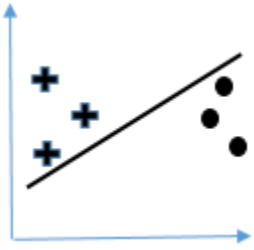
A)

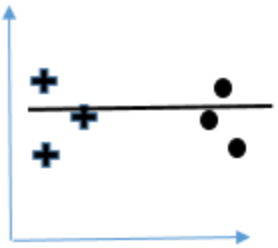


B)



C)



D) 

Answer: A (A has highest margin)

10. Which of the following is/are true about the Perceptron classifier?

A) It can learn a OR function

B) It can learn a AND function

C) The obtained separating hyperplane depends on the order in which the points are presented in the training process.

D) For a linearly separable problem, there exists some initialization of the weights which might lead to non-convergent cases.

Answer: A, B and C

(OR is a linear function, hence can be learnt by perceptron.

XOR is non linear function which cannot be learnt by a perceptron learning algorithm which can learn only linear functions.

The perceptron learning algorithm dependent on the order on which the data is presented, there are multiple possible hyperplanes, and depending on the order we will converge to any one of them.

We can also prove that the algorithm always converges to a separating hyperplane if it exists. Hence D is false.)