1.

Consider the dataset consisting of points (x, y), where x is a real value, and $y \in \{-1,\ 1\}$ is the class label. Let' us start with three points $(x_1, y_1) = (-1,\ -1)$, $(x_2, y_2) = (1, -1)$, $(x_3, y_3) = (0,\ 1)$. [By Hae Jin Song]

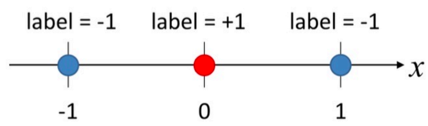


Figure 1: Three data points considered in this problem

(a) Can three points shown in Figure 1, in their current one-dimensional feature space, be perfectly separated with a linear classifier? Explain why or why not.

(b) Now we define a simple feature mapping $\phi(x) = [x, x^2]^T$ to transform the three points from one- to two-dimensional feature space. Plot the transformed points in the new two-dimensional feature space.

Is there a linear decision boundary that can separate the points in this new feature space? Explain why or why not.

(c) Given the feature mapping $\phi(x) = [x, x^2]^T$, write down the kernel function $k(x, x')$.

2.

We have trained a support vector machine with Laplace kernel determined by $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|_2)$. The support vectors are $\mathbf{x}_1^T = [2\ 1\ 0], \mathbf{x}_2^T = [-1\ 1\ 0], \mathbf{x}_3^T = [-5\ 0\ 0]$, whose corresponding α_i's are $\alpha_1 = -2, \alpha_2 = 3, \alpha_3 = -1$. Also, $\beta_0 = 1$. As a reminder, the discriminant function of a SVM can be written in terms of the α_i's and β_0. Determine the class to which the test vector $\mathbf{x}^T = [0\ 0\ 0]$ is classified.

3.

Consider the following data set: In class 1, we have $[1\ 1]^T, [1\ 2]^T, [2\ 2]^T$. In class 2, we have $[1.5\ 1.5]^T$.

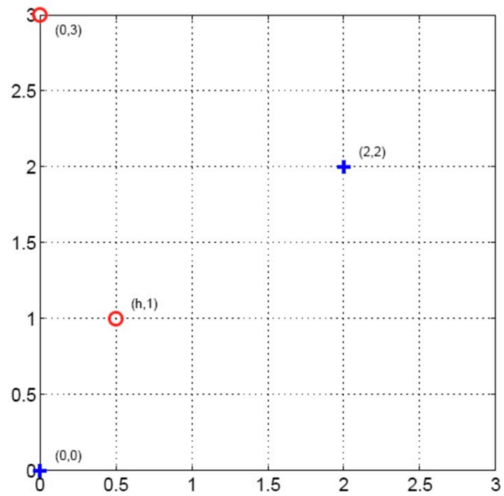
(a) Sketch the data set and determine whether or not it is linearly separable.

(b) Regardless of the answer to 2a, find a quadratic feature $X_3 = f(X_1, X_2) = aX_1^2 + bX_2^2 + cX_1X_2 + dX_1 + eX_2 + f$, that makes the data linearly separable; that is, $X_3 \geq 0$ for members of class 1, and $X_3 < 0$ for members of class 2. Find the maximum margin classifier only based on X_3. Hint: The equation of the maximum margin classifier based on only one feature is $X_3 = \beta_0$ and you should determine β_0.

(c) By solving $X_3 = f(X_1, X_2) = \beta_0$ for X_2, find the equation of the decision boundary in the original feature space and sketch it. Show the regions in the feature space that are classified as class 1 and class 2. You do not need to be very precise.

4.

3. Suppose we only have four training observations in two dimensions:



positive examples are $\mathbf{x}_1 = [0\ 0]^T, \mathbf{x}_2 = [2\ 2]^T$ and negative examples are $\mathbf{x}_3 = [h\ 1]^T, \mathbf{x}_4 = [0\ 3]^T$. h is a parameter.

(a) What is the largest value of h for which the training data are still linearly separable?

(b) Determine the support vectors when $h = 0.5$.

(c) When the training points are separable, does the slope of the maximum margin classifier change? Why?

(d) Assume that $h = .5$ and we have unlabeled data $\mathbf{x}_5 = [3\ 3]^T, \mathbf{x}_6 = [2\ 0.5]^T, \mathbf{x}_7 = [1\ 1.5]^T, \mathbf{x}_8 = [2.5\ 1.5]^T$. Which one will be labeled first, if we are performing self-training? Which one will be labeled first, if we are performing active learning?

5.

We wish to classify the following two dimensional dataset using a Maximum Margin Classifier:

$$\left\{\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right\}, \text{ with } y = 1 \text{ and } \left\{\mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}\right\} \text{ with } y = -1.$$

(a) Determine the equation of the linear decision boundary $f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

(b) Determine the set of support vectors $\mathcal{S}$.

Hint: you do not need to solve an optimization problem to answer this question.

6.

Consider a dataset with 3 points in 1-D:

Index	X	Y
1	0	$+$
2	-1	-
3	$+1$	-

(a) Carefully sketch these three training points. Are the classes linearly separable?

(b) Consider mapping each point to 2-D using new feature vectors $\boldsymbol{\varphi}(x) = [u_1(x), u_2(x)]^T$, in which $u_1(x)$ and $u_2(x)$ are polynomial functions of x. Find a $\boldsymbol{\varphi}(x)$ such that data are linearly separable in the new feature space.

(c) The maximum margin classifier in the new space has the equation $\mathbf{w}^T\boldsymbol{\varphi}(x)+b = 0$. Find $\mathbf{w}$ and b. Determine the decision boundary in the original 1-D space and the class assigned to $x = \frac{1}{3}$ by the classifier.

7.

Consider the XOR classsification problem (25 pts):

Index	X_1	X_2	Y
1	-1	-1	-1
2	-1	1	1
3	1	-1	1
4	1	1	-1

(a) Find a single feature $Z = f(X_1, X_2)$ that is a quadratic function and makes the data linearly separable. Find the maximum margin classifier using this single feature and project its decision boundary back onto the (X_1, X_2) space and sketch the decision boundary in this space.

(b) Show that adding adding the feature $X_3 = |X_1 - X_2|$ makes the data linearly separable. Find the maximum margin classifier in the space (X_1, X_2, X_3) for this data set and project it back onto (X_1, X_2) and sketch the decision boundary in this space.

8.

Consider the following data set (25 pts).

Index	X_1	X_2	Y
1	-2	0	Cross
2	-1	-1	Circle
3	-3	0	Cross
4	5	-2	Circle
5	0	-1	Circle
6	0.5	4	Cross
7	1	0	Circle
8	0	2	Cross

(a) Sketch or plot the data.

(b) Sketch or plot the maximum margin classifier and the margins.

(c) Find the equation of the maximum margin classifier and those of the margins.

(d) Provide the decision rule: If $\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$, classify to ..., otherwise, classify to ....

(e) Remove Observations 1,7. Does the equation of the separating hyperplane change?

(f) Write the equation of the separating hyperplane in 2e as $f(\mathbf{x}) = \sum_{i \in \mathcal{S}} \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle + \beta_0$ and determine α_i's.