

CS3244 Machine Learning

Homework 2 (Essay Questions)

Matric No. - A0153196B

October 4, 2016

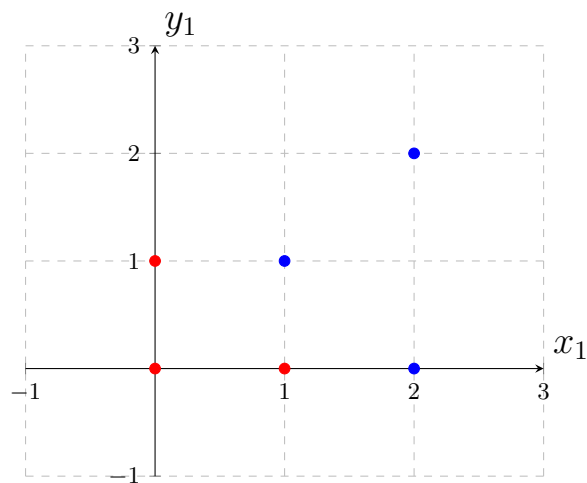
2.

Consider the following training data:

x_1	x_2	y
1	1	+
2	2	+
2	0	+
0	1	-
1	0	-
0	0	-

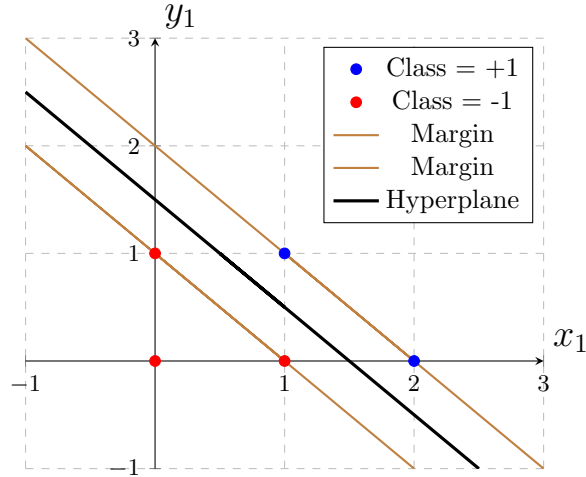
a. Plot these six training points. Are the classes $\{+, \}$ linearly separable?

ANS. Yes these classes are linearly separable. The plot is shown below :-



b. Construct the weight vector of the maximum margin hyper-plane by inspection and identify the support vectors.

ANS. The margin and hyperplane are given below :-



The equations of the margin, hyperplane, weight vector and support vectors are as follows:-

Margin (Class +1): $x_1 + x_2 - 2 = 0$

Margin (Class -1): $x_1 + x_2 - 1 = 0$

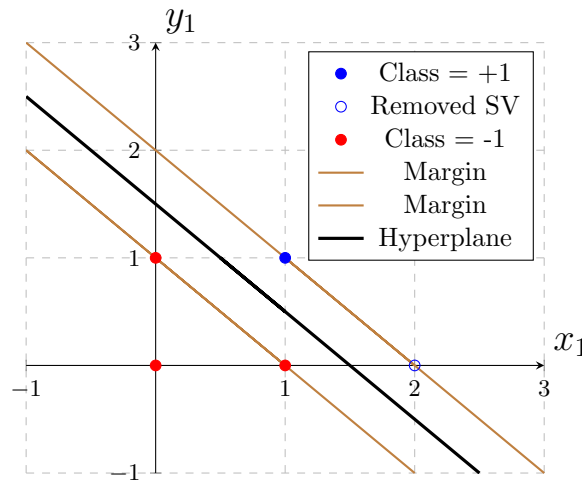
Hyperplane: $x_1 + x_2 - 1.5 = 0$

Normalized weight vector: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Support Vectors: $\begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

c. If you remove one of the support vectors, does the size of the optimal margin decrease, stay the same, or increase? Explain.

ANS. If we remove one of the support the margin will still **stay the same** because no matter which support vector we remove, the optimal margin is still passing through other three support vectors which will keep it intact. Since a support vector is the nearest point to the hyperplane, removing it means that we have to consider the next nearest point in our dataset which can only increase the margin or keep it fixed depending on the presence of other support vectors. A supporting figure is shown below:-



We can see in the above figure that the support vector at (2,0) has been removed but the margin stays the same because of the presence of the vector at (1,1).

d. Is your answer to (c) also true for any dataset? Provide a counterexample or give a short proof.

ANS. The answer to (c) is not true for every dataset. There can be cases when the margin increases. However, it will never decrease in case of a hard margin svm. We can show this with the help of a counterexample. Consider a dataset which consists of 3 points shown below:-

x_1	x_2	y
1	0	-
1	1	+
2	2	+

The plot of this data is given below. It has 2 support vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Now if we remove the support vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ we will get the figure shown below.

3.

A kernel is valid if it corresponds to a scalar product in some (perhaps infinite dimensional) feature space. Remember a necessary and sufficient condition for a function $K(\mathbf{x}, \mathbf{x})$ to be a valid kernel is that associated Gram matrix, whose elements are given by $k(\mathbf{x}_n, \mathbf{x}_m)$, should be positive semi-definite for all possible choices of the set \mathbf{x} . Show whether the following are also kernels:

Valid kernels follow the following properties which we are going to use in our proof:-

Property 1: Given two valid kernels $K_1(-, -)$ and $K_2(-, -)$:
 $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$

Property 2: Given two valid kernels $K_1(-, -)$ and $K_2(-, -)$:
 $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') * K_2(\mathbf{x}, \mathbf{x}')$

Property 3: Given a valid kernel $K_1(-, -)$:
 $K(\mathbf{x}, \mathbf{x}') = q[K_1(\mathbf{x}, \mathbf{x}')]$
 where $q[\cdot]$ is a polynomial function with non-negative coefficients.

1. $K(\mathbf{x}, \mathbf{x}') = c\langle \mathbf{x}, \mathbf{x}' \rangle$

ANS.

First of all, $\langle \mathbf{x}, \mathbf{x}' \rangle$ is a valid kernel since it represents a scalar product in the Euclidean space. Inner products are **symmetric** and their Gram Matrices are **positive semi-definite** which are the essential conditions of being a valid kernel.

Assuming c is non negative i.e. $c > 0$ we can use *Property 3* here as $c\langle \mathbf{x}, \mathbf{x}' \rangle$ is a polynomial function of $\langle \mathbf{x}, \mathbf{x}' \rangle$ with degree 1 and non negative coefficients (c).

$$\text{Let } K_1(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$$

$$\therefore q[K_1(\mathbf{x}, \mathbf{x}')] = q[\langle \mathbf{x}, \mathbf{x}' \rangle] = c\langle \mathbf{x}, \mathbf{x}' \rangle \quad \text{where } c > 0$$

Since $\langle \mathbf{x}, \mathbf{x}' \rangle$ is a valid kernel, $q[\langle \mathbf{x}, \mathbf{x}' \rangle]$ is also a valid kernel.

$\therefore c\langle \mathbf{x}, \mathbf{x}' \rangle$ is a valid kernel.

$$2. K(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle^2 + e^{(-\|\mathbf{x}\|^2)} e^{(-\|\mathbf{x}'\|^2)}$$

ANS.

Again as a base case $\langle \mathbf{x}, \mathbf{x}' \rangle$ is a valid kernel due to reasons specified in the previous question. We will prove that the given kernel function is valid by first showing that the individual components are valid:-

Proof for $\langle \mathbf{x}, \mathbf{x}' \rangle^2$:

$$\text{Let } K_1(\mathbf{x}, \mathbf{x}') = K_2(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$$

Since $\langle \mathbf{x}, \mathbf{x}' \rangle$ is valid $K_1(\mathbf{x}, \mathbf{x}')$ and $K_2(\mathbf{x}, \mathbf{x}')$ are also valid.

$$K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') * K_2(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle^2$$

Using **Property 2**, since $K_1(\mathbf{x}, \mathbf{x}')$ and $K_2(\mathbf{x}, \mathbf{x}')$ are valid $K(\mathbf{x}, \mathbf{x}')$ is also valid

$$\therefore \langle \mathbf{x}, \mathbf{x}' \rangle^2 \text{ is a valid kernel}$$

Proof for $e^{(-\|\mathbf{x}\|^2)} e^{(-\|\mathbf{x}'\|^2)}$:

We know that a kernel is valid if it is **symmetric** and **positive semidefinite**. Let's prove these properties individually:-

- Symmetry:

$$\text{Let } K(\mathbf{x}, \mathbf{x}') = e^{(-\|\mathbf{x}\|^2)} e^{(-\|\mathbf{x}'\|^2)}$$

$$K(\mathbf{x}, \mathbf{x}') = e^{(-\|\mathbf{x}\|^2)} e^{(-\|\mathbf{x}'\|^2)} = e^{(-\|\mathbf{x}'\|^2)} e^{(-\|\mathbf{x}\|^2)} = K(\mathbf{x}', \mathbf{x})$$

$$\therefore e^{(-\|\mathbf{x}\|^2)} e^{(-\|\mathbf{x}'\|^2)} \text{ is symmetric.}$$

- Positive semi-definiteness:

According to **Mercer's Theorem**:

A symmetric function $K(x, y)$ can be expressed as an inner product

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$$

for some mapping ϕ if and only if $K(\mathbf{x}, \mathbf{x}')$ is positive semidefinite

Thus, if we are able to find a mapping ϕ for the given function, it means that the function is positive semidefinite according to Mercer's theorem.

$$\text{Consider } \phi(x) = e^{(-\|\mathbf{x}\|^2)}$$

$$\text{Then } K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle = \langle e^{(-\|\mathbf{x}\|^2)}, e^{(-\|\mathbf{x}'\|^2)} \rangle = e^{(-\|\mathbf{x}\|^2)} e^{(-\|\mathbf{x}'\|^2)}$$

A symmetric function $K(\mathbf{x}, \mathbf{x}')$ is a kernel is valid if it can be expressed as an inner product in some feature space. In other words:-

$$K(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$$