CSCE 633 - Homework 3

Prakhar Suryavansh

Math: Solution 1

We are given $n = 7$ observations in $p = 2$ dimensions with associated class labels.
We assign the class labels Y as follows:

- $Y = +1$ for the "Red" class

- $Y = -1$ for the "Blue" class

The dataset is given as:

Index	X_1	X_2	Y
1	3	6	-1
2	2	2	-1
3	4	4	-1
4	1	3	-1
5	2	0	$+1$
6	4	2	$+1$
7	4	0	$+1$

Part (a):

To find the optimal separating hyperplane for a maximal margin classifier, we proceed as follows:

The maximal margin hyperplane depends directly on the points that lie on the margin, known as "support vectors". For a training set $D^{train} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, we have:

- Input vectors: $\mathbf{x}_1, ..., \mathbf{x}_N \in \mathbb{R}^D$

- Class labels: $y_1, ..., y_N \in \{-1, +1\}$

The optimization problem is formulated as:

$$\min_{w} \frac{1}{2} w^T w, \text{ such that } y_i(w^T x_i + w_0) \geq 1 \text{ for } i = 1, 2, \cdots, N$$

To solve this constrained optimization problem, we use the method of Lagrange multipliers:

1. The Lagrangian function (primal problem) is:

$$L = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{N} \alpha_i(y_i(w^T x_i + w_0) - 1)$$

2. Minimizing the Lagrangian with respect to the primal variables:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial w_0} = 0 \rightarrow 0 = \sum_{i=1}^{N} \alpha_i y_i$$

3. Substituting these back into the Lagrangian gives the dual form:

$$L = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{i'=1}^{N} \alpha_i \alpha_{i'} y_i y_{i'} x_i^T x_{i'}$$

4. The dual problem becomes:

$$\max_{\alpha_i} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{i'=1}^{N} \alpha_i \alpha_{i'} y_i y_{i'} x_i^T x_{i'} \right\}$$

subject to:

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0$$

5. The kernel matrix $K = x_i^T x_j$ for our data points is:

$$K = \begin{bmatrix} 45 & 18 & 30 & 21 & 6 & 20 & 12 \\ 18 & 8 & 16 & 9 & 4 & 12 & 8 \\ 30 & 16 & 32 & 15 & 8 & 20 & 16 \\ 21 & 9 & 15 & 10 & 2 & 10 & 8 \\ 6 & 4 & 8 & 2 & 4 & 12 & 8 \\ 20 & 12 & 20 & 10 & 12 & 20 & 16 \\ 12 & 8 & 16 & 8 & 8 & 16 & 16 \end{bmatrix}$$

After solving using quadratic optimization, we obtained:

- Weight vector $w = [1, -1]^T$

- Bias term $w_0 = -1$

- Optimal alpha values: $[0, 0.745, 0.255, 0, 0.245, 0.755, 0]$

Therefore, the equation for the optimal separating hyperplane is:

$$-1 + 1.000x_1 - 1.000x_2 = 0$$

Sketch for this optimal separating hyperplane is below:

Part (b):

The classification rule for the maximal margin classifier is:

Classify to Red if $-1 + 1.000x_1 - 1.000x_2 \geq 0$, and classify to Blue otherwise.

Part (c):

The margin width is $\frac{2}{\|w\|} = \frac{2}{\sqrt{2}} \approx 1.414$
The margin for the maximal margin hyperplane is as indicated in the figure below:

Part (d):

The Optimal alpha values we obtained were: $[0, 0.745, 0.255, 0, 0.245, 0.755, 0]$
We see that the alpha values that are non-zero are:

$$\alpha_2, \alpha_3, \alpha_5 \text{ and } \alpha_6$$

Therefore, the set of x_i whose $\alpha_i \neq 0$ are the support vectors, which are:
(2, 2), (4, 4), (2, 0) and (4, 2)
The support vectors for the maximal margin classifier are shown in the figure below:

Part (e):

A slight movement of the seventh observation will not affect the maximal marginal hyperplane because it is not a support vector. So, as long as it remains outside the margin boundary, it will not affect the maximal marginal hyperplane. Only observations on the margin boundary (support vectors) influence the hyperplane.

Part (f):

Since, there is only one optimal separating hyperplane, any other hyperplane that divides the given dataset would be non optimal. So, an alternative hyperplane that is not the optimal separating hyperplane could be as shown in the figure below.
The equation for this hyperplane is:

$$-1.5 + 1.2X_1 - 1.1X_2 = 0$$

Part (g):

Adding a new point, for example, $(2, 5)$ with class "Red" (i.e., $y = +1$), would make the two classes non-separable by a single hyperplane.