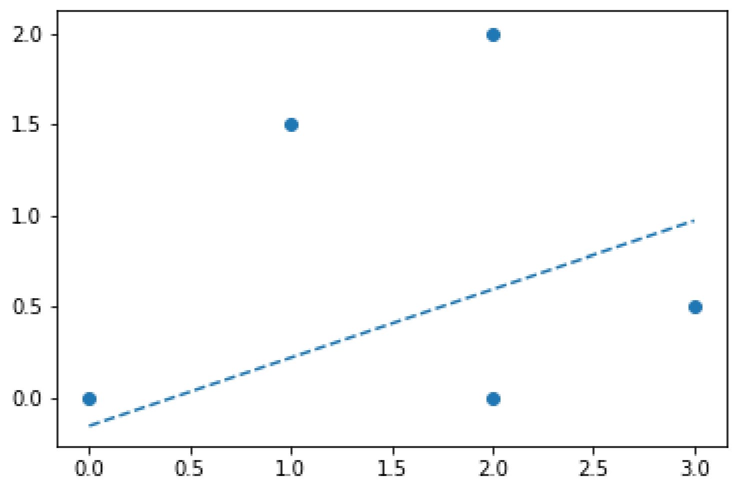
Theodore Kim
CS-UY 4563 – Introduction to Machine Learning
HW #5

1. Consider the training data:

| $x_1$ | $x_2$ | y |
|---|---|---|
| 0 | 0 | -1 |
| 2 | 2 | -1 |
| 2 | 0 | 1 |
| 1 | 1.5 | -1 |
| 3 | 0.5 | 1 |

(a) Plotted data with hyperplane: **w** = [12, -32], $\mathbf{w_0}$ = -5




(b) The data **is** linearly separable

(c) Given the above values

**i. Determine the functional margin:**

$$\gamma = \min_{i=1\ldots n} \gamma_i = \min_{i=1\ldots n} y_i(w^T x_i + w_0)$$

$$\gamma = \min_{i=1\ldots n} y_i([12 \quad -32]x_i - 5)$$

$$\gamma = \min\left\{-1\left([12 \quad -32]\begin{bmatrix}0\\0\end{bmatrix} - 5\right), -1\left([12 \quad -32]\begin{bmatrix}2\\2\end{bmatrix} - 5\right), 1\left([12 \quad -32]\begin{bmatrix}2\\0\end{bmatrix} - 5\right), -1\left([12 \quad -32]\begin{bmatrix}1\\1.5\end{bmatrix} - 5\right), 1\left([12 \quad -32]\begin{bmatrix}3\\0.5\end{bmatrix} - 5\right)\right\}$$

$$\gamma = \min\{-5, -45, 19, -41, 15\} = \mathbf{-5}$$

**ii. Determine the geometric margin:**

$$\gamma = \min_{i=1...n} \gamma_i = \min_{i=1...n} \frac{y_i(w^T x_i + w_0)}{\|w\|_2}$$

$$\gamma = \min_{i=1...n} \frac{y_i([12 \quad -32]x_i - 5)}{\sqrt{(12^2 + (-32)^2}} = \min_{i=1...n} \frac{y_i([12 \quad -32]x_i - 5)}{34.17}$$

$$\gamma = \min\left\{-\frac{5}{34.17}, -\frac{45}{34.17}, \frac{19}{34.17}, \frac{-41}{34.17}, \frac{15}{34.17}\right\} = \mathbf{-0.1463}$$

**iii. compute the canonical weights with respect to the training example**

The canonical weights with respect of the training set is the value for **w** and **w₀** such that the functional margin is 1.

Therefore, the functional margin was previous **5** therefore divide w by 5 to get the canonical weights:

$w' = \frac{1}{5} w = \begin{bmatrix} \mathbf{\frac{12}{5}} \\ \mathbf{-\frac{32}{5}} \end{bmatrix}$ and $w_0' = \frac{1}{5} w = \mathbf{-1}$

<u>(d) Identify which of the training examples are support vectors</u>

The training samples closest to the hyperplane are the support vectors, therefore the support vectors in this example are: (0, 0) on the negative side and (3, 0.5) of the positive side.

<u>(e)</u> Adding the point x = (1, 3) and y = -1 **does not** change the margin, hyperplane or support vector as it lies outside of the maximum margin as its functional margin is not equal to -1.

$$\gamma_i = y_i(w'^T x_i + w'_0)$$

$$\gamma_i = y_i\left(\begin{bmatrix} \frac{12}{5} & -\frac{32}{5} \end{bmatrix} x_i - 1\right)$$

$$\gamma_i = -1\left(\begin{bmatrix} \frac{12}{5} & -\frac{32}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} - 1\right)$$

<u>(f)</u> Removing the point x = (1, 1.5) **does not** change the margin, hyperplane or support vector as the point previously lay outside of the maximum margin as its functional margin was not equal to -1.

<u>(g)</u> Removing the point x = (0, 0) **does** change the support vector as it then moves it outwards to the next closest point: (1, 1.5). In response the maximum margin increases (becomes wider) as both

support vectors are now farther apart. Finally the separating hyperplane shifts in response to the wider maximum margin.

(h) Define a constraint optimization problem:

$\min w_1^2 + w_2^2$

Subject to $(-w_0) \geq 1$

$$-1(2w_1 + 2w_2 + w_0) \geq 1$$

$$(2x_1 + w_0) \geq 1$$

$$-1(w_1 + 1.5w_2 + w_0) \geq 1$$

$$(3w_1 + 0.5w_2 + w_0) \geq 1$$

2. Consider the training data:

| $x_1$ | y |
|---|---|
| 1 | -1 |
| 2 | 1 |
| 3 | 1 |
| 4 | -1 |

Create new features using the Gaussian Radial Basis Function, let σ = 1

$$\varphi(x) = K(x, x_i) = \exp\left(-\frac{\left\|x - l_j\right\|_2^2}{2\sigma^2}\right) = \exp\left(-\frac{\left\|\begin{bmatrix}1\\2\\3\\4\end{bmatrix} - l_j\right\|_2^2}{2}\right)$$

For all values $l \in \{l_1 = x_1, l_2 = x_2, \ldots l_n = x_n\}$

| $x_1$ | y | $z_1$ | $z_2$ | $z_3$ | $z_4$ |
|---|---|---|---|---|---|
| 1 | -1 | 1.00000 | 0.60653 | 0.13534 | 0.01111 |
| 2 | 1 | 0.60653 | 1.00000 | 0.60653 | 0.13534 |
| 3 | 1 | 0.13534 | 0.60653 | 1.00000 | 0.60653 |
| 4 | -1 | 0.01111 | 0.13534 | 0.60653 | 1.00000 |

The coefficients calculated using hard margin SVM on the above features:

$$w = \begin{bmatrix} -0.26924 \\ 0.86466 \\ 0.86466 \\ -0.26924 \end{bmatrix}, w_0 = -0.77929758$$

3. Running the hard margin SVM on the added data below using the weights would get the predictions:

For all values $l \in \{x_{11}, x_{12}, x_{13}, ((x_{12} + x_{13})/2)\}$, z has been calculated using a radial basis function kernel

| $x_1$ | $z_1$ | $z_2$ | $z_3$ | $z_4$ | $\hat{y}$ |
|---|---|---|---|---|---|
| 1.2 | 1 | 0.429557 | 0.122303 | 0.245904 | -1 |
| 2.5 | 0.429557 | 1 | 0.75484 | 0.932102 | 1 |
| 3.25 | 0.122303 | 0.75484 | 1 | 0.932102 | 1 |

3. Would the two constrained optimizations create the same decision boundary?

**The second one would result in the same decision boundary, while the first one won't because, rather than creating support vectors of functional margins of one, it is creating a line which achieves the maximum marginal SVM.**