Problem 4

Solve SVM for a data set with 3 data instances in 2 dimensions: (1,1, +), (-1,1, -), (0,-1, -).Here the first 2 number are the 2 dimension coordinates. '+' in 3rd place is positive class. And '-' in 3rd place is negative class. Your task is to compute alpha's, w, b.

Solution

We know that, the Lagrange multiplication for SVM is given by:

$$\max L_D(\alpha) = \frac{1}{2} * \sum_i \sum_j \alpha_i \alpha_j y_i y_j \overrightarrow{x_i x_j} - \sum_i \alpha_i$$

$$such\ that \sum_i \alpha_i y_i \neq 0 \quad \dots (1)$$

$$and\ \alpha_i \geq 0$$

And

$$w = \sum_i \alpha_i y_i x_i$$

$$y = wx + b$$

Solving for the above points:

$$L = \frac{1}{2}\{\alpha_1{}^2 y_1{}^2 x_1{}^T x_1 + \alpha_2{}^2 y_2{}^2 x_2{}^T x_2 + \alpha_3{}^2 y_3{}^2 x_3{}^T x_3 + 2\alpha_1\alpha_2 y_1 y_2 x_1{}^T x_2 + 2\alpha_1\alpha_3 y_1 y_3 x_1{}^T x_3 + 2\alpha_2\alpha_3 y_2 y_3 x_2{}^T x_3\} - (\alpha_1 + \alpha_2 + \alpha_3)$$

Substituting respective values:

$$L = \frac{1}{2}\{(\alpha_1{}^2 * 1^2 * 2) + (\alpha_2{}^2 * -1^2 * 2) + (\alpha_3{}^2 * -1^2 * 1) + (2 * \alpha_1\alpha_2 * 1 * -1 * 0) + (2 * \alpha_1\alpha_3 * 1 * -1 * -1) + (2 * \alpha_2\alpha_3 * -1 * -1 * -1)\} - (\alpha_1 + \alpha_2 + \alpha_3)$$

$$L = \frac{1}{2}\{(2\alpha_1{}^2) + (2\alpha_2{}^2) + (\alpha_3{}^2) + (0) + (2\alpha_1\alpha_3) + (-2\alpha_2\alpha_3)\} - (\alpha_0 + \alpha_1 + \alpha_2)$$

$$L = \alpha_1{}^2 + \alpha_2{}^2 + \frac{\alpha_3{}^2}{2} + \alpha_1\alpha_3 - \alpha_2\alpha_3 - \alpha_0 - \alpha_1 - \alpha_2$$

Taking derivatives with respect to alpha:

$$\frac{dL}{d\alpha_1} = 2\alpha_1 + \alpha_3 - 1 = 0 \quad => \quad 2\alpha_1 + \alpha_3 = 1 \quad \dots (2)$$

$$\frac{dL}{d\alpha_2} = 2\alpha_2 - \alpha_3 - 1 = 0 \quad => \quad 2\alpha_2 - \alpha_3 = 1 \quad \dots (3)$$

$$\frac{dL}{d\alpha_3} = \alpha_1 - \alpha_2 + \alpha_3 - 1 = 0 \quad => \quad \alpha_1 - \alpha_2 + \alpha_3 = 1 \quad \dots (4)$$

From (1)

$$\alpha_1 - \alpha_2 - \alpha_3 = 0 \quad => \quad \alpha_1 = \alpha_2 + \alpha_3 \quad \dots (5)$$

Naman Jain Vimal Kumar
1001670153

Case 1: Solving 2, 3 and 5:

- Substituting 5 in 2:

$$2(\alpha_2 + \alpha_3) + \alpha_3 = 1$$
$$2\alpha_2 + 3\alpha_3 = 1$$

- Subtracting 3 from above

$$(2\alpha_2 + 3\alpha_3) - (2\alpha_2 - \alpha_3) = 1 - 1$$
$$\alpha_3 = 0$$

- Substituting this in 3:

$$2\alpha_2 - 0 = 1$$
$$\alpha_2 = \frac{1}{2}$$

- Substituting these in 5:

$$\alpha_1 = \alpha_2 + \alpha_3 = \frac{1}{2} + 0 = \frac{1}{2}$$

- Therefore, $\alpha = \left[\frac{1}{2}, \frac{1}{2}, 0\right]$
- Solving for w:

$$w = \left[\frac{1}{2} * 1 * (1,1)\right] + \left[\frac{1}{2} * -1 * (-1,1)\right] + [0 * -1 * (0,1)]$$
$$w = (1, 0)$$
$$\|w\| = 1$$

Case 2: Solving 2, 4 and 5:

- Substituting 5 in 2:

$$2(\alpha_2 + \alpha_3) + \alpha_3 = 1$$
$$2\alpha_2 + 3\alpha_3 = 1 \qquad (a)$$

- Substituting 5 in 4:

$$\alpha_2 + \alpha_3 - \alpha_2 + \alpha_3 = 1$$
$$\alpha_3 = \frac{1}{2}$$

- Substituting this in (a):

$$2\alpha_2 + \frac{3}{2} = 1$$

$\alpha_2 = -\frac{1}{4}$ Discarding this case, as it is negative

Case 3: Solving 3, 4 and 5:

- Substituting 5 in 3:

$$2(\alpha_2 + \alpha_3) - \alpha_3 = 1$$
$$2\alpha_2 - \alpha_3 = 1 \qquad (a)$$

- Substituting 5 in 4:

$$\alpha_2 + \alpha_3 - \alpha_2 + \alpha_3 = 1$$
$$\alpha_3 = \frac{1}{2}$$

- Substituting this in (a):

$$2\alpha_2 - \frac{3}{2} = 1$$
$$\alpha_2 = \frac{5}{4}$$

- Substituting these in 5:

$$\alpha_1 = \alpha_2 + \alpha_3 = \frac{1}{2} + \frac{5}{4} = \frac{7}{4}$$

- Therefore, $\alpha = \left[\frac{7}{4}, \frac{5}{4}, \frac{1}{2}\right]$
- Solving for w:

$$w = \left[\frac{7}{4} * 1 * (1,1)\right] + \left[\frac{5}{4} * -1 * (-1,1)\right] + \left[\frac{1}{2} * -1 * (0,1)\right]$$
$$w = (3, 0)$$
$$\|w\| = 3$$

Considering Case 1 as 'w' should be minimized. Therefore w = (1, 0)

Solving for b:
We know that: y = wx + b

For Point 1:

$$+1 = (1,0)(1,1) + b$$
$$b = 0$$

For Point 2:

$$-1 = (1,0)(-1,1) + b$$
$$b = 0$$

For Point 3:

$$-1 = (1,0)(0,-1) + b$$
$$b = -1$$

Thus, point 1 and 2 are the support vectors for the given problem.

Problem 5

Solve SVM when data are non-separable, using k=2 when minimizing the violations of the misclassification, i.e., on those slack variables.

Solution

We need to minimize the weights and penalty for violating classification constrains

$$\min \left(\frac{\|w\|^2}{2} + C \sum_{i}^{n} {\zeta_i}^k \right)$$

subject to

$$y_i[wx_i + b] \geq 1 - \zeta_i$$
$$\zeta_i \geq 0$$

where ζ_i are slack variables

Therefore, by Lagrange multiplication rule:

$$L = \frac{\|w\|^2}{2} + C \sum_i^n {\zeta_i}^2 - \sum_i^n \alpha_i \{y_i[wx_i + b] - 1 + \zeta_i\} - \sum_i^n \mu_i \zeta_i \quad \text{...(1)}$$

Thus,

$$\frac{dL}{dw} = w + 0 - \sum_i^n \alpha_i\, y_i x_i = 0$$
$$w = \sum_i^n \alpha_i\, y_i x_i \qquad \text{... (2)}$$

$$\frac{dL}{db} = 0 + 0 - \sum_i^n \alpha_i\, y_i b = 0$$
$$\sum_i^n \alpha_i\, y_i = 0 \qquad \text{...(3)}$$

$$\frac{dL}{d\zeta_i} = 0 + 2C \sum_i^n \zeta_i - \sum_i^n \alpha_i\, \zeta_i - \sum_i^n \mu_i \zeta_i = 0$$
$$\sum_i^n \alpha_i\, \zeta_i + \sum_i^n \mu_i \zeta_i = 2C \sum_i^n \zeta_i \quad \text{...(4)}$$

Substituting (2) in (1):

$$L = \frac{1}{2} \sum_i^n \alpha_i y_i x_i \sum_j^n \alpha_j y_j x_j + C \sum_i^n {\zeta_i}^2 - \sum_i^n \alpha_i \left\{ y_i \left[\sum_j^n \alpha_j y_j x_j\, x_i + b \right] - 1 + \zeta_i \right\} - \sum_i^n \mu_i \zeta_i$$

$$L = \frac{1}{2} \sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j x_i x_j + C \sum_i^n {\zeta_i}^2 - \sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j x_i x_j - \sum_i^n \alpha_i y_i b + \sum_i^n \alpha_i - \sum_i^n \alpha_i \zeta_i - \sum_i^n \mu_i \zeta_i$$

From (3): $\sum_i^n \alpha_i\, y_i b = 0$

$$L = -\frac{1}{2} \sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j x_i x_j + C \sum_i^n {\zeta_i}^2 - 0 + \sum_i^n \alpha_i - \left(\sum_i^n \alpha_i \zeta_i + \sum_i^n \mu_i \zeta_i \right)$$

From (4): Substituting $\sum_i^n \alpha_i\, \zeta_i + \sum_i^n \mu_i \zeta_i = 2C \sum_i^n \zeta_i$

$$L = \sum_i^n \alpha_i - \frac{1}{2} \sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j x_i x_j + C \sum_i^n {\zeta_i}^2 - 2C \sum_i^n \zeta_i$$

Naman Jain Vimal Kumar
1001670153

$$L = \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} \alpha_i \alpha_j y_i y_j x_i x_j + C \sum_{i}^{n} \zeta_i \left(\sum_{i}^{n} \zeta_i - 2 \right)$$

References,

- Probability and Statistics for Computer Scientists, Second Edition, Michael Baron.
- http://www.ai.mit.edu/courses/6.867-f04/lectures/lecture-7-ho.pdf