Machine Learning
Assignment 2

Naman Jain Vimal Kumar
1001670153

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:

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

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

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

We know that,

$$\alpha_1 - \alpha \quad - \alpha_3 = 0 \quad => \quad \alpha_1 = \alpha_2 + \alpha_3$$

Therefore,

$$L = (\alpha_2 + \alpha_3)^2 + \alpha_2{}^2 + \frac{\alpha_3{}^2}{2} + (\alpha_2 + \alpha_3)\alpha_3 - \alpha_2\alpha_3 - (\alpha_2 + \alpha_3) - \alpha_2 - \alpha_3$$

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

$$L = 2\alpha_2{}^2 + \frac{5}{2}\alpha_3{}^{\;2} + 2\alpha_2\alpha_3 - 2\alpha_2 - 2\alpha_3$$

Taking derivatives with respect to alpha

Naman Jain Vimal Kumar
1001670153

$$\frac{dL}{d\alpha_2} = 4\alpha_2 + 2\alpha_3 - 2 = 0 \qquad => \qquad 2\alpha_2 + \alpha_3 = 1 \qquad \ldots(3)$$

$$\frac{dL}{d\alpha_3} = 2\alpha_2 + 5\alpha_3 - 2 = 0 \qquad => \qquad 2\alpha_2 + 5\alpha_3 = 2 \qquad \ldots(4)$$

Solving the above equations:

$$-4\alpha_3 = -1$$
$$\alpha_3 = 0.25$$

And

$$2\alpha_2 + 0.25 = 1$$
$$2\alpha_2 = 0.75$$
$$\alpha_2 = \frac{3}{8}$$

Thus,

$$\alpha_1 = \alpha_2 + \alpha_3$$
$$\alpha_1 = \frac{5}{8}$$

Solving for w:

$$w = \left[\frac{5}{8} * 1 * (1,1)\right] + \left[\frac{3}{8} * -1 * (-1,1)\right] + \left[\frac{1}{4} * -1 * (0,1)\right]$$
$$w = \left[\frac{5}{8}, \frac{5}{8}\right] + \left[\frac{3}{8}, -\frac{3}{8}\right] + \left[0, -\frac{1}{4}\right]$$
$$\mathrm{w} = [1, 0.5]$$

Considering Case 1 as ‘w’ should be maximized. Therefore w = (2, 0)

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

For Point 1:

$$+1 = (1, 0.5)(1, 1) + b$$
$$b = -\tfrac{1}{2}$$

For Point 2:

$$-1 = (1, 0.5)(-1, 1) + b$$
$$b = -\frac{1}{2}$$

For Point 3:

$$-1 = (1, 0.5)(0, -1) + b$$
$$b = -\frac{1}{2}$$

Naman Jain Vimal Kumar
1001670153

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

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

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:

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

Thus,

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

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

$$\frac{dL}{d\zeta_i} = 0 + kC\left(\sum_i^n \zeta_i\right)^{k-1} - \sum_i^n \alpha_i - \sum_i^n \mu_i = 0$$
$$\sum_i^n \alpha_i + \sum_i^n \mu_i = kC\left(\sum_i^n \zeta_i\right)^{k-1} \quad \ldots(4)$$

Substituting (2) in (1):

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

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

Naman Jain Vimal Kumar
1001670153

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

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

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

From (4): Substituting $\sum_i^n \alpha_i \, \zeta_i + \sum_i^n \mu_i\zeta_i = kC(\sum_i^n \zeta_i)^{k-1}$

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

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

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

When k = 2:

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

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

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