Date ___/___/___
Page ______

11/5/2021 Homework: Convex Optimization and SVM

Step 1:

Formulating the model / optimization function for a soft margin SVM:

$$\cancel{L(\omega, b, \beta_J, \lambda_J, \varepsilon_J)} \arg\min \frac{1}{2}\|W\|_2^2 + C\sum_{i=1}^{n} \varepsilon_i$$

s.t.:

$$y_i(\omega^T x_i + b) \geq 1 - \varepsilon_i \quad \forall i = 1 \ldots n$$

$$\varepsilon_i \geq 0$$

To remove some of these constraints, we apply Lagrangian function

$$L(\omega, b, \beta_J, \lambda_J, \varepsilon_J) = \frac{1}{2}\|W\|_2^2 - \sum_{J=1}^{n} \beta_J [y_J(\omega^T x_J + b) - 1 + \varepsilon_j]$$

$$+ C\sum_{i=1}^{N} \varepsilon_i - \sum_{J=1}^{n} \lambda_J \varepsilon_J$$

$$\therefore L = \frac{1}{2}\|W\|_2^2 + C\sum_{i=1}^{N} \varepsilon_i - \omega^T \sum_{J=1}^{n} \beta_J y_J x_J - b\sum_{J=1}^{n} \beta_J y_J$$

$$+ \sum_{J=1}^{N} \beta_J - \sum_{J=1}^{n} \beta_J \varepsilon_J - \sum_{J=1}^{n} \lambda_J \varepsilon_J$$

Date ___/___/___
Page ___

Step 2: Solving for ω^*, b^* and ε^*_J,

$$\frac{\partial L}{\partial \omega} = \omega^* + 0 - \sum_{J=1}^{n} \beta_J y_J x_J - 0 + 0 - 0 - 0 = 0$$

$$\therefore \omega^* = \sum_{J=1}^{n} \beta_J y_J x_J \rightarrow (1)$$

$$\frac{\partial L}{\partial b} = 0 + 0 - 0 - \sum_{J=1}^{N} \beta_J y_J + 0 - 0 - 0 = 0$$

$$\therefore \sum_{J=1}^{N} \beta_J y_J = 0 \rightarrow (2)$$

$$\frac{\partial L}{\partial \varepsilon} = \cancel{0 + nc - 0 - 0 + 0 -}$$

$$= 0 + \sum_{J=1}^{n} c - 0 - 0 + 0 - \sum_{J=1}^{n} \beta_J - \sum_{J=1}^{n} \lambda_J = 0$$

$$\therefore c - \beta_J - \lambda_J = 0 \rightarrow (3)$$

Step 3: Replacing the findings in step 2 ~~to~~ in the original equation to get the dual function.

$$L^* = \frac{1}{2} \sum_{i=1}^{n} \sum_{J=1}^{n} \beta_i \beta_J y_i y_J (x_i^T x_J) + \cancel{c \sum_{J=1}^{n} \varepsilon_J}$$
$$- \sum_{i=1}^{n} \sum_{J=1}^{n} \beta_i \beta_J y_i y_J (x_i^T x_J) + \sum_{J=1}^{n} \beta_J$$
$$\cancel{- \sum_{J=1}^{n} \beta_J \varepsilon_J} - \cancel{\sum_{J=1}^{n} \lambda_J \varepsilon_J} \quad \text{(from (3))}$$

Date / /
Page

$$\therefore L^* = -\frac{1}{2} \sum_{J=1}^{n} \sum_{i=1}^{n} \beta_i \beta_J (x_i^T x_J) + \sum_{J=1}^{n} \beta_J$$

Step 4: Writing the dual function with it's associated constraints.

$$\text{s.t}: \quad \beta_J \geq 0, \rightarrow (4)$$

$$\sum \beta_J y_J \leq 0 \rightarrow (5)$$

$$C - \lambda_J - \beta_J = 0 \rightarrow (6)$$

Now, note that

$\beta_J \geq 0$ and $\lambda_J \geq 0$ ~~(Primal constraints)~~

Now, there can be two cases:

(a) $\lambda_J > 0$,

then $C - \beta_J > \lambda_J$

$\Rightarrow \beta_J < C$ (Based on (6))

(b) $\lambda_J = 0$,

then $C = \beta_J \Rightarrow \beta_J = C$

Date ___/___/___

Page ___

From this we can go to step 4.

Step 4 : Dual function with new constraint.

$$\max_{\beta_J} \sum \beta_J - \frac{1}{2} \sum_{i=1}^{n} \sum_{J=1}^{n} \beta_i \beta_J y_i y_J (x_i^T x_J)$$

s.t :

$$\sum \beta_J y_J \geq 0 ,$$

also called Box constraint. ← $0 \leq \beta_J \leq C$ [we just derived this in step 3]

As we can see, the dual function is free from w and b, we only need to solve for β_J now.

Step 5 : Applying SMO to solve for the dual variable β and λ.

~~Imagine we solved for w^* and b^* using SMO.~~

Imagine we solved for $\beta_1^*, \beta_2^*, \ldots \beta_n^*$ by using SMO.

~~Also we solved for $\lambda_1^* \ldots \lambda_J^*$ using SMO as well.~~

Date ___/___/______

Page ______

Step 6: Checking for KKT constraints

Complementary Slackness

$$\beta_J^*[y_J(w^{*T}x_J + b^*) - 1 + \varepsilon] = 0$$

$$\lambda_J \varepsilon_J = 0$$

There can be the following cases:

1) $\beta_J > 0$,

Then $y_J(w^{*T}x + b^*) - 1 + \varepsilon = 0$

$\therefore y_J(w^{*T}x + b^*) = 1 - \varepsilon$

This can further have two cases:

a) If $\lambda_J > 0$, then $\varepsilon_J = 0$

which means $y_J(w^{*T}x + b^*) = 1$

i.e, the points lie on the margin.

b) If $\varepsilon_J > 0$, $\Rightarrow \lambda_J = 0$

Then $C = \beta_J$.

Date ___/___/___
Page ___

$\beta^{*}_{j} = C$

=> These are the ~~misclassified~~ points which lie inside the margin and can be classified or misclassified depending on ξ.

=> These are the misclassified points.

2) $\beta_j = 0$,

then => $y_j(w^{*T}x_j + b) > 1 - \xi_j$

Now, $\lambda_j = C$ => $\xi_j = 0$

Scanned with CamScanner