

Homework 3: Support Vector Machines

Lecturer: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

 Instructions: For Formative assessment, submit the solutions to all of the parts of the Exercise

Exercise 1. Consider a training data set $\mathcal{S} = \{z_i = (x_i, y_i)\}_{i=1}^m$. Consider the Soft-SVM Algorithm that requires the solution of the following quadratic minimization problem (in a slightly modified but equivalent form to what we have discussed)

Primal problem:

$$(0.1) \quad (w^*, b^*, \xi^*) = \arg \min_{(w, b, \xi)} \left(\frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^m \xi_i \right)$$

$$(0.2) \quad \text{subject to: } y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i, \quad \forall i = 1, \dots, m$$

$$(0.3) \quad \xi_i \geq 0, \quad \forall i = 1, \dots, m$$

for some user-specified fixed parameter $C > 0$.

- (1) Specify the Lagrangian function L associated to the above primal quadratic minimization problem, where $\{\alpha_i\}$ are the Lagrange coefficients wrt (0.2), and $\{\beta_i\}$ are the Lagrange coefficients wrt (0.3). Write down any possible restrictions on the Lagrange coefficients.
- (2) Compute the dual Lagrangian function denoted as \tilde{L} as a function of the Lagrange coefficients and the data points \mathcal{S} .
- (3) Apply the Karush–Kuhn–Tucker (KKT) conditions to the above problem, and write them down.
- (4) Derive and write down the dual Lagrangian quadratic maximization problem, along with the inequality and equality constraints, where you seek to find $\{\alpha_i\}$.
- (5) Justify why the i -th point x_i lies on the margin boundary when $\alpha_i \in (0, C)$ (beware it is $\alpha_i \neq C$), and why the i -th point x_i can lie inside the margin when $\alpha_i = C$.
- (6) Given optimal values $\{\alpha_i^*\}$ for Lagrangian coefficients $\{\alpha_i\}$ as they are derived by solving the dual Lagrangian maximization problem in part 4, derive the optimal values w^* and b^* for the parameters w and b as function of the support vectors. Regarding parameter b it should be in the derived in the form

$$b^* = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \left(y_i - \sum_{j \in \mathcal{S}} \alpha_j^* y_j \langle x_j, x_i \rangle \right)$$

where you determine the sets \mathcal{M} and \mathcal{S} .

- (7) Report the halfspace predictive rule $h_{w,b}(x)$ of the above problem as a function of α^* and b^* .

Solution.

(1) It is

$$(0.4) \quad L(w, b, \xi, \alpha, \beta) = \frac{1}{2} \|w\|_2^2 + \sum_{i=1}^m C\xi_i + \sum_{i=1}^m \alpha_i (1 - y_i (\langle w, x_i \rangle + b) - \xi_i) - \sum_{i=1}^m \beta_i \xi_i$$

(2) Let α, β be fixed. We minimize (0.4) wrt w, b and we get

$$(0.5) \quad 0 = \frac{\partial L}{\partial w}(w, b, \xi, \alpha, \beta) \implies w = \sum_{i=1}^m \alpha_i y_i x_i$$

$$0 = \frac{\partial L}{\partial b}(w, b, \xi, \alpha, \beta) \implies 0 = \sum_{i=1}^m \alpha_i y_i$$

$$(0.6) \quad 0 = \frac{\partial L}{\partial \xi_i}(w, b, \xi, \alpha, \beta) \implies \alpha_i = C - \beta_i$$

and we substitute (0.5)-(0.6) in (0.4) and we get

$$\tilde{L}(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \langle x_j, x_i \rangle$$

(3) The Karush–Kuhn–Tucker (KKT) conditions applied to the above problem are

$$0 = \nabla \frac{1}{2} \|w\|_2^2 \nabla \sum_{i=1}^m C\xi_i + \nabla \sum_{i=1}^m \alpha_i (1 - y_i (\langle w, x_i \rangle + b) - \xi_i) - \nabla \sum_{i=1}^m \beta_i \xi_i \quad \text{Stationarity}$$

$$1 - y_i (\langle w, x_i \rangle + b) - \xi_i \leq 0, \quad \forall i = 1, \dots, m \quad \text{Primal feasibility}$$

$$\xi_i \geq 0$$

(0.7)

$$\alpha_i \geq 0 \quad \forall i = 1, \dots, m \quad \text{Dual feasibility}$$

(0.8)

$$\beta_i \geq 0 \quad \forall i = 1, \dots, m$$

(0.9)

$$\alpha_i (1 - y_i (\langle w, x_i \rangle + b) - \xi_i) = 0, \quad \forall i = 1, \dots, m \quad \text{Complementary slackness}$$

(0.10)

$$\beta_i \xi_i = 0, \quad \forall i = 1, \dots, m$$

(4) It is

$$(0.11) \quad \alpha^* = \arg \max_{\alpha \in \mathbb{R}^m: \alpha \geq 0} \left(\sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \langle x_j, x_i \rangle \right)$$

$$\text{subject to } 0 = \sum_{i=1}^m \alpha_i y_i$$

$$(0.12) \quad \alpha_i \in [0, C] \quad \forall i = 1, \dots, m$$

constrain (0.12) results from (0.6), (0.8), and (0.7).

(5)

- By (0.5), if $\alpha_i = 0$ then x_i does not contribute to the computation of the weights.
- By (0.5), if $\alpha_i \neq 0$ (aka $\alpha_i > 0$), then x_i is a support vector and contributes.
- If $\alpha_i \in (0, C)$ (where $\alpha_i \neq C$) then (0.6) implies that $\beta_i > 0$. By (0.10) if $\beta_i > 0$ then $\xi_i = 0$. Hence, given these, from (0.9), it is $1 = y_i (\langle w, x_i \rangle + b)$ i.e. x_i lies on the boundary.
- If $\alpha_i = C$, then x_i can lie inside the boundary. This is because, from (0.6) $\beta_i = 0$. Hence

- from (0.10) it may be $\xi_i = 0$ hence lie inside the margin, or
- it may be $\xi_i > 0$ and hence be correctly classified when $\xi_i \in (0, 1]$, or
- it may be $\xi_i > 0$ and hence be misclassified when $\xi_i \in (1, \infty)$.

- (6) From (0.9), it is either $\alpha_i = 0$ or $(1 - y_i (\langle w, x_i \rangle + b) - \xi_i) = 0$. Let $\mathcal{S} = \{i : y_i (\langle w, x_i \rangle + b) = 1 - \xi_i\}$. From (0.5), it is

$$(0.13) \quad w^* = \sum_{i \in \mathcal{S}} \alpha_i^* y_i x_i$$

Using (0.9) and summing up indexes in $\mathcal{M} = \{i : \alpha_i \in (0, C)\}$ for which $\xi_i = 0$ it is

$$b^* = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \left(y_i - \sum_{j \in \mathcal{S}} \alpha_j^* y_j \langle x_j, x_i \rangle \right)$$

(7) The formula is

$$(0.14) \quad \begin{aligned} h_{w,b}(x) &= \text{sign}(\langle w^*, x \rangle + b^*) \\ &= \text{sign} \left(\sum_{i=1}^m \alpha_i^* y_i \langle x_i, x \rangle + b^* \right) \end{aligned}$$