Unit 11 - Week 8 - Support Vector Machine (SVM) - II

Assignment 08

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2019-09-25, 23:59 IST.

<u>Instructions</u>:

1. **Attempt all questions.**
2. **Submission deadline**: 25th September 2019 23:59 IST
3. **Solutions to be posted**: 26th September 2019
4. Older browsers might show **unnecessary vertical bars** at the end of math equations.

1) Consider the problem of finding an optimal hyperplane for non-seperable patterns, we introduce a new set of variables, $\{\xi_i\}_{i=1}^{N}$ into the definition of the separating hyperplane as $d_i(w^T x_i + b) \geq 1 - \xi_i$. Choose the correct statements from the options given below. ***2 points***

- The slack variable ξ_i can take both positive and negative values.
- For $0 < \xi_i \leq 1$, the data point falls inside the region of separation, but on the correct side of the decision surface.
- For $\xi_i > 1$,the data point falls on the wrong side of the separating hyperplane.
- For support vectors ξ_i will be always zero.

No, the answer is incorrect.
Score: 0

Accepted Answers:
For $0 < \xi_i \leq 1$, the data point falls inside the region of separation, but on the correct side of the decision surface.
For $\xi_i > 1$,the data point falls on the wrong side of the separating hyperplane.

2) For the nonseparable case, we minimize the cost function defined as ***1 point***
$L = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$
(True/False) The optimal value of C is obtained by minimizing the cost function with respect to C.

- True
- False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False

3) In continuation with question 2, consider the following statements: ***2 points***
a) The parameter C can be chosen using cross validation approach.
b) When C is assigned a small value, the training samples are considered to be noisy, and less emphasis should therefore be placed on it.
c)The optimization problem for linearly separable patterns can be considered as a special case of optimization problem for nonseparable patterns, by setting $\xi_i = 0$ for all i.
d) When C is assigned a large value, the implication is that the designer of the SVM has high confidence in the quality of the training samples.
Which of the above statements are correct?

- Only a and c
- Only b and d
- Only a,b and c
- a, b, c and d

No, the answer is incorrect.
Score: 0

Accepted Answers:
a, b, c and d

4) (True/False) For solving the dual problem for nonseparable patterns, the constraints on the slack variables i.e., $\xi_i \geq 0$ and its corresponding Lagrange multiplier $\lambda_i \geq 0$ is essential. ***1 point***

- True
- False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False

5) Given the training sample where the data are $\{x_i, d_i\}_{i=1}^{N}$ statistically independent and identically distributed. The frequently used loss function for regression is the epsilon sensitive loss defined as ***2 points***
$L_\epsilon(d_i, y_i) = \begin{cases} |d_i - y_i| - \epsilon & |d_i - y_i| \geq \epsilon \\ 0 & \text{otherwise.} \end{cases}$
where d_i is the desired response and y_i is the corresponding estimator output where $y_i = \mathbf{w}^T \phi(\mathbf{x}_i)$. The optimization problem is defined as

$$\min \quad \frac{1}{N} \sum_{i=0}^{N-1} L_\epsilon(d_i, y_i)$$
$$\text{subject to} \quad \begin{cases} \|\mathbf{w}\|^2 & \leq c_0 \\ d_i - \mathbf{w}^T \phi(\mathbf{x}_i) & \leq \epsilon + \xi_i \\ \mathbf{w}^T \phi(\mathbf{x}_i) - d_i & \leq \epsilon + \xi_i' \\ \xi_i & \geq 0 \\ \xi_i' & \geq 0 \end{cases} \quad \forall i = 0, 1, 2, \ldots, N-1.$$

Find the Lagrangian for the problem.

- $$L = \frac{1}{N} \sum_{i=0}^{N-1} (\xi_i + \xi_i') - \alpha \left(\|\mathbf{w}\|^2 - c_0\right) - \sum_{i=0}^{N-1} \beta_i \left(\epsilon + \xi_i + d_i - \mathbf{w}^T \phi(\mathbf{x}_i)\right) - \sum_{i=0}^{N-1} \beta_i' \left(\epsilon + \xi_i' + \mathbf{w}^T \phi(\mathbf{x}_i) + d_i\right) + \sum_{i=0}^{N-1} \gamma_i \xi_i + \sum_{i=0}^{N-1} \gamma_i' \xi_i'$$
- $$L = \frac{1}{N} \sum_{i=0}^{N-1} (\xi_i + \xi_i') + \alpha \left(c_0 - \|\mathbf{w}\|^2\right) + \sum_{i=0}^{N-1} \beta_i \left(\epsilon + \xi_i - d_i + \mathbf{w}^T \phi(\mathbf{x}_i)\right) + \sum_{i=0}^{N-1} \beta_i' \left(\epsilon + \xi_i' - \mathbf{w}^T \phi(\mathbf{x}_i) + d_i\right) + \sum_{i=0}^{N-1} \gamma_i \xi_i + \sum_{i=0}^{N-1} \gamma_i' \xi_i'$$
- $$L = \frac{1}{N} \sum_{i=0}^{N-1} (\xi_i + \xi_i') - \alpha \left(c_0 - \|\mathbf{w}\|^2\right) - \sum_{i=0}^{N-1} \beta_i \left(\epsilon + \xi_i - d_i + \mathbf{w}^T \phi(\mathbf{x}_i)\right) - \sum_{i=0}^{N-1} \beta_i' \left(\epsilon + \xi_i' - \mathbf{w}^T \phi(\mathbf{x}_i) + d_i\right) - \sum_{i=0}^{N-1} \gamma_i \xi_i - \sum_{i=0}^{N-1} \gamma_i' \xi_i'$$
- $$L = \frac{1}{N} \sum_{i=0}^{N-1} (\xi_i + \xi_i') - \alpha \left(c_0 - \|\mathbf{w}\|^2\right) - \sum_{i=0}^{N-1} \beta_i \left(\epsilon + \xi_i - d_i + \mathbf{w}^T \phi(\mathbf{x}_i)\right) - \sum_{i=0}^{N-1} \beta_i' \left(\epsilon + \xi_i' - \mathbf{w}^T \phi(\mathbf{x}_i) + d_i\right)$$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$$L = \frac{1}{N} \sum_{i=0}^{N-1} (\xi_i + \xi_i') - \alpha \left(c_0 - \|\mathbf{w}\|^2\right) - \sum_{i=0}^{N-1} \beta_i \left(\epsilon + \xi_i - d_i + \mathbf{w}^T \phi(\mathbf{x}_i)\right) - \sum_{i=0}^{N-1} \beta_i' \left(\epsilon + \xi_i' - \mathbf{w}^T \phi(\mathbf{x}_i) + d_i\right) - \sum_{i=0}^{N-1} \gamma_i \xi_i - \sum_{i=0}^{N-1} \gamma_i' \xi_i'$$

6) In continuation with question 5, compute $\frac{\partial L}{\partial \mathbf{w}}$, $\frac{\partial L}{\partial \xi_i}$ and $\frac{\partial L}{\partial \xi_i'}$. ***2 points***

- $$\frac{\partial L}{\partial \mathbf{w}} = 2\alpha \mathbf{w} - \sum_{i=0}^{N-1} \beta_i \phi(\mathbf{x}_i) + \sum_{i=0}^{N-1} \beta_i' \phi(\mathbf{x}_i), \quad \frac{\partial L}{\partial \xi_i} = N - \beta_i - \gamma_i, \quad \frac{\partial L}{\partial \xi_i'} = N - \beta_i' - \gamma_i'$$
- $$\frac{\partial L}{\partial \mathbf{w}} = \alpha \mathbf{w} + \sum_{i=0}^{N-1} \beta_i \phi(\mathbf{x}_i) - \sum_{i=0}^{N-1} \beta_i' \phi(\mathbf{x}_i), \quad \frac{\partial L}{\partial \xi_i} = \frac{1}{N} + \beta_i + \gamma_i, \quad \frac{\partial L}{\partial \xi_i'} = \frac{1}{N} - \beta_i' - \gamma_i'$$
- $$\frac{\partial L}{\partial \mathbf{w}} = \alpha \mathbf{w} + \sum_{i=0}^{N-1} \beta_i \phi(\mathbf{x}_i) - \sum_{i=0}^{N-1} \beta_i' \phi(\mathbf{x}_i), \quad \frac{\partial L}{\partial \xi_i} = \frac{1}{N} - \beta_i - \gamma_i, \quad \frac{\partial L}{\partial \xi_i'} = \frac{1}{N} - \beta_i' - \gamma_i'$$
- $$\frac{\partial L}{\partial \mathbf{w}} = 2\alpha \mathbf{w} - \sum_{i=0}^{N-1} \beta_i \phi(\mathbf{x}_i) + \sum_{i=0}^{N-1} \beta_i' \phi(\mathbf{x}_i), \quad \frac{\partial L}{\partial \xi_i} = \frac{1}{N} - \beta_i - \gamma_i, \quad \frac{\partial L}{\partial \xi_i'} = \frac{1}{N} - \beta_i' - \gamma_i'$$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$$\frac{\partial L}{\partial \mathbf{w}} = 2\alpha \mathbf{w} - \sum_{i=0}^{N-1} \beta_i \phi(\mathbf{x}_i) + \sum_{i=0}^{N-1} \beta_i' \phi(\mathbf{x}_i), \quad \frac{\partial L}{\partial \xi_i} = \frac{1}{N} - \beta_i - \gamma_i, \quad \frac{\partial L}{\partial \xi_i'} = \frac{1}{N} - \beta_i' - \gamma_i'$$

7) In continuation with question 5, find the Lagrangian dual for the problem. ***3 points***

- $$\max \quad c_0 - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\beta_i - \beta_i')(\beta_j - \beta_j') \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \sum_{i=0}^{N-1} (\beta_i + \beta_i')\epsilon + \sum_{i=0}^{N-1} (\beta_i + \beta_i') d_i$$ $$\text{subject to} \begin{cases} \alpha \geq 0 \\ 0 \leq \beta_i \leq \frac{1}{N} \\ 0 \leq \beta_i' \leq \frac{1}{N} \end{cases} \forall i = 0, 1, 2, \ldots, N-1.$$
- $$\max \quad -\alpha c_0 - \frac{1}{4\alpha} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\beta_i - \beta_i')(\beta_j - \beta_j') \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \sum_{i=0}^{N-1} (\beta_i + \beta_i')\epsilon + \sum_{i=0}^{N-1} (\beta_i + \beta_i') d_i$$ $$\text{subject to} \begin{cases} \alpha \geq 0 \\ 0 \leq \beta_i \leq \frac{1}{N} \\ 0 \leq \beta_i' \leq \frac{1}{N} \end{cases} \forall i = 0, 1, 2, \ldots, N-1.$$
- $$\max \quad +\alpha c_0 + \frac{1}{4\alpha} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\beta_i - \beta_i')(\beta_j + \beta_j') \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) + \sum_{i=0}^{N-1} (\beta_i + \beta_i')\epsilon + \sum_{i=0}^{N-1} (\beta_i + \beta_i') d_i$$ $$\text{subject to} \begin{cases} \alpha \geq 0 \\ 0 \leq \beta_i \leq \frac{1}{N} \\ 0 \leq \beta_i' \leq \frac{1}{N} \end{cases} \forall i = 0, 1, 2, \ldots, N-1.$$
- $$\max \quad -\alpha c_0 - \frac{1}{4\alpha} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\beta_i - \beta_i')(\beta_j - \beta_j') \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) + \sum_{i=0}^{N-1} (\beta_i + \beta_i')\epsilon + \sum_{i=0}^{N-1} (\beta_i + \beta_i') d_i$$ $$\text{subject to} \begin{cases} \alpha \geq 0 \\ 0 \leq \beta_i \leq N \\ 0 \leq \beta_i' \leq N \end{cases} \forall i = 0, 1, 2, \ldots, N-1.$$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$$\max \quad -\alpha c_0 - \frac{1}{4\alpha} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\beta_i - \beta_i')(\beta_j - \beta_j') \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \sum_{i=0}^{N-1} (\beta_i + \beta_i')\epsilon + \sum_{i=0}^{N-1} (\beta_i + \beta_i') d_i$$
$$\text{subject to} \begin{cases} \alpha \geq 0 \\ 0 \leq \beta_i \leq \frac{1}{N} \\ 0 \leq \beta_i' \leq \frac{1}{N} \end{cases} \forall i = 0, 1, 2, \ldots, N-1.$$

8) In continuation with question 5, find the estimate y for an unseen sample x by using an inner product kernel defined as $k(x_i, x_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ ***1 point***

- $y = \alpha \sum_{i=0}^{N-1} (\beta_i - \beta_i') k(x_i, x)$
- $y = \frac{1}{2\alpha} \sum_{i=0}^{N-1} (\beta_i - \beta_i') k(x_i, x)$
- $y = \frac{1}{\alpha^2} \sum_{i=0}^{N-1} (\beta_i - \beta_i') k(x_i, x)$
- $y = 2\alpha \sum_{i=0}^{N-1} (\beta_i + \beta_i') k(x_i, x)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$y = \frac{1}{2\alpha} \sum_{i=0}^{N-1} (\beta_i - \beta_i') k(x_i, x)$

9) If we are using a kernel function k to evaluate the inner products in a feature space with feature map ϕ, the associated Gram matrix $\mathbf{G}$ has entries $\mathbf{G}_{ij} = k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. Then the kernel matrix $\mathbf{G}$ is ***1 point***

- Positive definite.
- Negative definite.
- Positive semi-definite.
- Negative semi-definite.

No, the answer is incorrect.
Score: 0

Accepted Answers:
Positive semi-definite.

10) Let k_1 and k_2 be kernels over $X \times X$, $X \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^+$, $f(.)$ a real valued function on X, $\phi : X \to \mathbb{R}^N$ with k_3 a kernel over $\mathbb{R}^N \times \mathbb{R}^N$. Then the following function are kernels ***3 points***

- k(x,z)=k_{3}(\phi(x),\phi(z))
- k(x,z)=f(x)f(z)
- k(x,z)=ak_{1}(x,z)
- k(x,z) = k_{1}(x,z)+k_{2}(x,z)

No, the answer is incorrect.
Score: 0

Accepted Answers:
k(x,z)=k_{3}(\phi(x),\phi(z))
k(x,z)=f(x)f(z)
k(x,z)=ak_{1}(x,z)
k(x,z) = k_{1}(x,z)+k_{2}(x,z)

11) Following are the examples for Hilbert spaces ***2 points***

- Any finite dimensional inner product space.
- $S = \{(x_1, x_2, x_3, \ldots) : x_k \in \mathbb{C}, \sum_{k=1}^{\infty} |x_k|^2 < \infty\}$ with $\langle x, y \rangle = \sum_{k=1}^{\infty} x_k \bar{y}_k$ where $\mathbb{C}$ is a complex field, $\bar{y}$ is the complex conjugate of y
- The space of measurable functions on $[a, b]$ with inner product defined as $\langle f, g \rangle = \int_a^b w(t) f(x) \overline{g(x)}$ where $w(t) > 0 \quad \forall t$ is some weighting function.
- $C([a, b])$ is the set of complex valued functions defined on the interval $[a, b]$ such that $\langle f, g \rangle = \int_a^b f(x) \overline{g(x)}$

No, the answer is incorrect.
Score: 0

Accepted Answers:
Any finite dimensional inner product space.
$S = \{(x_1, x_2, x_3, \ldots) : x_k \in \mathbb{C}, \sum_{k=1}^{\infty} |x_k|^2 < \infty\}$ with $\langle x, y \rangle = \sum_{k=1}^{\infty} x_k \bar{y}_k$ where $\mathbb{C}$ is a complex field, $\bar{y}$ is the complex conjugate of y
The space of measurable functions on $[a, b]$ with inner product defined as $\langle f, g \rangle = \int_a^b w(t) f(x) \overline{g(x)}$ where $w(t) > 0 \quad \forall t$ is some weighting function.