QUIZ1

1. Recall that N is the size of the data set and d is the dimensionality of the input space. The primal formulation of the linear soft-margin support vector machine problem, without going through the Lagrangian dual problem, is

    A. a quadratic programming problem with N variables

    B. a quadratic programming problem with $d+1$ variables

    C. none of the other choices

    D. a quadratic programming problem with $2N$ variables

    E. a quadratic programming problem with $N+d+1$ variables

2. Consider the following training data set:

$$\mathbf{x}_1 = (1,0), y_1 = -1 \qquad \mathbf{x}_2 = (0,1), y_2 = -1 \qquad \mathbf{x}_3 = (0,-1), y_3 = -1$$

$$\mathbf{x}_4 = (-1,0), y_4 = +1 \qquad \mathbf{x}_5 = (0,2), y_5 = +1 \qquad \mathbf{x}_6 = (0,-2), y_6 = +1$$

$$\mathbf{x}_7 = (-2,0), y_7 = +1$$

Use following nonlinear transformation of the input vector $\mathbf{x} = (x_1, x_2)$ to the transformed vector $\mathbf{z} = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}))$:

$$\phi_1(\mathbf{x}) = x_2^2 - 2x_1 + 3 \qquad \phi_2(\mathbf{x}) = x_1^2 - 2x_2 - 3$$

What is the equation of the optimal separating "hyperplane" in the $\mathcal{Z}$ space?

    A. $z_1 + z_2 = 4.5$

    B. $z_1 - z_2 = 4.5$

    C. $z_1 = 4.5$

    D. $z_2 = 4.5$

    E. none of the other choices

3. Consider the same training data set as Question 2, but instead of explicitly transforming the input space $\mathcal{X}$ to $\mathcal{Z}$, apply the hard-margin support vector machine algorithm with the kernel function

$$K(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^T \mathbf{x}')^2,$$

which corresponds to a second-order polynomial transformation. Set up the optimization problem using $(\alpha_1, \cdots, \alpha_7)$ and numerically solve for them (you can use any package you want). Which of the followings are true about the optimal $\boldsymbol{\alpha}$?

    A. $\sum_{n=1}^{7} \alpha_n \approx 2.8148$

    B. $\min_{1 \le n \le 7} \alpha_n = \alpha_7$

    C. there are 6 nonzero α_n

    D. none of the other choices

    E. $\max_{1 \le n \le 7} \alpha_n = \alpha_7$

4. Following Question 3, what is the corresponding nonlinear curve in the $\mathcal{X}$ space?

    A. $\frac{1}{9}(8x_1^2 - 16x_1 + 6x_2^2 - 15) = 0$

1

B. none of the other choices

C. $\frac{1}{9}(8x_2^2 - 16x_2 + 6x_1^2 + 15) = 0$

D. $\frac{1}{9}(8x_2^2 - 16x_2 + 6x_1^2 - 15) = 0$

E. $\frac{1}{9}(8x_1^2 - 16x_1 + 6x_2^2 + 15) = 0$

5. Compare the two nonlinear curves found in Questions 2 and 4, which of the following is true?

    A. none of the other choices

    B. The curves should be the same in the $\mathcal{X}$ space, because they are learned with respect to the same $\mathcal{Z}$ space

    C. The curves should be different in the $\mathcal{X}$ space, because they are learned with respect to different $\mathcal{Z}$ spaces

    D. The curves should be different in the $\mathcal{X}$ space, because they are learned from different raw data $\{(\mathbf{x}_n, y_n)\}$

    E. The curves should be the same in the $\mathcal{X}$ space, because they are learned from the same raw data $\{(\mathbf{x}_n, y_n)\}$

6. Recall that for support vector machines, d_{vc} is upper bounded by $\frac{R^2}{\rho^2}$, where ρ is the margin and R is the radius of the minimum hypersphere that $\mathcal{X}$ resides in. In general, R should come from our knowledge on the learning problem, but we can estimate it by looking at the minimum hypersphere that the training examples resides in. In particular, we want to seek for the optimal R that solves

$$(P) \quad \min_{R \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^d} \quad R^2 \quad \text{subject to } \|\mathbf{x}_n - \mathbf{c}\|^2 \leq R^2 \text{ for } n = 1, 2, \cdots, N.$$

Let λ_n be the Lagrange multipliers for the n-th constraint above. Following the derivation of the dual support vector machine in class, write down (P) as an equivalent optimization problem

$$\min_{R \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^d} \quad \max_{\lambda_n \geq 0} \quad L(R, \mathbf{c}, \boldsymbol{\lambda}).$$

What is $L(R, \mathbf{c}, \boldsymbol{\lambda})$?

    A. $R^2 + \sum_{n=1}^{N} \lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2)$

    B. $R^2 - \sum_{n=1}^{N} \lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2)$

    C. $R^2 + \sum_{n=1}^{N} \lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 + R^2)$

    D. $R^2 - \sum_{n=1}^{N} \lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 + R^2)$

    E. none of the other choices

7. Using (assuming) strong duality, the solution to (P) in Question 6 would be the same as the Lagrange dual problem

$$(D) \quad \max_{\lambda_n \geq 0} \quad \min_{R \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^d} \quad L(R, \mathbf{c}, \boldsymbol{\lambda}).$$

Which of the following can be derived from the KKT conditions of (P) and (D) at the optimal $(R, \mathbf{c}, \boldsymbol{\lambda})$?

    A. if $\sum_{n=1}^{N} \lambda_n \neq 0$, then $\mathbf{c} = \left(\sum_{n=1}^{N} \lambda_n \mathbf{x}_n\right) \Big/ \left(\sum_{n=1}^{N} \lambda_n\right)$

    B. if $\lambda_n = 0$, then $\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2 = 0$

    C. if $R \neq 0$, then $R \neq 0$

    D. none of the other choices

    E. if $\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2 < 0$, then $\lambda_n = 0$