Homework 02: The Bias-variance Tradeoff, SVMs and Kernel Methods

CS 6316 Machine Learning

Due on Oct. 27, 2020 11:59 PM

1. **Bias-Variance Tradeoff** (12 points)

   Please refer to the attached iPython notebook file.

2. **KKT Conditions** (4 points) The Lagrangian form of SVMs with slack variable $\boldsymbol{\xi}$ is formulated as

$$\begin{aligned} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) =& \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{m}\xi_i \\ & - \sum_{i=1}^{m}\alpha_i(y_i(\boldsymbol{w}^\mathsf{T}\boldsymbol{x}_i + b) - 1 + \xi_i) \\ & - \sum_{i=1}^{m}\beta_i\xi_i \end{aligned} \tag{1}$$

   Similar to SVMs in separable cases (page 19), to find the KKT conditions (as in page 29, we need to compute the derivative with respect to all parameters $\{\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}\}$. Overall, please show that the KKT conditions can be represented to the following equations

$$\boldsymbol{w} = \sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i \tag{2}$$

$$\sum_{i=1}^{m}\alpha_i y_i = 0 \tag{3}$$

$$\alpha_i + \beta_i = C \tag{4}$$

$$\alpha_i = 0 \quad \text{or} \quad y_i(\boldsymbol{w}^\mathsf{T}\boldsymbol{x}_i + b) = 1 - \xi_i \tag{5}$$

$$\beta_i = 0 \quad \text{or} \quad \xi_i = 0 \tag{6}$$

3. **Kernel Methods** (4 points) In our lecture, we show that a special case of the polynomial kernels

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + c)^d \tag{7}$$

   with $d = 2$ and $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^2$. On page 53 of the slides, we should how this special case can be decomposed as a dot product with a nonlinear mapping $\Phi(\cdot)$

$$K(\boldsymbol{x}, \boldsymbol{x}') = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{x}') \rangle. \tag{8}$$

   In this problem, consider $d = 3$ with $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^2$ and show how the $\Phi(\boldsymbol{x})$ is defined in this case.