

CS6316 Homework 02: The Bias-variance Tradeoff, SVMs and Kernel Methods

Gabriel Hanson

Due on Oct. 27, 2022 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 ξ is formulated as

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

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 $\{\mathbf{w}, b, \xi, \alpha, \beta\}$. Overall, please show that the KKT conditions can be represented to the following equations

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

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

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

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

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

Solution:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i = 0 \quad \longrightarrow \quad \mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0 \quad \longrightarrow \quad \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi} = C - \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \beta_i = 0 \quad \longrightarrow \quad C = \alpha_i + \beta_i$$

and

$$\forall i \quad \alpha_i (y_i (\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i) = 0 \quad \longrightarrow \quad \alpha_i = 0 \quad \text{or} \quad y_i (\mathbf{w}^\top \mathbf{x}_i + b) = 1 - \xi_i$$

$$\forall i \quad \beta_i \xi_i = 0 \quad \longrightarrow \quad \beta_i = 0 \quad \text{or} \quad \xi_i = 0$$

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

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

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

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

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

Solution:

$$\begin{aligned} K(\mathbf{x}, \mathbf{x}') &= (\langle \mathbf{x}, \mathbf{x}' \rangle + c)^3 \quad \forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^2 \\ K(\mathbf{x}, \mathbf{x}') &= (\langle \mathbf{x}, \mathbf{x}' \rangle + c)^2 * (\langle \mathbf{x}, \mathbf{x}' \rangle + c) \\ &= (x_1 x'_1 + x_2 x'_2 + c)^2 * (x_1 x'_1 + x_2 x'_2 + c) \end{aligned}$$

This first term was solved for in slide 56 of the SVM lecture, which I will substitute in here

$$\begin{aligned} &= (x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x'_1 x_2 x'_2 + 2cx_1 x'_1 + 2cx_2 x'_2 + c^2) * (x_1 x'_1 + x_2 x'_2 + c) \\ &= (x_1^3 x_1'^3 + x_1 x'_1 x_2^2 x_2'^2 + 2x_1^2 x_1'^2 x_2 x'_2 + 2cx_1^2 x_1'^2 + 2cx_1 x'_1 x_2 x'_2 + c^2 x_1 x'_1 \\ &\quad + x_1^2 x_1'^2 x_2 x'_2 + x_2^3 x_2'^3 + 2x_1 x'_1 x_2^2 x_2'^2 + 2cx_1 x'_1 x_2 x'_2 + 2cx_2^2 x_2'^2 + c^2 x_2 x'_2 \\ &\quad + cx_1^2 x_1'^2 + cx_2^2 x_2'^2 + 2cx_1 x'_1 x_2 x'_2 + 2c^2 x_1 x'_1 + 2c^2 x_2 x'_2 + c^3) \end{aligned}$$

Combining terms here leaves us with

$$= (x_1^3 x_1'^3 + x_2^3 x_2'^3 + 3x_1^2 x_1'^2 x_2 x'_2 + 3x_1 x'_1 x_2^2 x_2'^2 + 3cx_1^2 x_1'^2 + 3cx_2^2 x_2'^2 + 6cx_1 x'_1 x_2 x'_2 + 3c^2 x_1 x'_1 + 3c^2 x_2 x'_2 + c^3)$$

which can be rewritten as

$$\begin{aligned} &[x_1^3, x_2^3, \sqrt{3}x_1^2 x_2, \sqrt{3}x_1 x_2^2, \sqrt{3}cx_1^2, \sqrt{3}cx_2^2, \sqrt{6}cx_1 x_2, \sqrt{3}cx_1, \sqrt{3}cx_2, c^{3/2}] \\ &\quad \begin{bmatrix} x_1'^3 \\ x_2'^3 \\ \sqrt{3}x_1'^2 x_2' \\ \sqrt{3}x_1' x_2'^2 \\ \sqrt{3}cx_1'^2 \\ \sqrt{3}cx_2'^2 \\ \sqrt{6}cx_1' x_2' \\ \sqrt{3}cx_1' \\ \sqrt{3}cx_2' \\ c^{3/2} \end{bmatrix} \end{aligned}$$

Let $K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$, then

$$\phi(\mathbf{x}) = [x_1^3, x_2^3, \sqrt{3}x_1^2 x_2, \sqrt{3}x_1 x_2^2, \sqrt{3}cx_1^2, \sqrt{3}cx_2^2, \sqrt{6}cx_1 x_2, \sqrt{3}cx_1, \sqrt{3}cx_2, c^{3/2}]^\top$$