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 $\boldsymbol{\xi}$ 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

$$
\begin{aligned}
\boldsymbol{w} &= \sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i && (2)\\
\sum_{i=1}^{m}\alpha_i y_i &= 0 && (3)\\
\alpha_i + \beta_i &= C && (4)\\
\alpha_i = 0 \quad &\text{or} \quad y_i(\boldsymbol{w}^\mathsf{T}\boldsymbol{x}_i + b) = 1 - \xi_i && (5)\\
\beta_i = 0 \quad &\text{or} \quad \xi_i = 0 && (6)
\end{aligned}
$$

Solution:

$$
\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{m}\alpha_i y_i x_i = 0 \qquad \longrightarrow \qquad \boldsymbol{w} = \sum_{i=1}^{m}\alpha_i y_i x_i
$$

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

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

and

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

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

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

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

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.

Solution:

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + c)^3 \ \forall \ x, x' \in \mathbb{R}^2$$
$$K(\boldsymbol{x}, \boldsymbol{x}') = (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + c)^2 * (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + c)$$
$$= (x_1x_1' + x_2x_2' + c)^2 * (x_1x_1' + x_2x_2' + c)$$

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

$$= (x_1^2x_1'^2 + x_2^2x_2'^2 + 2x_1x_1'x_2x_2' + 2cx_1x_1' + 2cx_2x_2' + c^2) * (x_1x_1' + x_2x_2' + c)$$
$$\begin{aligned} = (&x_1^3x_1'^3 + x_1x_1'x_2^2x_2'^2 + 2x_1^2x_1'^2x_2x_2' + 2cx_1^2x_1'^2 + 2cx_1x_1'x_2x_2' + c^2x_1x_1' \\ &+ x_1^2x_1'^2x_2x_2' + x_2^3x_2'^3 + 2x_1x_1'x_2^2x_2'^2 + 2cx_1x_1'x_2x_2' + 2cx_2^2x_2'^2 + c^2x_2x_2' \\ &+ cx_1^2x_1'^2 + cx_2^2x_2'^2 + 2cx_1x_1'x_2x_2' + 2c^2x_1x_1' + 2c^2x_2x_2' + c^3) \end{aligned}$$

Combining terms here leaves us with

$$= (x_1^3x_1'^3 + x_2^3x_2'^3 + 3x_1^2x_1'^2x_2x_2' + 3x_1x_1'x_2^2x_2'^2 + 3cx_1^2x_1'^2 + 3cx_2^2x_2'^2 + 6cx_1x_1'x_2x_2' + 3c^2x_1x_1' + 3c^2x_2x_2' + c^3)$$

which can be rewritten as

$$[x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, \sqrt{3c}x_1^2, \sqrt{3c}x_2^2, \sqrt{6c}x_1x_2, \sqrt{3}cx_1, \sqrt{3}cx_2, c^{3/2}] \begin{bmatrix} x_1'^3 \\ x_2'^3 \\ \sqrt{3}x_1'^2x_2' \\ \sqrt{3}x_1'x_2'^2 \\ \sqrt{3c}x_1'^2 \\ \sqrt{3c}x_2'^2 \\ \sqrt{6c}x_1'x_2' \\ \sqrt{3}cx_1' \\ \sqrt{3}cx_2' \\ c^{3/2} \end{bmatrix}$$

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

$$\phi(\boldsymbol{x}) = [x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, \sqrt{3c}x_1^2, \sqrt{3c}x_2^2, \sqrt{6c}x_1x_2, \sqrt{3}cx_1, \sqrt{3}cx_2, c^{3/2}]^\intercal$$