Homework 03: The Bias-variance Tradeoff, SVMs and Kernel Methods

CS 4774 Machine Learning

Due on April 11, 11:59 PM

Submission Instructions

- This assignment consists two written questions (5 points, in total) and one programming question (9 points, in total).
- Please submit the written part in a pdf file with name *[ComputingID]-hw03.pdf* and the programming part in a iPython notebook file with name *[ComputingID]-hw03.ipynb*
- For the written part, if you prefer to use handwriting (and then scan it to pdf), make sure it is clear enough for grading.

Questions

Please finish the programming part first.

1. **RBF Kernel** (2 points) As discussed in class, the RBF kernel is defined as

$$K(\boldsymbol{x}, \boldsymbol{x}') = e^{-\gamma \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2} \tag{1}$$

   Hopefully, from the programming part, you have already gotten a sense about how the hyper-parameter γ impact the model performance. Based on your observation and Equation 1, please give an intuitive explanation about how γ could impact model complexity. Your answer should cover

   - Whether higher or lower values leads to more flexible models, and
   - Why?

2. **Polynomial Kernels** (3 points) In our lecture on kernel methods, we show that a special case of the polynomial kernels

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + c)^d \tag{2}$$

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

$$K(\boldsymbol{x}, \boldsymbol{x}') = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{x}') \rangle. \tag{3}$$

   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. Note that, before splitting the kernel function to be a dot product of two high-dimensional vectors, make sure merge the same items as much as you can, as we demonstrated in class.