CMU 10-715: Homework 3

Kernel methods

DUE: Oct. 3, 2020, 11:59 PM.

Instructions:

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books, again after you have thought about the problems on your own. Please don't search for answers on the web, previous years' homeworks, etc. (please ask the TAs if you are not sure if you can use a particular reference). There are two requirements: first, cite your collaborators fully and completely (e.g., "Alice explained to me what is asked in Question 4.3"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.

- **Submitting your work:** Assignments should be submitted as PDFs using Gradescope unless explicitly stated otherwise. Each derivation/proof should be completed on a separate page. Submissions can be handwritten, but should be labeled and clearly legible. Else, submission can be written in LaTeX.

- **Late days:** For each homework you get three late days to be used only when anything urgent comes up. No points will be deducted for using these late days. We will consider an honor system where we will rely on you to use the late days appropriately.

1 Validity of a kernel [30]

Prove that K_1 an K_2 are valid kernel functions.

(a) (15 points) $K_1(\boldsymbol{x_1}, \boldsymbol{x_2}) = exp(\frac{-||\boldsymbol{x_1}-\boldsymbol{x_2}||^2}{\sigma^2})$, where σ is a constant, $K_1 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

(b) (15 points) $K_2(x_1, x_2) = \sigma^2 exp(cos(\frac{2\pi(x_1-x_2)}{p}))$, where σ, p are constants, $K_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

You can assume the following statements hold without proving them:

1. Kernels are closed under summation, multiplication, combination of polynomials with non-negative coefficients, and exponentiation.
2. $K(\boldsymbol{x_1}, \boldsymbol{x_2}) = \boldsymbol{x_1}^T \boldsymbol{x_2}$ is a valid kernel function, where $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.
3. $K(\boldsymbol{x_1}, \boldsymbol{x_2}) = f(\boldsymbol{x_1}) f(\boldsymbol{x_2})$ is a valid kernel function, where $f : \mathbb{R}^d \to \mathbb{R}$, $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

2 Kernelized soft SVM [40]

Given a training set $S = \{(\boldsymbol{x_1}, y_1), \cdots, (\boldsymbol{x_n}, y_n)\}$ where $\boldsymbol{x_i} \in \mathbb{R}^d, y_i \in \{-1, 1\}$. Recall that the soft SVM problem can be formulated as follows:

$$\underset{\mathbf{w}}{\text{minimize}} \;\; \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{i=1}^{n} \max\big(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i)\big) \tag{1}$$

(a) (10 points) Given a kernel function $K(\boldsymbol{x_1}, \boldsymbol{x_2}) = \langle\psi(\boldsymbol{x_1}), \psi(\boldsymbol{x_2})\rangle$. Rewrite equation (1) to solve for the kernelized soft SVM problem.

(b) (30 points) Write down the stochastic gradient descent algorithm (with mini-batch) for solving the problem in (a), assuming the batch size is b. Use the kernel trick instead of directly applying soft SVM to the transformed data, $\psi(\boldsymbol{x_i})$. The algorithm should include the initialization step, the update step, and the output. Only the pseudo code of the algorithm is needed. No coding is required.

3 Kernels and linear separability [30]

(a) (15 points) Show an example of a separator $g : \mathbb{R} \to \{-1, 1\}$ such that it is not perfectly separable even with infinite data using an identity kernel and a linear separator, but is separable if using either K_1 or K_2 from Q1 and a linear separator.

(b) (15 points) Show an example of a separator $g : \mathbb{R} \to \{-1, 1\}$ such that it is not perfectly separable even with infinite data using K_1 and a linear separator, but is separable if using K_2 and a linear separator.

Note: You can either write down the separator algebraically or show a hand-drawn plot of the data representing the separator.