STAT 542: Homework 3

Due: Mar. 8 midnight on Canvas

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

[3pts] Consider a regression problem with data $(x_i, y_i)_{i=1}^n$, where each x_i is one-dimensional. Consider a feature map $\phi(x) = (1, \frac{x}{2}, \frac{x^2}{4}, \dots, \frac{x^{p-1}}{2^{p-1}})$.

- Find an explicit expression for the kernel function $k(x, x') = \langle \phi(x), \phi(x') \rangle$, where $\langle , \rangle$ is the inner product in $\mathbb{R}^p$.
- Suppose that we want to estimate the regression function by solving

$$\hat{\theta} := \arg\min_{\theta \in \mathbb{R}^p} \left\{ \sum_{i=1}^n | \langle \phi(x_i), \theta \rangle - y_i |^2 + \lambda \|\theta\|_2^2 \right\} \tag{1}$$

  where $\lambda > 0$, and setting $\hat{f}(x) = \left\langle \phi(x), \hat{\theta} \right\rangle$. Find an expression of $\hat{f}$ using only $K = [k(x_i, x_j)]_{1 \le i,j \le n}$ (and not using ϕ directly). Note that the resulting algorithm is more computationally efficient than directly solving (1) when n is much smaller than p.
- If $\lambda \le 0$, is the expression of $\hat{f}$ you found in the previous question still equivalent to (1)?

Problem 2.

[2pts] Consider 3-NN in the setting of P6 in `s8_knn.pdf`. Show that as $n \to \infty$, the asymptotic error of 3-NN is upper bounded by

$$P_{3nn} \le p(1-p)(4p - 4p^2 + 1) \tag{2}$$

where p is the Bayes probability of error. Show that the bound can be weakened to $P_{3nn} \le 1.4p$, and compare it with the case of 1-NN.

Hint: `Notes_knn.pdf` *contains the main ingredients of the analysis.*

Problem 3.

[2pts] Kernel functions can be defined over objects as diverse as graphs, sets, strings, and text documents. Consider, for instance, a fixed set and define a nonvectorial space consisting of all possible subsets of this set. If A_1 and A_2 are two such subsets then one simple choice of kernel would be

$$k(A_1, A_2) = 2^{|A_1 \cap A_2|} \tag{3}$$

where $A_1 \cap A_2$ denotes the intersection of sets A_1 and A_2, and $|A|$ denotes the number of elements in A.

- Show that this is a valid kernel function, by constructing a feature map ϕ such that $k(A_1, A_2) = \langle \phi(A_1), \phi(A_2) \rangle$.
- If $k(A_1, A_2) = 4^{|A_1 \cap A_2|}$ instead, construct a ϕ such that $k(A_1, A_2) = \langle \phi(A_1), \phi(A_2) \rangle$.
- [1 additional bonus point] Construction for $k(A_1, A_2) = a^{|A_1 \cap A_2|}$, where $a > 1$ is arbitrary?

Hint: Part 1) is taken from Ex 6.12 in Bishop's book: "https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf". The constructions are not unique, but a convenient method for part 2), 3) is to use part 1) and then use the construction in P21 `s7_svm.pdf` *for a polynomial transform of a kernel.*