STAT 542: Homework 3

Due: Mar. 10 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.

[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

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

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

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

Problem 2.

[2pts] An important contribution to arise from the kernel viewpoint has been the extension to inputs that are symbolic, rather than simply vectors of real numbers. 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

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

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 $a > 0$ 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.*

Problem 3.

[3pts] Solve Ex. 9.5 (a)-(c) in ESL about "Degrees of freedom of a tree". `https://hastie.su.domains/Papers/ESLII.pdf`

Hint: Regression tree may be solved by `tree()` *in R; see e.g. p353 in* ***`https://hastie.su.domains/ISLR2/ISLRv2_website.pdf`*** *. To get an intuition for Dof, first consider the case where the m partition regions are fixed, so that $\hat{Y}$ is simply a projection of Y on an m-dimensional space. In a regression tree since the regions are partitioned adaptively, this intuition does not give the exact Dof.*