ECEN 649: Patten Recognition

Homework 1

Assign date: Wednesday September 4, 2019
Due date: Tuesday September 17, 2019

Problem 1 (Naive Bayes, 10 Points): Say you have 1000 fruits which could be either "Banana", "Orange", or "Other". For each fruit you look at the following three features: "Long", "Sweet", and "Yellow", and the value for each feature can be either "Yes" or "No". Let us aggregate the training data to form a counts table as follows:

	Long=Yes	Long=No	Sweet=Yes	Sweet=No	Yellow=Yes	Yellow=No
Banana	400	100	350	150	450	50
Orange	0	300	150	150	300	0
Other	100	100	150	50	50	150

a) (5 Points) Use the above counts table to estimate the probability of each class and the conditional probability of each feature value given each class. While the probability of each class can be reasonably estimated using relative frequency, such an approach may lead to a lot of zero-estimates when it comes to the conditional probability of each feature value given each class. Note that a zero-estimate for one feature can wipe out entirely the information from the other features, and this could be very problematic for naive Bayes. Therefore, the conditional probability of each feature value given each class is usually estimated with *Laplace's correction*. You can learn about Laplace's correction from the Wiki page `https://en.wikipedia.org/wiki/Additive_smoothing`.

b) (5 Points) Use the estimates from Part a) and naive Bayes to answer the following question: If a fruit is "long=yes", "sweet=yes", and "yellow=yes", what fruit is it?

Problem 2 (VC-dimension of intervals, 10 Points) Let the domain set $\mathcal{X} = \mathbb{R}$ and consider the class of *interval indicators* that predict 1 over an open interval and 0 outside of it. What is the VC-dimension of this class? Carefully justify your answer.

Problem 3 (VC-dimension of parity functions, 10 Points) Let $\mathcal{X}$ be the Boolean hypercube $\{0,1\}^d$. For a set $\mathcal{I} \subseteq [d]$, we define a *parity function* $h_{\mathcal{I}}$ as follows:

$$h_{\mathcal{I}}(\mathbf{x}) = \left(\sum_{i=1}^{d} x_i\right) \mod 2$$

where $\mathbf{x} = (x_1, \ldots, x_d) \in \{0,1\}^d$, i.e., $h_{\mathcal{I}}$ computes the parity of bits in $\mathcal{I}$. What is the VC-dimension of the class of all such parity functions? Carefully justify your answer.

Problem 4 (Efficient implementation of ERM for decision stumps, 10 Points) Read Chapter 10.1.1 of the textbook on efficient implementation of ERM for decision stumps. Implement the described algorithm using a programming language of your choice. To verify the correctness of your implementation, the grader will send you a dataset and ask you to return an ERM hypothesis.