Assignment 2

Due Thursday October 16 at the start of class.

(Question 1) [9 points] Let \((x_1, x_2)\) be a point in \(\mathbb{R}^2\). The label of \(x\) is defined to be the sign of the product \(x_1 \cdot x_2\). Given enough labeled training examples, can the perceptron algorithm learn this concept? Explain why this is impossible or possible.

(Question 2) [21 points] For each statement below, state clearly “True” if it is mostly true, or “False” if it is mostly false. Then write two or three sentences explaining why or how the statement is true or false. The maximum score for each answer is three points.

1. \(P(X = a | Y = b \text{ and } X = c) \in \{0, 1\}\) for all values of \(a, b,\) and \(c\).
2. If \(A\) and \(B\) are binary events then \(P(A \text{ or } B) = P(A) + P(B) - 2P(A \text{ and } B)\).
3. Bayes’ rule can be derived from the product rule \(P(A \text{ and } B) = P(A)P(B|A)\).

The next statements all refer to the following scenario. Given a training set \(D\), let \(n\) be its number of examples and let \(d\) be its dimensionality. \(D\) and \(D'\) are two very large datasets of the same size but different shapes: \(nd = n'd'\) and \(n < n'\) and \(d > d'\). In other words, \(D\) has high dimensionality and few examples; \(D'\) has low dimensionality and many examples.

4. Training a perceptron classifier, for one epoch, will be faster for \(D\) than for \(D'\).
5. A naive Bayes classifier will use more memory for \(D\) than for \(D'\).
6. Running a nearest-neighbor algorithm will be faster for \(D\) than for \(D'\).
7. A linear separator is more likely to exist for \(D\) than for \(D'\).

(Question 3) [20 points] [Adapted from an exam question written by Andrew Moore of CMU.] Consider two binary events named \(S\) for “person smokes” and \(C\) for “person gets cancer.” Half of all people smoke and half of all smokers get cancer: \(Pr(S) = 0.5\) and \(Pr(C|S) = 0.5\). Only a fifth of non-smokers get cancer: \(Pr(C|\text{not } S) = 0.2\).

(a) [5 points] Compute the chance of being a non-smoker non-cancer victim, \(Pr((\text{not } S) \text{ and } (\text{not } C))\).

A lawyer for a tobacco company agrees with the facts stated above. However, he claims that a certain fraction of people \(z = Pr(G)\) have an evil gene. The gene causes a certain fraction of people \(x = Pr(S|G)\) to smoke, while those who do not have the gene never smoke: \(Pr(S|\text{not } G) = 0\). And the gene causes a certain fraction of people \(y = Pr(C|G)\) to get cancer, but those without the gene never get cancer: \(Pr(C|\text{not } G) = 0\).

(b) [10 points] Find numerical values for \(x, y,\) and \(z\) that make the lawyer’s argument work. Explain your reasoning.

(c) [5 points] Is the answer to part (b) unique, or not? Explain your reasoning.