CS 189/289A Introduction to Machine Learning
Fall 2024 Jennifer Listgarten, Saeed Saremi

DIS9

1 Concerns about Randomness

One may be concerned that the randomness introduced in random forests may cause trouble. For example, some features or sample points may never be considered at all. In this problem we will be exploring this phenomenon.

(a) Consider n training points in a feature space of d dimensions. Consider building a random forest with T binary trees, each having exactly h internal nodes. Let m be the number of features randomly selected (from among d input features) at each tree node. For this setting, compute the probability that a certain feature (say, the first feature) is never considered for splitting in any tree node in the forest.

(b) Now let us investigate the possibility that some sample point might never be selected. Suppose each tree employs $n' = n$ bootstrapped (sampled with replacement) training sample points. Compute the probability that a particular sample point (say, the first sample point) is never considered in any of the trees.

(c) Compute the values of the two probabilities you obtained in parts (b) and (c) for the case where there are $n = 50$ training points with $d = 5$ features each, $T = 25$ trees with $h = 8$ internal nodes each, and we randomly select $m = 1$ potential splitting features in each treenode. You may leave your answer in a fraction and exponentiated form, e.g., $\left(\frac{51}{100}\right)^2$. What conclusions can you draw about the concerns of not considering a feature or sample mentioned at the beginning of the problem?

2 Probabilistic Graphical Models

Recall that we can represent joint probability distributions with directed acyclic graphs (DAGs). Let G be a DAG with vertices $X_1, ..., X_k$. If P is a (joint) distribution for $X_1, ..., X_k$ with (joint) probability mass function p, we say that G represents P if

$$p(x_1, \cdots, x_k) = \prod_{i=1}^{k} P(X_i = x_i | \text{pa}(X_i)), \tag{1}$$

where $\text{pa}(X_i)$ denotes the parent nodes of X_i. (Recall that in a DAG, node Z is a parent of node X iff there is a directed edge going out of Z into X.)

Consider the following DAG

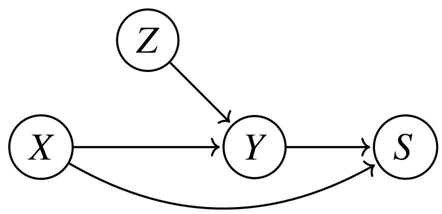

Figure 1: G, a DAG

(a) Write down the joint factorization of $P_{S,X,Y,Z}(s, x, y, z)$ implied by the DAG G shown in Figure 1.

(b) Is $S \perp Z \mid Y$?

(c) Is $S \perp X \mid Y$?

3 PGMs: Sleeping in Class

In this question, you'll be reasoning about a Dynamic Bayesian Network (DBN), a form of a Probabilistic Graphical Model.

Your favorite discussion section TA wants to know if their students are getting enough sleep. Each day, the TA observes the students in their section, noting if they fall asleep in class or have red eyes. The TA makes the following conclusions:

1. The prior probability of getting enough sleep, S, with no observations, is 0.7.
2. The probability of getting enough sleep on night t is 0.8 given that the student got enough sleep the previous night, and 0.3 if not.
3. The probability of having red eyes R is 0.2 if the student got enough sleep, and 0.7 if not.
4. The probability of sleeping in class C is 0.1 if the student got enough sleep, and 0.3 if not.

(a) Formulate this information as a dynamic Bayesian network that the professor could use to filter or predict from a sequence of observations. If you were to reformulate this network as a hidden Markov model instead (that has only a single observation variable), how would you do so? Give a high-level description (probability tables for the HMM formulation are not necessary.)

(b) Consider the following evidence values at timesteps 1, 2, and 3:

(a) $e_1 = (R_1 = 0, C_1 = 0)$ = not red eyes, not sleeping in class
(b) $e_2 = (R_2 = 1, C_2 = 0)$ = red eyes, not sleeping in class
(c) $e_3 = (R_3 = 1, C_3 = 1)$ = red eyes, sleeping in class

Find the likelihood of this sequence of observations. Assume a prior on $P(S_1)$ that is consistent with the prior in the previous part; that is, $P(S_1 = 1) = 0.7$.

(c) Consider the same evidence values at timesteps 1, 2, and 3 as the previous part:

(a) $e_1 = (R_1 = 0, C_1 = 0)$ = not red eyes, not sleeping in class

(b) $e_2 = (R_2 = 1, C_2 = 0)$ = red eyes, not sleeping in class

(c) $e_3 = (R_3 = 1, C_3 = 1)$ = red eyes, sleeping in class

Find the most likely sequence of hidden states S_t that produced the evidence above.