CS57800 Statistical Machine Learning
Homework 4

November 27, 2015

1 Probability (20 pts)

Let X, Y, and Z be random variables taking values in {0, 1}. The following table lists the probability of each possible assignment of 0 and 1 to the variables (e.g., P(X = 0,Y = 1,Z = 0) = 1/10)

	Z=0	Z=0	Z=1	Z=1
	X=0	X=1	X=0	X=1
Y=0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{2}{15}$
Y=1	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{8}{45}$	$\frac{4}{45}$

1. Is X independent of Y ? Why or why not?
2. Is X conditionally independent of Y given Z (explain)?
3. Calculate $P(X = 0 \mid X + Y > 0)$.

2 Hidden Markov Model (30 pts)

We would like to build a very simple machine-learning-based weather report. To accomplish that we will build a probabilistic model, based on one sensory input tracking the location of a groundhog, which can either be Inside its burrow (I), Outside (O), or in-between, only its Head sticking out of the burrow (H). Each day can either be Nice (N) or Cold (C). We will build a first-order Markov Model, by assuming that on nice days the groundhog will be Outside with a probability of 2/3, Inside with a probability of 1/6, or stick its head out with a probability of 1/6. Cold days will have the groundhog Outside with a probability of 1/4, Inside with a probability of 1/2, and have its Head out with a probability of 1/4. If today is Nice, it will stay nice tomorrow with a probability of 2/3. If it's Cold, it will stay cold tomorrow with a probability of 1/2. We will start our weather station next year in the first day of the Fall, when it's known that the probability of Nice and Cold days is equal.

1. Write the HMM model described above formally, provide the probability distributions required to fully define the model.

2. Suppose that we observe the following (daily) groundhog behaviors: I,H,I,O. What is the most likely states (daily weather reports), that would produce these observations?

3 Naïve Bayes (50 pts)

We define the Boolean threshold function over a 7 dimensional Boolean cube $f_{TH(3,7)}$, as follows: given an instance x, $f_{TH(3,7)} = 1$ if and only if 3 or more of x's components are 1.

1. Show that $f_{TH(3,7)}$ has a linear decision surface over the 7 dimensional Boolean cube.
2. Assume that you are given data sampled according to the uniform distribution over the Boolean cube $\{0,1\}^7$ and labeled according to $f_{TH(3,7)}$. Use naïve Bayes to learn a hypothesis that predicts these labels. What is the hypothesis generated by the naïve Bayes algorithm? (Assume you are given the data required to get accurate probability estimates. No need to run the algorithm though).
3. Show that the hypothesis produced in the previous question does not represent this function.
4. Are the naïve Bayes assumptions satisfied by $f_{TH(3,7)}$? Justify your answer.