Weekly Homework 7

Ava Chong
CS 1675: Intro to Machine Learning

March 21, 2019

Problem 1. Decision trees

(a) The restricted decision tree had an error of .2576 while the unrestricted decision tree had an error of .2751. Based on these raw numbers alone, we would chose the restricted tree because it has a better prediction rate. Backpruning will help reduce overfitting and therefore increase the model's power.

(b) First, I played with the minimum parent size and found that I was able to lower the error to .2314 – this was true for all parameter sizes between 50 and 100. Anything above 100 would cause the error to jump up higher and I assumed that this was because this exceeded the total number of parameters. This makes sense because we were no longer providing enough parameters. Changing the other variables made minimal effect and .2314 was the lowest error I could find.

Problem 2. Probabilities: Bayes theorem

	Disease (0.01%)	No Disease
Test Pos	49.5%	.05%
Test Neg	.05%	49.5%

Since 99% of the time the test gives the correct result and disease-positive-test and no-disease-negative-test are the same, we assumed that each we 99% divided by 2 which equals 49.5% I also assumed that since they were symmetric, they would also be symmetric the other way. There really isn't anymore information to assume otherwise.

Problem 3. Bayesian belief networks: foundations
Given these two equations
1. $P(X, Y|Z) = P(X|Z)P(Y|Z)$
2. $P(X|Y, Z) = P(X|Z)$

- Divide by the first equation by $P(Y|Z)$ and expand using the definition of conditional probability.

$P(X,Y|Z)/P(Y|Z) = P(X|Z)$
$[P(Z|X)P(X)P(Z|Y)P(Y)]/[P(Z) * P(Y|Z)] = P(X|Z))$

Simplify using $P(Z|X) + P(Z|Y) = P(Z)$ and $P(Y|Z)P(Z) = P(Y + Z) = P(Y)P(Z)$
$P(Y,X,Z)/P(Y,Z) = P(X|Z)$
$P(X|Y,Z) = P(X|Z)$

We find that using the definition of conditional probability we can prove the first and second equations imply each other.