CS1675 - Assignment 7

Zachary M. Mattis

March 21, 2019

I. Problem 1 - Decision Trees

a. Restricted vs. Unrestricted

The unrestricted tree is very dense, and may be prone to over-fitting. In this case, the unrestricted tree had a test error of 0.2751, while the restricted tree had a 0.2576 error. In general, to backpruning is a useful tool, as it improves predictive accuracy by the reduction of overfitting.

b. fitctree

Min Parents	**20**	**25**	**30**
15	0.2314	-	0.2227
20	0.2358	0.2358	0.2358
25	0.2402	0.2402	0.2707

Table 1: Error = f(Min_Parents, Max_Split)

Min Leaf	**5**	**10**	**15**
15	0.2358	0.2358	0.2227
20	0.2445	0.2489	0.2533
25	0.2489	0.2489	0.2445

Table 2: Error = f(Min_Leaf, Max_Split)

Prune, Parent, Leaf	**1, 20, 10**	**1, 20, 15**	**2, 20, 10**
15	0.2314	0.2314	0.2314
20	0.2227	0.2227	0.2314
25	0.2227	0.2140	0.2314

Table 3: Error = f(Prune, Parent, Leaf, Max_Split)

The lowest error result I obtained was 0.2009, using max splits=35, leaf=7, parent=20, pruning level 1.

II. Problem 2 - Probabilities: Bayes' Theorem

$P(disease) = 0.0001$
$P(healthy) = 0.9999$
$P(test = +|disease) = 0.99$
$P(test = +|healthy) = 0.01$
$P(test = -|disease) = 0.01$

$P(test = -|healthy) = 0.99$

$P(disease|test = +) = \frac{P(test=+|disease)P(disease)}{P(test=+)}$

$$\begin{aligned} P(test = +) &= P(test = +|disease)P(disease) + P(test = +|healthy)P(healthy) \\ &= (0.99)(0.0001) + (0.01)(0.9999) \\ &= 0.010098 \end{aligned}$$

$P(disease|test = +) = \frac{(.99)(.0001)}{0.010098} = 0.0098$

Given a less than 1% chance that somebody who tested positive for disease actually suffers from the disease, it is not advisable for this test to become widely adopted.

III. Bayesian Belief Networks

1. $P(X, Y|Z) = P(X|Z)P(Y|Z)$
2. $P(X|Y, Z) = P(X|Z)$

$$\begin{aligned} P(X|Y,Z) &= \frac{P(X,Y,Z)}{P(Y,Z)} && \text{(cond. prob.)} \\ &= \frac{P(X,Y|Z)P(Z)}{P(Y,Z)} && \text{(product rule)} \\ &= \frac{P(X|Z)P(Y|Z)P(Z)}{P(Y,Z)} && \text{(cond. ind.)} \\ &= \frac{P(X|Z)P(Y,Z)}{P(Y,Z)} && \text{(cond. prob.)} \\ &= P(X|Z) \end{aligned}$$