Programming Assignment 1

Siddarth Sampangi
September 22, 2015

1 DECISION TREES

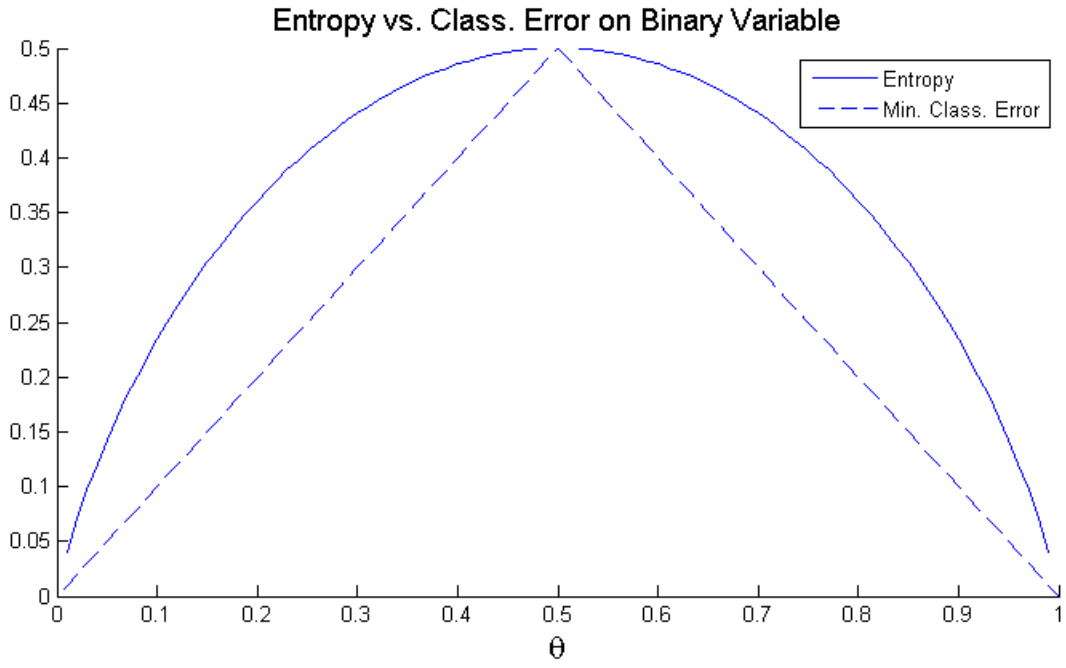
1.a ENTROPY AND CLASSIFICATION ERROR

1.a.1

$$\begin{aligned} H(Y) &= -\sum_{y \in \{yes,no\}} p(y) log_e p(y) \\ &= -[\theta log_e(\theta) + (1-\theta) log_e(1-\theta)] \\ &= -\theta log_e(\theta) - (1-\theta) log_e(1-\theta) \end{aligned}$$

1.a.2 The best classification error would be 0.

1.a.3



1.b TRAIN A DECISION TREE

$$P(Y = y|X = x) = \frac{Count(Y = y, X = x)}{Count(X = x)} \qquad P(X = x_1, Y = y_1) = \frac{Count(X = x_1, Y = y_1)}{\sum_{x \in X, y \in Y} Count(X = x, Y = y)}$$

P(Y=Yes)	0.32
P(Y=No)	0.68

P(Y|X)	A=Child	A=Adult	C=First	C=Lower	G=Male	G=Female
Y=Yes	0.52	0.31	0.62	0.27	0.21	0.73
Y=No	0.48	0.69	0.38	0.73	0.79	0.27

P(Y,X)	A=Child	A=Adult	C=First	C=Lower	G=Male	G=Female
Y=Yes	0.03	0.30	0.09	0.23	0.17	0.16
Y=No	0.02	0.65	0.06	0.62	0.62	0.06

$$\begin{aligned} H(Y) &= -\theta log_e(\theta) - (1-\theta) log_e(1-\theta) \\ &= (0.32) log_e(0.32) - (1-(0.32)) log_e(1-(0.32)) \\ &= 0.63 \end{aligned}$$

$$\begin{aligned} I(A;Y) &= H(Y) - H(Y|A) \\ &= H(Y) + \sum_{a,y} p(a,y) log_e p(y|a) \\ &= 0.63 + [0.03 log_e 0.52 + 0.02 log_e 0.48 + 0.30 log_e 0.31 + 0.65 log_e 0.69] \\ &= 0.003 \end{aligned}$$

$$\begin{aligned} I(C;Y) &= H(Y) - H(Y|C) \\ &= H(Y) + \sum_{c,y} p(c,y) log_e p(y|c) \\ &= 0.63 + [0.09 log_e 0.62 + 0.06 log_e 0.38 + 0.23 log_e 0.27 + 0.62 log_e 0.73] \\ &= 0.03 \end{aligned}$$

$$\begin{aligned} I(G;Y) &= H(Y) - H(Y|G) \\ &= H(Y) + \sum_{g,y} p(g,y) log_e p(y|g) \\ &= 0.63 + [0.17 log_e 0.21 + 0.62 log_e 0.79 + 0.16 log_e 0.73 + 0.06 log_e 0.27] \\ &= 0.09 \end{aligned}$$

The feature with the highest information gain is gender.

CODE

PROBLEM 1.A.3

```
x = linspace(0,1,101);
for i=1:101
    y(i)=-1*(x(i)*log(x(i))+(1-x(i))*log(1-x(i)));
    z(i)=min(x(i),1-x(i));
end
m = 0.5/max(y);
for i=1:101
    y(i) = y(i)*m;
end
f = figure;
hold on;
plot(x,y);
plot(x,z, '--');
title('Entropy vs. Class. Error on Binary Variable', 'FontSize',14);
xlabel('\theta', 'FontSize',15);
legend('Entropy','Min. Class. Error');
```