Vibhor Jain
3/17/13
CSE 176
Prof. David Noelle
TA Jimei Yang

Problem Set #2

1.
a)
Entropy_S= -(3/6)*log_2(3/6)-(3/6)*log_2(3/6)=1

b)
Entropy_a2_T=-(2/4)*log_2(2/4)-(2/4)*log_2(2/4)=1
Entropy_a2_F==-(1/2)*log_2(1/2)-(1/2)*log_2(1/2)=1

Gain=Entropy_S-(4/6)*Entropy_a2_T-(2/6)*Entropy_a2_F=0

2.
In the lazy learning version of ID3, you start off with any arbitrary example and draw a chain (the nodes representing the attribute number and the links representing the attribute value), where the end leaf node contains the classification of the first example. Then for the subsequent examples, you build off your initial chain by drawing subtrees from nodes where the subsequent examples deviate from the previous examples (in terms of their attributes and classifications).

The advantage is that this scheme will accurately classify any possible example the algorithm has seen before. The disadvantage is that the algorithm might cost too much space and will only be able to classify correctly the examples it has seen before.

5.
a) A perceptron may not be able to learn this task. A perceptron hopes to learn a single plain that can separate all positive from all negative categorizations, which may not be possible with a single plane.

b) Yes, a decision tree learning algorithm can learn this task because the tree would be able to weed out the attributes that would not have a relevant impact on the classification of the data.

c) The decision tree learning algorithm would do a better job of generalizing to novel inputs. The perceptron learns a linear equation (which may or may not help in categorizing the data well) while the decision tree learning algorithm constructs its trees based on the variation of the data based on its attributes.

6.
a)

$g(x)=1/(1+\exp(-x))$

$$\begin{aligned}g'(x) &= \exp(x)/((1+\exp(x))^2) \\ &= \exp(-x)/((1+\exp(-x))^2) \\ &= (\exp(-x)/ (1+\exp(-x)))*g(x) \\ &= [1-g(x)]g(x)\end{aligned}$$

$g'(x)=[1-g(x)]g(x)$

b)

$g(x)=\tanh(x)=\sinh(x)/\cosh(x)=\sinh(x)*\text{sech}(x)$

$$\begin{aligned}g'(x) &= \sinh(x)'*\text{sech}(x)+\sinh(x)*\text{sech}(x)' \\ &= \cosh(x)*\text{sech}(x)+\sinh(x)*(-\tanh(x)*\text{sech}(x)) \\ &= \cosh(x)/\cosh(x)-\sinh(x)*\tanh(x)/\cosh(x) \\ &= 1-\tanh(x)*\tanh(x)\end{aligned}$$

$g'(x)=1-g(x)^2$

comments:
I used Tom Mitchell's *Machine Learning* to answer these questions