HW Assignment 4 (Due date: Dec 2, Friday)

1 Problems

1. [**Decision Trees**, 50 points]
   Consider the following set of training examples:

A_1	A_2	A_3	Label
F	F	F	$-$
F	T	F	$-$
T	F	F	$+$
T	F	T	$+$
F	T	F	$-$
F	F	T	$-$
T	F	T	$+$
F	T	F	$-$

   Table 1: Training examples

   1. What is the entropy of the class label distribution?
   2. What attribute gets selected as root by the ID3 algorithm?

2. [**Naive Bayes**, 50 points]
   The Naive Bayes algorithm for text categorization presented in class (slide 16) treats all sections of a document equally, ignoring the fact that words in the title are often more important than words in the text in determining the document category. Describe how you would modify the Naive Bayes algorithm for text categorization to reflect the constraint that words in the title are K times more important than the other words in the document for deciding the category, where K is an input parameter (include pseudocode).

3. [**Logistic Regression**, 50 points]
   By setting the gradient of the log-likelihood to zero, prove that the ML solution $\mathbf{w}_{ML}$ for the multiple class logistic regression problem satisfies the constraint that, for every feature ϕ_i, the observed value of ϕ_i on the training data D is the same as its expected value on D under the probability distribution parameterized by $\mathbf{w}_{ML}$ (i.e. the constraint on slide 14).

4. [**Logistic Regression**, 50 points]
Assume that a binary feature ϕ_i is equal to 1 for all training examples belonging to a particular class C_k, and zero otherwise (i.e. ϕ_i perfectly separates examples from class C_k from all other examples). Show that in this case the magnitude of the ML solution for $\mathbf{w}$ goes to infinity, thus motivating the use of a prior over the parameters (Hint: use constraint equation on slide 14).

5. [**k-Means Clustering**, 50 points]
Prove that the value of the objective function of k-Means decreases at every iteration i.e. $J^{(t+1)} \leq J^{(t)}$. Use this observation to prove that k-Means converges after a finite number of iterations.

6. [**Hidden Markov Models**, 50 *bonus points*]
Using the notation introduced in class, show that:

$$p(O|\mu) = \sum_{i=1}^{N} \alpha_i(t)\beta_i(t) \qquad\qquad \forall t \in [1..T]$$