First Name: ____________________ Last Name: ____________________

Sample Machine Learning Questions, 2010

Note: This sample exam is to indicate the style of questions you should expect. Check the topic list for the particular topics to be covered on your exam. The actual exam will have space for your answers and is likely to contain 50-60 points of questions.

1. (8 pts) Short answer.

   (a) What is *overfitting*?

   (b) Formally, when are two events A and B (statistically) *independent*?

   (c) Assume we have a two-class (+, −) classification problem and are asked to predict the label of a new instance x where we know $P(+ \mid x) = 0.6$ and $P(- \mid x) = 0.4$. How can the *Bayes optimal prediction* on x be "−"?

   (d) Which one of the following statements is most applicable for the perceptron algorithm?

   Which statement is most applicable for logistic regression?

   (note: x represents the features and y represents the label)

   - learns a $P(y \mid x)$ model?
   - uses a $P(x \mid y)$ assumption?
   - is not associated with conditional probabilities?

2. (3 pts) Describe the Backpropagation algorithm used to train Neural Networks.

3. (6 pts) Short answer, but show enough work so I can follow your calculations.

   (a) Assume that the instance consist of a single real-valued feature. What Gaussian (Normal distribution with mean μ and standard deviation σ) maximizes the likelihood of the four point dataset $\{-1, 1, 1, 7\}$?

   (b) Naive Bayes. Consider the Naive Bayes algorithm when trained on the following dataset:

x_1	x_2	x_3	label
1	1	1	+
1	0	1	+
1	1	0	+
1	0	1	−
0	0	0	−

   What prediction does it make on the instance $x_1 = 1$, $x_2 = 0$, $x_3 = 0$.

4. (5 pts) Assume we have a coin that has some unknown probability h of coming up heads (and probability $1 - h$ of coming up tails). If the coin is flipped four times getting three heads and a tail (HHHT) then:

   (a) (2 pts) What is the maximum likelihood estimate for h?

   (b) (3 pts) Assume that (before flipping the coin) we have a prior density $p(h)$ for the various values of $h \in [0, 1]$. Give the formula for the posterior probability density $p(h \mid \text{HHHT})$ as a function of the prior $p(h)$.

5. (4 pts) Assume that we are running AdaBoost on a sample of 3 examples $\{(x_1, +1), (x_2, +1), (x_3, -1)\}$, and after iteration T we have a master hypothesis $H_T(x_i) = \sum_t^T \alpha_t h_t(x_i)$ as follows (recall that the h_t's are the weak hypotheses produced during previous iterations):

x_i	$H(x(i))$
x_1	$\ln(2)$
x_2	$-\ln(3)$
x_3	$-\ln(2)$

   Find the distribution on these points used by AdaBoost in the next iteration (2 pts).

   Next (2 pt), suppose the next weak hypothesis h_{T+1} seen by AdaBoost correctly labels all four of the points. How will h_{T+1} be incorporated into the new master hypotheses H_{T+1}?