Introduction to Machine Learning **Fall Semester, 2013/14**

Homework 3: Nov 12th, 2013

Due: Nov 26th (See the submission guidelines in the course web site)

Theory Questions

1. PAC-learn Parity:
   Let the Parity concept class C be the set of all xor's over n literals $\{x_1, \ldots, x_n\}$:
   $C = \{x_{i_1} \oplus x_{i_2} \oplus \ldots \oplus x_{i_l} | i_j \in \{1, \cdots n\}\}$

   (a) What would be a sufficient size m of a sample set $S = \{(x^i, h(x^i))\}_{i=1}^m$ that is consistent with an (unknown, to be learned) $h \in C$ such that with probability at least $1 - \delta$ an algorithm that outputs a hypothesis h_S consistent with S has $err(h_S) < \epsilon$?
   (note $x^i = (x_1^i, \ldots, x_n^i)$ where x_j^i is the binary value of the literal x_j in sample i).

   (b) Describe an efficient (ϵ, δ)-PAC algorithm that learns the Parity class (that is, that efficiently finds a consistent hypothesis and requires a number of samples that is ploynomial in $\frac{1}{\epsilon}$ and $\log \frac{1}{\delta}$ to achieve δ-confidence and ϵ-accuracy).

2. Learning Triangles:
   Let the Triangles concept class be the set
   $TR = \{ET \subset R^2 | ET$ is an equilateral triangle with a base parallel to the X-axis$\}$.
   Describe an efficient PAC-learning algorithm for the class TR. Prove the required sample complexity.

3. Perceptron:
   Show that the perceptron mistake bound that was proved in class is tight in a sense. That is, for any m find a linearly separable sequence with margin γ of m samples $S = \{(x_i, y_i)\}_{i=1}^m$ where $||x_i|| = 1, y_i \in \{-1, 1\}$ such that the number of updates made by the perceptron algorithm is exactly $\frac{1}{\gamma^2}$. Hint: let m be the number of features (i.e. the dimenssion of the samples space).

Programming Assignment

Write a program to implement the Perceptron algorithm. Run the program on the mnist data sets in http://www.cs.tau.ac.il/∼ mansour/ml-course-10/mnist.html. (The data has handwritten digits of the number "4" and number "7".) File format: Each line is a different image. The images are 28 28 with each entry having a gray level. This implies that an image has 784 entries. There are 1000 training examples and 200 test examples. The training examples are in Xtrain and their labels are in Ytrain. The testing examples are in Xtest and their labels are in Ytest. Run the algorithm on six random permutations of the training data. Indicate test and training error for each run. Discuss your findings.