Machine Learning: Foundations **Fall Semester, 2012/13**

Homework 1: Nov 4, 2012

Lecturer: Yishay Mansour

Homework number 1.

Theory question I: Consider a collection of points, $x_1, \cdots x_m$, sampled i.i.d. according to an Exponential Distribution. (Recall that an exponential distribution has a parameter λ and the density of x is $\lambda e^{-\lambda x}$.)

1. What is the Maximum Likelihood (ML) estimate of λ.
2. What is the Maximum A Posteriori (MAP) value of λ given that its prior distribution is exponential with parameter 1.
3. **(BONUS)** Compute the Bayesian posterior distribution for an exponential distribution where λ is distributed according to a Gamma distribution, i.e.,

$$\mathrm{Gamma}(\lambda\,;\,\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\,\lambda^{\alpha-1}\,\exp(-\lambda\,\beta).$$

(Comment: when maximizing, remember to check also the second derivative.)

Theory question II: Consider a countable hypothesis class H with a distribution $p(h) > 0$ for $h \in H$, and an arbitrary target function f. Show that with probability $1-\delta$, for every $h \in H$, we have

$$|\widehat{error}(h) - error(h)| \leq \sqrt{\frac{-c\log(p(h)\delta)}{m}}$$

for some constant $c > 0$, and where the probability is taken over a sample of size m. (Comment: Note that H might be infinite.)

Theory question III: Let $X = [0,1]$. The class $C = \{c_z : z \in [0,1]\}$ where $c_z(x) = 1$ iff $x \leq z$. Let $H = C$. Give an efficient PAC learning algorithm for learning C using H.
The algorithm needs to be both polynomial time and polynomial sample. Try to get the best sample complexity, up to a multiplicative constant.

Theory question IV: Let $X = [0,1)$. The class C_k includes concepts that switches k times. Each $c \in C_k$ has k parameters, $0 = z_1 < z_2 < \cdots < z_k = 1$. The label of a point x which is in the interval $[z_i, z_{i+1})$ is $c(x) = i \bmod 2$. Consider a set S of m examples (x_i, b_i) where $x_i \in [0,1)$ and $b_i \in \{0,1\}$. Compute efficiently a concept $c \in C_k$ which minimizes the error on S. The running time should be polynomial in both m and k. (HINT: Use dynamic programming. Also, note that the x_is need not be unique.)

Programming assignment:

You will need to write the programming assignments in MATLAB.

Write a program to implement the Naive Bayes algorithm with Gaussian distribution. Run the program on the `iris` data set at the home page of the course.

File format: For each database you have a data file and an information file, that describing the input. In a nutshell, each line in the data file is a record (example), and the last value in the line is the one you would like to predict.

Methodology: You split the file randomly to two parts. One part is the *training set*, which is used in order to learn the classifier. The other part is the *test set*, which evaluates the classifier you learned.

First, you take the training set and learn a classifier. Then you take the test set and evaluate the error rate of the classifier. The size of the training set should be a parameter that you can adjust.

Output: In your output plot two different error functions as a function of the fraction of points in the training set. (Start with very small training set, say 0.05.) Each point should be an average of 10 runs (random splits of training and testing). The error functions are:

1. The error frequency on the test set.

2. The error frequency on the training set.

See the instructions for handing in programming assignment. Also, look at the comments on homework 1 in the previous course, especially on the programming.

The homework is due in two weeks