University of Pittsburgh
CS2750 Machine Learning
Professor Milos Hauskrecht

Handout 2
January 29, 2003

Problem assignment 2

Due: Wednesday, February 5, 2003

Problem 1. Comparison of predictors

We have two predictors A,B that predict the value of a target real-valued attribute. The predictors are run on the test data and the results are listed in file 'predictor.txt', where the first column represents the target value, the second is the value predicted by A and the third column is the output of B.

Using Matlab calculate and report:

- (a) Mean squared error of both predictors;
- (b) Confidence intervals for $p = 0.95$ for the mean squared errors of both predictors. Hint: use Matlab's function tinv or norminv.
- (c) Which predictor appears to be better in the squared error sense. What is the (probabilistic) support for this hypothesis? Hint: use Matlab's functions tcdf or normcdf.

Problem 2. Poisson distribution

The Poisson distribution is used to model the number of random arrivals to a system over a fixed period of time. Examples of systems in which events are determined by random arrivals are: arrivals of customers requesting the service, occurence of natural disasters, such as floods, etc. The Poisson distribution is defined as:

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2, \cdots.$$

Answer the following questions:

- (a) Using the definition of the Possion distribution show that the sum of probabilities of all events is 1. (Hint: check how e^λ is defined in terms of a sum).

- (b) Derive the mean of the Poisson distribution.
- (c) Assume we have n independent samples of x. What is the ML estimate of the parameter λ.
- (d) The conjugate prior for the Poisson distribution is Gamma distribution. It is defined as:
$$p(\lambda|a,b) = \frac{1}{b^a \Gamma(a)} \lambda^{(a-1)} e^{-\frac{\lambda}{b}}.$$
Show that the posterior density of the parameter λ is again a Gamma distribution.
- (e) Show that the Poission distribution is a member of the exponential family of distributions. Give $\eta, T(x), Z(\eta)$ and $h(x)$ components.

Now we are ready to do some Matlab experiments:

- (f) plot the probability function for Poisson distributions with parameters $\lambda = 2$ and $\lambda = 6$. Note that the Poisson model is defined over nonenegative integres only.
- (g) Assume the data in 'poisson.txt' that represent the number of incoming phone calls received over a fixed period of time. Compute and report the ML estimate of the parameter λ.
- (h) Assume the prior on λ is given by $\lambda \sim Gamma(a, b)$. Plot the Gamma distribution for the following set of parameters $(a = 1, b = 2)$ and $(a = 3, b = 5)$.
- (g) Plot the posterior density for λ after observing samples in 'poission.txt' and using priors in part (h). What changes in the distribution do you observe?