MLE and MAP

Introduction to Machine Learning (CSCI 1950-F), Summer 2011

Solve the following problems. Provide mathematical justification for your answers.

In the following series of exercises you will derive ML and MAP estimators for a simple model of an uncalibrated sensor. Suppose the sensor output, X, is a random variable that ranges over the real numbers. Assume that $X \sim \text{Uniform}(0, \theta)$ for some unknown $\theta > 0$. That is, the sensor's outputs are uniformly distributed on some unknown interval $(0, \theta)$, so that the density is

$$\begin{aligned} p(x|\theta) &= \begin{cases} 1/\theta & \text{if } 0 < x < \theta \\ 0 & \text{otherwise} \end{cases} \\ &= \frac{1}{\theta} I_{(0,\theta)}(x) \end{aligned}$$

Here, $I_{(0,\theta)}(x)$ denotes an *indicator function* which equals 1 when $0 < x < \theta$, and 0 otherwise.

You are given data $D = (x_1, \dots, x_n)$, and to characterize the sensor's sensitivity, you would like to estimate θ.

(1) MLE for θ

Modeling the data using $X_1, \dots, X_n \sim \text{Uniform}(0, \theta)$ i.i.d., what is the likelihood function $p(D|\theta)$? What is the maximum likelihood estimate (MLE) for θ? Give an informal proof that your answer is in fact the MLE.

(2) Pareto prior

Suppose that we place the following prior distribution on θ:

$$p(\theta) = \alpha \beta^\alpha \theta^{-\alpha-1} I_{(\beta,\infty)}(\theta).$$

(This is known as a *Pareto distribution*. We denote that θ has this distribution by writing $\theta \sim \text{Pareto}(\alpha, \beta)$.) Plot the three prior probability densities corresponding to the following three hyperparameter choices: $(\alpha, \beta) = (0.1, 0.1)$; $(\alpha, \beta) = (2.0, 0.1)$; $(\alpha, \beta) = (1.0, 2.0)$.

(3) Posterior distribution for θ

Now, you model the data using $\theta \sim \text{Pareto}(\alpha, \beta)$ and $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$ conditionally independent given θ. Derive the posterior distribution $p(\theta|D)$. Does it belong to any family of distributions that you recognize?

(4) MAP estimate for θ

Using the posterior derived in the previous exercise, what is the MAP estimate of θ? How does this compare to the MLE?

(5) Optimal θ under square loss

Recall that the square loss is defined as $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$. For the posterior derived in exercise (3), what estimator of θ minimizes the posterior expected square loss? Simplify your answer as much as possible. Is it the same as the MLE and/or the MAP?

(6) Visualizing an example

Suppose you observe the data $D = (0.7, 1.3, 1.7)$. Plot the density of the posterior distribution of θ for each of the three priors in exercise (2). For each of these three priors, what is the MAP estimate? For each of these three priors, what estimate (as derived in exercise (5)) minimizes the posterior expected square loss?