19 June 2023

The 2nd quiz in Pattern Recognition and Machine Learning

Due 26 June 2023

Exercise 3.1 MLE for the Bernoulli/ binomial model

Derive Equation 3.22 by optimizing the log of the likelihood in Equation 3.11.

$$p(\mathcal{D}|\theta) = \theta^{N_1}(1-\theta)^{N_0} \tag{3.11}$$

$$\hat{\theta}_{MLE} = \frac{N_1}{N} \tag{3.22}$$

Exercise 3.11 Bayesian analysis of the exponential distribution

A lifetime X of a machine is modeled by an exponential distribution with unknown parameter θ. The likelihood is $p(x|\theta) = \theta e^{-\theta x}$ for $x \geq 0$, $\theta > 0$.

a. Show that the MLE is $\hat{\theta} = 1/\overline{x}$, where $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

b. Suppose we observe $X_1 = 5, X_2 = 6, X_3 = 4$ (the lifetimes (in years) of 3 different iid machines). What is the MLE given this data?

Use the values in b. in what follows

c. Assume that an expert believes θ should have a prior distribution that is also exponential

$$p(\theta) = \text{Expon}(\theta|\lambda) = \text{Ga}(\theta|1, \lambda) \tag{3.98}$$

Choose the prior parameter, call it $\hat{\lambda}$, such that $\mathbb{E}[\theta] = 1/3$. Hint: recall that the Gamma distribution has the form

Check the normalizing constant

$$\text{Ga}(\theta|a, b) \quad \propto \quad \theta^{a-1}e^{-\theta b} \tag{3.99}$$

and its mean is a/b.

Check the mean of Gamma dist.

d. What is the posterior, $p(\theta|\mathcal{D}, \hat{\lambda})$?

e. Is the exponential prior conjugate to the exponential likelihood?

f. What is the posterior mean, $\mathbb{E}\left[\theta|\mathcal{D}, \hat{\lambda}\right]$?

g. Explain why the MLE and posterior mean differ. Which is more reasonable in this example?

Exercise 3.14 Posterior predictive for Dirichlet-multinomial

(Source: Koller.).

a. Suppose we compute the empirical distribution over letters of the Roman alphabet plus the space character (a distribution over 27 values) from 2000 samples. Suppose we see the letter "e" 260 times. What is $p(x_{2001} = e|\mathcal{D})$, if we assume $\boldsymbol{\theta} \sim \text{Dir}(\alpha_1, \ldots, \alpha_{27})$, where $\alpha_k = 10$ for all k?

b. Suppose, in the 2000 samples, we saw "e" 260 times, "a" 100 times, and "p" 87 times. What is $p(x_{2001} = p, x_{2002} = a|\mathcal{D})$, if we assume $\boldsymbol{\theta} \sim \text{Dir}(\alpha_1, \ldots, \alpha_{27})$, where $\alpha_k = 10$ for all k? Show your work.