Foundations of Machine Learning

Brett Bernstein

August 22, 2018

Conditional Probability Models: Concept Check

Conditional Probability Models

MLE Learning Objectives

- Define the likelihood of an estimate of a probability distribution for some data $\mathcal{D}$.
- Define a parameteric model, and some common parameteric families.
- Define the MLE for some parameter θ of a probability model.
- Be able to find the MLE using first order conditions on the log-likelihood.

Conditional Probability Models

- Describe the basic structure of a linear probabilistic model, in terms of (i) a parameter θ of the probablistic model, (ii) a linear score function, (iii) a transfer function (kin to a "response function" or "inverse link" function, though we've relaxed requirements on the parameter theta).
- Explain how we can use MLE to choose w, the weight vector in our linear function (in (ii) above).
- Give common transfer functions for (i) bernoulli, (ii) poisson, (iii) gaussian, and (iv) categorical distributions. Explain why these common transfer functions make sense (in terms of their codomains).
- Explain the equivalence of EMR and MLE for negative log-likelihood loss.

MLE/Conditional Probability Model Concept Check Question

1. In each of the following, assume $X_1, \ldots, X_n$ are an i.i.d. sample from the given distribution.

(a) Compute the MLE for p assuming each $X_i \sim \text{Geom}(p)$ with PMF $f_X(k) = (1-p)^{k-1}p$ for $k \in \mathbb{Z}_{\geq 1}$.

(b) Compute the MLE for λ assuming each $X_i \sim \text{Exp}(\lambda)$ with PDF $f_X(x) = \lambda e^{-\lambda x}$.

2. We want to fit a regression model where $Y|X = x \sim \text{Unif}([0, e^{w^T x}])$ for some $w \in \mathbb{R}^d$. Given i.i.d. data points $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$, give a convex optimization problem that finds the MLE for w.

3. Explain why softmax is related to computing the maximum of a list of values.

4. Suppose x has a Poisson distribution with unknown mean θ:

$$p(x|\theta) = \frac{\theta^x}{x!}\exp(-\theta), \qquad x = 0, 1, \cdots$$

Let the prior for θ be a gamma distribution:

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)}\exp(-\beta\theta), \qquad \theta > 0$$

where Γ is the gamma function. Show that, given an observation x, the posterior $p(\theta|x, \alpha, \beta)$ is a gamma distribution with updated parameters $(\alpha', \beta') = (\alpha + x, \beta + 1)$. What does this tell you about the Poisson and gamma distributions?