DS 102 Discussion 4

Monday, February 24, 2020

In this discussion, we'll continue to develop intuition and experience with how expectation-maximization (EM) allows us to fit model parameters by approximating maximum likelihood estimation. In particular, EM comes in handy when our models involve *latent variables*, or variables we never actually observe in the data, which are common when we try to model complex phenomena.

Consider the beta-binomial model:

$$\begin{aligned} Z &\sim \text{Beta}(\alpha, \beta) \\ X &\sim \text{Binomial}(n, Z) \end{aligned}$$

where the integer n is considered fixed and known, and $\alpha, \beta > 0$ are the two parameters. You can think of the beta-binomial as randomly picking the bias Z of a coin, then flipping that coin n times and observing how many heads show up. In practice, it's often used to model data where a binomial would seem appropriate, but the data has higher variance than a vanilla binomial random variable. The randomness in picking p captures that increased variance.

Suppose we're interested in the distribution of batting averages in Major League Baseball, which we model as a beta distribution with unknown positive parameters α and β. That is, each player's true batting average is a value $X \in [0, 1]$ drawn from this distribution. However, we don't actually observe each player's true batting average. Instead, over the course of a season we observe X, the number of hits out of n total pitches.

The two steps of EM are motivated by two insights.

- *Expectation (E) step*: $q^{(t)} \leftarrow \mathbb{P}(Z \mid X, \alpha^{(t)}, \beta^{(t)})$. If you knew α, β, it'd be straightforward to compute $\mathbb{P}(Z \mid X, \alpha, \beta)$ (which we'll show). This can be interpreted as imputing the "missing values" of Z that you didn't observe.

- *Maximization (M) step*: $\alpha^{(t)}, \beta^{(t)} \leftarrow \operatorname{argmax}_{\alpha,\beta>0} \mathbb{E}_{Z\sim q^{(t)}}[\log \mathbb{P}(X, Z \mid \alpha, \beta)]$. The insight behind this is that if you knew Z, it'd be straightforward to find the α, β that maximize $\mathbb{P}(X, Z \mid \alpha, \beta)$ (which we'll show).

1. For the E-step, derive the probability density function of the posterior $p(z \mid x, \alpha, \beta)$. Recall that the probability density function of the Beta(α, β) distribution is given by

   $$p(z \mid \alpha, \beta) = \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha, \beta)}$$

   where $B(\alpha, \beta)$ is the normalizing constant.

   What fact about the beta and binomial distributions have we recovered?

2. Now we derive the maximization step.