CPSC 540 Notes on Univariate Gaussian MLE

Mark Schmidt

Fall 2014

Consider a dataset $\mathcal{D} = \{x_i\}_{i=1}^n$, where each $x_i \in \mathbb{R}$. We'll model the x_i as IID draws from a Gaussian distribution,

$$x_i \sim \mathcal{N}(\mu, \sigma^2),$$

so that

$$p(x_i|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

and we'll fit the parameters μ and σ using the MLE. The likelihood is

$$p(\mathcal{D}|\mu, \sigma^2) = \prod_{i=1}^N p(x_i|\mu, \sigma^2),$$

and recall that to maximize the likelihood we can equivalently maximize the log-likelihood

$$\begin{aligned}
\log p(\mathcal{D}|\mu, \sigma^2) &= \log \prod_{i=1}^N p(x_i|\mu, \sigma^2) \\
&= \sum_{i=1}^N \log p(x_i|\mu, \sigma^2) && (\log(ab) = \log(a) + \log(b)) \\
&= \sum_{i=1}^N \log\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right) && (\text{Gaussian assumption}) \\
&= \sum_{i=1}^N -\log(\sigma) - \log(\sqrt{2\pi}) + \log(\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) && (\log(a/b) = \log(a) - \log(b)) \\
&= \sum_{i=1}^N -\log(\sigma) - \log(\sqrt{2\pi}) - \frac{(x_i-\mu)^2}{2\sigma^2} && (\log(\exp(a)) = a)
\end{aligned}$$

The derivative with respect to μ is

$$\nabla_\mu \log p(\mathcal{D}|\mu, \sigma^2) = \sum_{i=1}^N \frac{x_i - \mu}{\sigma^2}.$$

Setting the derivative to zero to find a staionary point we get

$$0 = \sum_{i=1}^N \frac{x_i - \mu}{\sigma^2},$$

and by multiplying by σ^2 and re-arranging that

$$\sum_{i=1}^{N} \mu = \sum_{i=1}^{N} x_i.$$

Noting that $\sum_{i=1}^{N} \mu = N\mu$ (μ is repeated N times), we get the MLE for μ,

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

This is a kind of tedious way to show that the estimate of μ that maximizes the likelihood is just the average of the data points.[1]

Plugging this in and taking the derivative with respect to σ^2 is

$$\nabla_\mu \log p(\mathcal{D}|\mu, \sigma^2) = \sum_{i=1}^{N} -\frac{1}{\sigma} + \frac{(x_i - \hat{\mu})^2}{\sigma^3}.$$

Setting this equal to zero and re-arranging we get

$$\sum_{i=1}^{n} \frac{1}{\sigma} = \sum_{i=1}^{N} \frac{(x_i - \hat{\mu})^2}{\sigma^3},$$

and multiplying both sides by σ^3 we get

$$\sum_{i=1}^{n} \sigma^2 = \sum_{i=1}^{N} (x_i - \hat{\mu})^2,$$

or that

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^2.$$

[1]Technically we have only shown that this is a stationary point, but it does happen to be a maximizer for $\sigma > 0$ since $\nabla^2 \log p(\mathcal{D}|\mu, \sigma^2) = \sum_{i=1}^{N} \frac{1}{\sigma^2}$, which is positive.