Homework 1 Solutions

May 16, 2011

1 Bishop 1.11

We start with the likelihood of a single point under a (univariate) Gaussian

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{1}$$

for a number of data points $\mathbf{x}$ distributed i.i.d. The i.i.d point is important because it means the data points don't interact with each other [1], so we can use the product rule and write the likelihood as

$$\mathrm{p}(\mathbf{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2) \tag{2}$$

.

The book skips the next step, but we can collect all terms from multiplying out each individual term in equation 1 to yield

$$\mathrm{p}(\mathbf{x}|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$$

.

To get to the next part, we apply a few rules of logarithms

$$\begin{aligned}
\ln(\mathrm{p}(\mathbf{x}|\mu,\sigma^2)) &= \ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}\right) \\
&= \ln(1) - \frac{N}{2}(\ln(2\pi\sigma^2) + \ln\left(\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}\right) \\
&= 0 - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2
\end{aligned}$$

[1] If only this were true in real data!

,

so at the end we get the log likelihood

$$\ln(\mathrm{p}(\mathbf{x}|\mu,\sigma^2)) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2)$$

.

To obtain the maximum likelihood estimator for μ, we set the derivative equal to zero and solve for μ, which will be the maximum [2].

$$\begin{aligned}\frac{\partial f}{\partial \mu}\ln(\mathrm{p}(\mathbf{x}|\mu,\sigma^2)) = -\frac{\partial f}{\partial \mu}\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{\partial f}{\partial \mu}\frac{N}{2}\ln(2\pi) - \frac{\partial f}{\partial \mu}\frac{N}{2}\ln(\sigma^2)\\ \frac{\partial f}{\partial \mu}\ln(\mathrm{p}(\mathbf{x}|\mu,\sigma^2)) = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\frac{\partial f}{\partial \mu}(x_n-\mu)^2 - 0 - 0\\ \frac{\partial f}{\partial \mu}\ln(\mathrm{p}(\mathbf{x}|\mu,\sigma^2)) = \frac{2}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)\end{aligned}$$

We then set this to zero

$$\begin{aligned}0 &= \frac{2}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)\\ 0 &= \sum_{n=1}^{N}(x_n-\mu)\\ 0 &= \sum_{n=1}^{N}x_n - N\mu\\ N\mu &= \sum_{n=1}^{N}x_n\\ \mu &= \frac{1}{N}\sum_{n=1}^{N}x_n\end{aligned}$$

and we have shown that the mean of the observed data points is the maximum likelihood estimator of μ.

[2]Now if you think back to calculus, you'll remember that a derivative of zero is enough to be clear that you are at a maximum, because you have to verify certain things are true of your function to make sure you aren't at a saddle point or a minimum.