Maximum Likelihood Estimates of the Univariate Gaussian Distribution

Maxwell Aladago
February 6, 2019

1 Writing out the Likelihood function

Maximum likelihood estimation asks us to find the parameters θ of a given probability distribution which maximize the probability of an observed data D. More formally, we are asked to determine

$$\arg\max_{\theta} p(\mathcal{D};\theta) \tag{1}$$

The derivations provided here are for the specific case of the univariate Gaussian distribution which has the form

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

where μ, the mean and σ^2, the variance as the parameters controlling the distribution.
If we assume that each $x^{(i)}$ from $\mathcal{D} = \{x^{(1)}, ..., x^{(m)}\}$ is independently sampled from the same Gaussian, we can write the probability of the data as

$$\mathcal{N}(\mathcal{D};\mu,\sigma^2) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right) \tag{3}$$

Eq 3 is the probability of the data. Often, however, we called it the likelihood of the data. For our specific case, the likelihood is given as

$$\mathcal{L}(\mathcal{D}|\mu,\sigma^2) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right) \tag{4}$$

2 The log likelihood

From Eq 1, you can see that all Maximum Likelihood estimation is only interested in the parameters which maximize probability of the data, not the actual maximum probability. Generally, products are more difficult to work with compared to summations. We will benefit enormously if we can somehow convert the products in Eq 4 to sums without changing the values which maximize the likelihood of the data. This is where the logarithmic function becomes very important. First, it empowers us to move from products to sums. Second and more importantly, it's monotonic. This means, that if $a > b$, then $\log a > \log b$. Hence, the parameters which will give us the maximum value of the function in Eq 4 will also give the maximum value of the log of that function. The log of the likelihood is called the log-likelihood and has the form

$$\begin{aligned}
l(\mathcal{D}|\mu,\sigma^2) &= \log\left(\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right)\right) \\
&= \sum_{i=1}^{m} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right)\right) \\
&= \sum_{i=1}^{m} \log\left(2\pi\sigma^2\right)^{-\frac{1}{2}} + \sum_{i=1}^{m} \log\left(\exp\left(-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right)\right) \\
&= \sum_{i=1}^{m} -\frac{1}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{m} -\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2 \\
&= -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(x^{(i)}-\mu)^2
\end{aligned} \tag{5}$$

3 Estimating the parameters which give the maximum likelihood

As a reminder, maximum likelihood for the Gaussian simply asks us to determine

$$\{\mu_{ML}, \sigma^2_{ML}\} = \arg\max_{\mu,\sigma^2} l(D|\mu, \sigma^2) \tag{6}$$

We know that the derivative of log-likelihood at the maximum will be zero. Hence, we can compute the parameters which maximizes the log-likelihood (and by extension the likelihood) by setting their partial derivatives to zero. Let's start with the mean, μ

3.1 Estimating μ_{ML}

$$\begin{aligned} \frac{\partial l(D|\mu, \sigma^2}{\partial \mu}) &= \frac{\partial}{\partial \mu}\left(-\frac{m}{2}\log(2\pi\sigma^2)\right) + \frac{\partial}{\partial \mu}\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x^{(i)} - \mu)^2\right) \\ &= 0 + -\frac{1}{\sigma^2}\sum_{i=1}^{m}(\mu - x^{(i)}) \end{aligned} \tag{7}$$

If we set $\frac{\partial l(D|\mu,\sigma^2}{\partial \mu} = 0$, then we'll have

$$\begin{aligned} -\frac{1}{\sigma^2}\sum_{i=1}^{m}(\mu - x^{(i)}) &= 0 \\ \sum_{i=1}^{m}(\mu - x^{(i)}) &= 0 \\ \sum_{i=1}^{m}\mu &= \sum_{i=1}^{m}x^{(i)} \\ m\mu &= \sum_{i=1}^{m}x^{(i)} \\ \mu_{ML} &= \frac{1}{m}\sum_{i=1}^{m}x^{(i)} \end{aligned} \tag{8}$$

And that gives us the ML estimate for the parameter μ.

3.2 Estimating σ^2_{ML}

Let's now derive the variance σ^2 too. In deriving the variance, we will assume that we already have the ML estimate of the mean μ_{ML}. Our partial derivative with respect to the variance will be of the form

$$\begin{aligned} \frac{\partial l(D|\mu, \sigma^2}{\partial \sigma^2}) &= \frac{\partial}{\partial \sigma^2}\left(-\frac{m}{2}\log(2\pi\sigma^2)\right) + \frac{\partial}{\partial \sigma^2}\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x^{(i)} - \mu_{ML})^2\right) \\ &= \frac{\partial}{\partial \sigma^2}(-\frac{m}{2}\log(2\pi)) + \frac{\partial}{\partial \sigma^2}(-\frac{m}{2}\log\sigma^2) + \frac{\partial}{\partial \sigma^2}\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x^{(i)} - \mu_{ML})^2\right) \\ &= 0 - \frac{m}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{m}(x^{(i)} - \mu_{ML})^2 \\ &= -\frac{1}{2\left(\sigma^2\right)^2}\left(\sigma^2 m - \sum_{i=1}^{m}(x^{(i)} - \mu_{ML})^2\right) \end{aligned} \tag{9}$$

If we set $\frac{\partial l(D|\mu,\sigma^2}{\partial \sigma^2})$ to, 0 then we can find the variance which maximizes the likelihood of the data. Thus

$$
\begin{aligned}
-\frac{1}{2\left(\sigma^2\right)^2}\left(\sigma^2 m-\sum_{i=1}^{m}(x^{(i)}-\mu_{ML})^2\right) &= 0 \\
\sigma^2 m-\sum_{i=1}^{m}(x^{(i)}-\mu_{ML})^2 &= 0 \qquad (10)\\
\sigma^2_{ML} &= \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu_{ML})^2
\end{aligned}
$$

Hence, we have computed in closed-form, the parameters which maximize the likelihood the data for the univariate Gaussian. The two estimates make sense right? If we can access the true parameters, then our next best shot is to use the sample parameters. That is exactly what maximum likelihood gave us. Can you guys derive the parameters for the multivariate Gaussian distribution using similar processes? Give it a try and the let me know how far you go.