MoHBF Analytical Sheet 3

Stephan Gabler

June 14, 2010

I cooperated with Duncan Blythe and Rafael Schultze-Kraft.

1 Exercise 1 (Linear Regression and LMSE)

The error function is given by $E = \sum_{i=1}^{N}(y_i - \alpha x - \beta)^2$. Differentiate E with respect to α and β to get:-

$$\frac{\partial E}{\partial \beta} = -2\sum_{i=1}^{N}(y_i - \alpha x_i - \beta) = 0 \tag{1}$$

Which is zero at a minimum.
so for a minimum w.r.t.beta:

$$\beta = \frac{1}{N}(\sum_{i=1}^{N} y_i - \alpha \sum_{i=1}^{N} x_i) \tag{2}$$

But also if we substitute this value for β into the error function and differentiate w.r.t. α, then we get:-

$$\begin{aligned}
\frac{\partial E}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\sum_{i=1}^{N}(y_i - \alpha x_i - \frac{1}{N}\sum_{j=1}^{N} y_j + \frac{\alpha}{N}\sum_{j=1}^{N} x_j)^2 \\
&= \frac{\partial}{\partial \alpha}\sum_{i=1}^{N}(y_i - \alpha x_i - \frac{1}{N}\sum_{j=1}^{N} y_j + \frac{\alpha}{N}\sum_{j=1}^{N} x_j) \times (-x_i + \frac{1}{N}\sum_{j=1}^{N} x_j) \\
&= 2\sum_{i=1}^{N}(-y_i x_i + \alpha {x_i}^2 + \frac{x_i}{N}\sum_{j=1}^{N} y_i - \frac{\alpha}{N} x_i \sum_{j=1}^{N} x_j) + \sum_{i=1}^{N}(y_i \frac{1}{N}\sum_{j=1}^{N} x_j - \frac{1}{N^2}\sum_{j=1}^{N} y_j \sum_{j=1}^{N} x_j + \frac{\alpha}{N^2}(\sum_{j=1}^{N} x_j)^2)
\end{aligned}$$

But this last expression is zero whenever

$$\begin{gathered}
\alpha(\sum_{i=1}^{N} {x_i}^2 - \frac{1}{N}(\sum_{i=1}^{N} x_i)^2 - \frac{1}{N}(\sum_{i=1}^{N} x_i)^2 + \frac{N}{N^2}(\sum_{i=1}^{N} x_i)^2) \\
= \sum_{i=1}^{N} y_i x_i - \frac{1}{N}\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N} y_i \sum_{i=1}^{N} x_i + \frac{N}{N^2}\sum_{i=1}^{N} y_i \sum_{i=1}^{N} x_i
\end{gathered}$$

And this is the result we wanted:-

$$\alpha = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2} \tag{3}$$

2 Exercise 2 (LMSE and Maximum Likelihood)

All data-points $y_1 \cdots y_i$ are a assumed to be gaussian distributed around the true model $f(\vec{x}, \vec{\alpha})$. The standard distributions σ are the same for all all data points. The joint density for all data-points is then:

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i-\mu)^2}{2\sigma^2}} \tag{4}$$

with μ being equal to the value of the true model $f(\vec{x}, \vec{\alpha})$. The negative log-likelihood of the model is

$$L(\vec{\alpha}, \vec{x}) = \sum_{i=1}^{N} ln(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i-\mu)^2}{2\sigma^2}}) \tag{5}$$

which we then want to maximize.

$$\begin{aligned} \max_{\vec{\alpha}} L(\vec{\alpha}, \vec{x}) &= \sum_{i=1}^{N} ln(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i-f(\vec{x},\vec{\alpha}))^2}{2\sigma^2}}) \\ &= -n\, ln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \sum_{i=1}^{N} ln(e^{\frac{-(y_i-f(\vec{x},\vec{\alpha}))^2}{2\sigma^2}}) \\ &= \underbrace{-\frac{n}{2\sigma^2} ln(\frac{1}{\sqrt{2\pi\sigma^2}})}_{constant} + \sum_{i=1}^{N} (y_i - f(\vec{x}, \vec{\alpha}))^2 \end{aligned} \tag{6}$$

Because we can neglect the constant term, only the last part has to be maximized which is the same as the LMSE.

$$\sum_{i=1}^{N} (y_i - f(\vec{x}, \vec{\alpha}))^2 \tag{7}$$

I think that for both behavioral and psychophysical experiments assuming σ to be constant for all $\vec{x}$ might not be the best assumption. For example could the std of a effect increase with time because the subjects get tired.

3 Exercise 3 (Maximum Likelihood and Binomial Data)

The binomial distribution has the form $Pr(K = k) = \binom{N}{k} p^k (1-p)^{N-k}$. Here we set N to the no. of trials, K to the number of successes on the part of the observer. and p to the actual probability of detection. Then our likelihood function is $\binom{N}{k} \tilde{p}^k (1-\tilde{p})^{N-k}$. Differentiating this w.r.t $\tilde{p}$ gives:- $\binom{N}{k} k\tilde{p}^{k-1}(1-\tilde{p})^{N-k} - \binom{N}{k}(N-k)\tilde{p}^k(1-\tilde{p})^{N-k-1}$. This is zero if and only if:-

$$k\tilde{p}^{k-1}(1-\tilde{p})^{N-k} = (N-k)\tilde{p}^k(1-\tilde{p})^{N-k-1} \tag{8}$$

$$k(1-\tilde{p}) = (N-k)\tilde{p} \tag{9}$$

$$k = N\tilde{p} \tag{10}$$

$$\tilde{p} = \frac{k}{N} \tag{11}$$

Now try the log likelihood, which is given by:- $log(\binom{N}{k}) + log(\tilde{p}^k) + log((1-\tilde{p})^{N-k})$. Differentiating this w.r.t. $\tilde{p}$ and setting to zero gives:-

$$\frac{k\tilde{p}^{k-1}}{(\tilde{p}^k)} = \frac{(N-k)(1-\tilde{p})^{N-k-1}}{(1-\tilde{p})^{N-k}} \tag{12}$$

This clearly gives the same result as before.

4 Exercise 4 (Binomial Data and Adaptive Procedures)

First we need to define what we mean here by convergence since we know that at the least after every 3 presentations our stimulus intensity will be changed by a discrete amount according to how the 3 down 1 up procedure is defined, so the sequence of stimulus values themselves don't converge in a mathematical sense.

We could do this in terms of average of sum of the preceding terms. So we might say (for example) that the sequence of stimulus presentations converges to a if for any $\epsilon > 0$ there exists a time point T such that if $t > T$ $|\frac{1}{M}\sum_{i=1}^{t} s(i) - a| < \epsilon$.

But this won't work since it's conceivable that these terms s(i) could alternate between tow distinct values, and the *convergence* point will lie somewhere in between.