CS420, Tutorial 5. SSSE and Statistical efficiency (derivations).

-ffm, with references from A+'s notes

October 13, 2011

1 SSSE

1.1 Derivatives over the parameters

The Sum of Standard Square Errors function (S^3E) parameterized by $\vec{q}$, for an affine model is:

$$S^3E(\vec{q}) = \frac{1}{2}\sum_k (z_k - A_k\vec{q})^\top \Sigma_k^{-1} (z_k - A_k\vec{q}), \tag{1}$$

To minimize we need:

$$\frac{\partial}{\partial\vec{q}} S^3E(\vec{q}) = 0. \tag{2}$$

More explicitly:

$$0 = \frac{\partial}{\partial q_i} S^3E(\vec{q}) = \frac{1}{2}\sum_k \left[\frac{\partial}{\partial q_i}(z_k - A_k\vec{q})\right]^\top \Sigma_k^{-1} (z_k - A_k\vec{q}) + (z_k - A_k\vec{q})^\top \Sigma_k^{-1} \left[\frac{\partial}{\partial q_i}(z_k - A_k\vec{q})\right]. \tag{3}$$

Since we know that

$$\frac{\partial}{\partial q_i}(z_k - A_k\vec{q}) = -A_k \frac{\partial}{\partial q_i}\vec{q} = -A_k e_i, \tag{4}$$

with $e_i = [0, 0, \ldots, 1, \ldots, 0, 0]^\top$ with the one at the i-th position. Then we can substitute back into Equation 3 to get:

$$\frac{\partial}{\partial q_i} S^3E(\vec{q}) = \frac{1}{2}\sum_k \left[(-A_k e_i)^\top \Sigma_k^{-1} (z_k - A_k\vec{q}) + (z_k - A_k\vec{q})^\top \Sigma_k^{-1} (-A_k e_i)\right]. \tag{5}$$

The sizes of these matrices are:

$$(-A_k e_i)^\top_{[2\times 1]^\top} (\Sigma_k^{-1})_{[2\times 2]} (z_k - A_k\vec{q})_{[2\times 1]} = p_{[1\times 1]},$$

and for the other half:

$$(z_k - A_k\vec{q})^\top_{[2\times 1]^\top} (\Sigma_k^{-1})_{[2\times 2]} (-A_k e_i)_{[2\times 1]} = p_{[1\times 1]},$$

which are both the same constants. So we can re-write (and re-arrange) as:

$$\frac{\partial}{\partial q_i} S^3E(\vec{q}) = -\sum_k e_i^\top \left[A_k^\top \Sigma_k^{-1}(z_k - A_k\vec{q})\right]. \tag{6}$$

The i-th component of the partial derivative with respect to $\vec{q}$ is the i-th component of

$$-\sum_k \left[A_k^\top \Sigma_k^{-1}(z_k - A_k\vec{q})\right]. \tag{7}$$

Or in a more general way, we have written the entire (column) vector of partial derivatives as:

$$\frac{\partial}{\partial\vec{q}} S^3E(\vec{q}) = -\sum_k \left[A_k^\top \Sigma_k^{-1}(z_k - A_k\vec{q})\right] \tag{8}$$

■

2 Statistical efficiency

2.1 Log-likelihood of $p(\vec{z}|m)$

The log likelihood, $log(p(z_1, ..., z_K|m))$, is

$$
\begin{aligned}
log(p(\vec{z}_1, \ldots, \vec{z}_K)|\vec{m}) &= log\left[\prod_{k=1}^{K} p(\vec{z}_k - m|0, \Sigma_k)\right] \\
&= \sum_{k=1}^{K} log\left[p(\vec{z}_k - m|0, \Sigma_k)\right] \\
&= \sum_{k=1}^{K} log\left[\frac{1}{(2\pi|\Sigma_k|)} e^{-\frac{1}{2}(z_k - m)^\top \Sigma_k^{-1}(z_k - m)}\right] \\
&= -\sum_{k=1}^{K}\left[\frac{1}{2}(z_k - m)^\top \Sigma_k^{-1}(z_k - m) - log(2\pi|\Sigma_k|)\right] \\
&= -\frac{1}{2}\sum_{k=1}^{K}\left[(z_k - m)^\top \Sigma_k^{-1}(z_k - m)\right] + [\text{Constant independent of } \vec{m}]
\end{aligned}
$$

So, minimizing the above (to obtain maximum likelihood estimates for $\vec{m}$) is equivalent to minimize the sum of the standard squared errors:

$$SSSE(m) = -\frac{1}{2}\sum_{k=1 \ldots K}(z_k - m)^\top \Sigma_k^{-1}(z_k - m)$$

■

2.2 Normal equations

To find maximum likelihood estimates of the mean $\vec{m}$, we use the partial derivatives of S^3E and make them equal to zero (algorithm from the Parameter Estimation lecture notes, page 20.):

$$\frac{\partial S^3E}{\partial \vec{q}}(\vec{q^\star}) = 0$$

Similarly to Equation 8 (except in this case A is the identity), one gets:

$$
\begin{aligned}
0 = \frac{\partial}{\partial \vec{q}} S^3E(\vec{q}) &= -\sum_k\left[\Sigma_k^{-1}(z_k - \vec{m})\right] \\
0 &= -\sum_k\left[\Sigma_k^{-1} z_k\right] + \sum_k\left[\Sigma_k^{-1}\vec{m}\right] \\
\sum_k \Sigma_k^{-1}\vec{m} &= \sum_k \Sigma_k^{-1} z_k \\
\left[\sum_k \Sigma_k^{-1}\right]\vec{m} &= \sum_k \Sigma_k^{-1} z_k \\
F\vec{m} &= r
\end{aligned}
$$

■

with $F = \sum_k \Sigma_k^{-1}$ and $r = \sum_k \Sigma_k^{-1} z_k$. The matrix F is the Fisher information matrix.