Model Selection is a Linear Function of λ*

Jason D. M. Rennie
jrennie@gmail.com

February 27, 2006

Abstract

We show that the model score is a linear function of λ.

Let $X \in \mathbb{R}^{n \times d}$. Let $A_i \in \mathbb{R}^{a_i \times d}$ with $\sum_i a_i = n$ and $B_j \in \mathbb{R}^{b_j \times d}$ with $\sum_j b_j = n$. We define two partitionings: $X = \begin{bmatrix} A_1 \\ \vdots \\ A_l \end{bmatrix} = \begin{bmatrix} B_1 \\ \vdots \\ B_m \end{bmatrix}$.

Consider trace norm distribution models corresponding to these partitionings:

$$P(A_1, \ldots, A_l) = \prod_i P(A_i) = \prod_i \frac{1}{Z_{A_i}} \exp(-\lambda \|A_i\|_\Sigma), \quad \text{and} \tag{1}$$

$$P(B_1, \ldots, B_m) = \prod_j P(B_j) = \prod_j \frac{1}{Z_{B_j}} \exp(-\lambda \|B_j\|_\Sigma). \tag{2}$$

Then, we prefer model A if

$$f_{A,B}(\lambda) = \sum_j \lambda \|B_j\|_\Sigma + \sum_j \log Z_{B_j} - \sum_i \lambda \|A_i\|_\Sigma - \sum_i \log Z_{A_i} > 0 \tag{3}$$

Definte a trace norm distribution on X, $P(X)$. Note that

$$Z = \int \exp(-\lambda \|X\|_\Sigma) = \frac{1}{\lambda^{nd}} \int \exp(-\|X\|_\Sigma). \tag{4}$$

Define $Z'_{A_i} \equiv \int \exp(-\|A_i\|_\Sigma) = \lambda^{a_i d} Z_{A_i}$. Similarly, $Z'_{B_j} = \lambda^{b_j d} Z_{B_j}$. Then,

$$f_{A,B}(\lambda) = \lambda \left(\sum_j \|B_j\|_\Sigma - \sum_i \|A_i\|_\Sigma \right) + \sum_j \log Z'_{B_j} - \sum_i \log Z'_{A_i}. \tag{5}$$

The model score is a linear function of λ. Hence, for each model, there is a contiguous range of λ for which that model is preferred over all others. Note that the range may be empty—i.e. a model may never be preferred.

*Joint work with John Barnett and Tommi Jaakkola.