Maximum Likelihood for Polytomous Rasch Model for Multi-Scale Speech Quality Judgment

Iman

Tuesday, November 4, 2008 12:12

1 Introduction

In this section we compute the first and the second derivatives of the likelihood function

$$S = \sum_{i=1}^{N} \log(P(d_i)), \tag{1}$$

where N is the total number of data points. The first derivatives are

$$\begin{aligned}
\frac{\partial S}{\partial \omega_k} &= \sum_{i=1}^{N} 1/P(d_i) \frac{\partial P(d_i)}{\partial \omega_k} \\
\frac{\partial S}{\partial \tau_m} &= \sum_{i=1}^{N} 1/P(d_i) \frac{\partial P(d_i)}{\partial \tau_m} \\
\frac{\partial S}{\partial \alpha_m} &= \sum_{i=1}^{N} 1/P(d_i) \frac{\partial P(d_i)}{\partial \alpha_m}
\end{aligned} \tag{2}$$

where $k = 1, \cdots, p$ for p features, and $m = 1, \cdots, M-1$ for M thresholds. The second derivatives are

$$
\begin{aligned}
\frac{\partial^2 S}{\partial \omega_i \partial w_j} &= \sum_{i=1}^{N} \frac{-1}{P^2(d_i)} \cdot \frac{\partial P(d_i)}{\partial \omega_i} \cdot \frac{\partial P(d_i)}{\partial \omega_j} + \frac{1}{P(d_i)} \cdot \frac{\partial^2 P(d_i)}{\partial \omega_i \partial \omega_j} \\
\frac{\partial^2 S}{\partial \tau_l \partial \tau_m} &= \sum_{i=1}^{N} \frac{-1}{P^2(d_i)} \cdot \frac{\partial P(d_i)}{\partial \tau_l} \cdot \frac{\partial P(d_i)}{\partial \tau_m} + \frac{1}{P(d_i)} \cdot \frac{\partial^2 P(d_i)}{\partial \tau_m \partial \tau_l} \\
\frac{\partial^2 S}{\partial \alpha_l \partial \alpha_m} &= \sum_{i=1}^{N} \frac{-1}{P^2(d_i)} \cdot \frac{\partial P(d_i)}{\partial \alpha_l} \cdot \frac{\partial P(d_i)}{\partial \alpha_m} + \frac{1}{P(d_i)} \cdot \frac{\partial^2 P(d_i)}{\partial \alpha_l \partial \alpha_m} \qquad (3)
\end{aligned}
$$

Before proceeding to compute the above derivatives, we rephrase the probability distribution as

$$
\begin{aligned}
P_{i,n} &= \frac{\exp(\beta_{i,n})}{1 + \sum_{l=2}^{M} \exp(\beta_{i,l})}, \\
P_{i,n} &= P(d_i = n) \\
\beta_{i,n} &= \sum_{k=1}^{n-1} \alpha_k (\Delta_i - \tau_k) \qquad (4)
\end{aligned}
$$

1.1 First order derivatives

$$
\begin{aligned}
\frac{\partial P_{i,n}}{\partial \Delta_i} &= \sigma_{i,n} P_{i,n} \\
\sigma_{i,n} &= \gamma_n - \sum_{k=2}^{M} \gamma_k P_{i,k} \\
\gamma_n &= \sum_{k=1}^{n-1} \alpha_k \qquad (5)
\end{aligned}
$$

$$
\frac{\partial P_{i,n}}{\partial \omega_j} = \phi_{i,j} \frac{\partial P_{i,n}}{\partial \Delta_i} \qquad (6)
$$

where $\phi_{i,j}$ denotes the value of j-th feature in the i-th data point.

$$
\begin{aligned}
\frac{\partial P_{i,n}}{\partial \tau_m} &= -\alpha_m \sum_{k=m+1}^{M} \frac{\partial P_{i,n}}{\partial \beta_{i,k}} \\
\frac{\partial P_{i,n}}{\partial \alpha_m} &= (\Delta_i - \tau_m) \sum_{k=m+1}^{M} \frac{\partial P_{i,n}}{\partial \beta_{i,k}}
\end{aligned} \tag{7}
$$

$$
\frac{\partial P_{i,n}}{\partial \beta_{i,k}} = \begin{cases} -P_{i,n} \cdot P_{i,k} & k \neq n; \\ P_{i,n} - P_{i,n}^2 & k = n. \end{cases} \tag{8}
$$

1.2 Second Order Derivatives

These are used to compute the hessian and the covariance matrix for Laplace approximations. There exist at most six forms of differentiation namely $\frac{\partial^2 P_{i,n}}{\partial \omega_{j_1} \partial \omega_{j_2}}$, $\frac{\partial^2 P_{i,n}}{\partial \omega_j \partial \tau_m}$, $\frac{\partial^2 P_{i,n}}{\partial \omega_j \partial \alpha_m}$, $\frac{\partial^2 P_{i,n}}{\partial \tau_m \partial \tau_l}$, $\frac{\partial^2 P_{i,n}}{\partial \alpha_m \partial \alpha_l}$, and $\frac{\partial^2 P_{i,n}}{\partial \tau_m \partial \alpha_l}$. Let us first derive how to compute these based on $\frac{\partial^2 P_{i,n}}{\partial \beta_{i,m} \partial \beta_{i,l}}$, $\frac{\partial^2 P_{i,n}}{\partial \Delta_i \partial \beta_{i,m}}$, and $\frac{\partial^2 P_{i,n}}{\partial \Delta_i^2}$.

$$
\begin{aligned}
\frac{\partial^2 P_{i,n}}{\partial \omega_{j_1} \partial \omega_{j_2}} &= \phi_{i,j_1} \cdot \phi_{i,j_2} \cdot \frac{\partial^2 P_{i,n}}{\partial \Delta_i^2} \\
\frac{\partial^2 P_{i,n}}{\partial \omega_j \partial \tau_m} &= -\phi_{i,j} \cdot \alpha_m \cdot \sum_{k=m+1}^{M} \frac{\partial^2 P_{i,n}}{\partial \Delta_i \partial \beta_{i,m}} \\
\frac{\partial^2 P_{i,n}}{\partial \omega_j \partial \alpha_m} &= \phi_{i,j} \cdot (\Delta_i - \tau_m) \cdot \sum_{k=m+1}^{M} \frac{\partial^2 P_{i,n}}{\partial \Delta_i \partial \beta_{i,k}} \\
\frac{\partial^2 P_{i,n}}{\partial \tau_m \partial \tau_l} &= \alpha_m \cdot \alpha_l \cdot \sum_{k_1=m+1}^{M} \sum_{k_2=l+1}^{M} \frac{\partial^2 P_{i,n}}{\partial \beta_{i,k_1} \partial \beta_{i,k_2}} \\
\frac{\partial^2 P_{i,n}}{\partial \alpha_m \partial \alpha_l} &= (\Delta_i - \tau_m) \cdot (\Delta_i - \tau_l) \cdot \sum_{k_1=m+1}^{M} \sum_{k_2=l+1}^{M} \frac{\partial^2 P_{i,n}}{\partial \beta_{i,k_1} \partial \beta_{i,k_2}} \\
\frac{\partial^2 P_{i,n}}{\partial \tau_m \partial \alpha_l} &= -\alpha_m \cdot (\Delta_i - \tau_l) \cdot \sum_{k_1=m+1}^{M} \sum_{k_2=l+1}^{M} \frac{\partial^2 P_{i,n}}{\partial \beta_{i,k_1} \partial \beta_{i,k_2}}
\end{aligned} \tag{9}
$$

Now, let us show how to compute $\frac{\partial^2 P_{i,n}}{\partial \Delta_i^2}$, $\frac{\partial^2 P_{i,n}}{\partial \Delta_i \partial \beta_{i,m}}$, and $\frac{\partial^2 P_{i,n}}{\partial \beta_{i,m} \partial \beta_{i,l}}$.

$$
\begin{aligned}
\frac{\partial^2 P_{i,n}}{\partial \Delta_i^2} &= P_{i,n}(\sigma_n^2 - \sum_{k=2}^{M} \sigma_k \cdot \gamma_k \cdot P_{i,k}) \\
\frac{\partial^2 P_{i,n}}{\partial \Delta_i \partial \beta_{i,m}} &= \begin{cases} -P_{i,n} \cdot P_{i,m} \cdot (\sigma_n + \sigma_m) & m \neq n; \\ P_{i,n} \cdot \sigma_n - 2P_{i,n}^2 \sigma_n & m = n; \end{cases} \\
\frac{\partial^2 P_{i,n}}{\partial \beta_{i,m} \partial \beta_{i,l}} &= \begin{cases} -P_{i,n} \cdot P_{i,m} + 2P_{i,n}^2 \cdot P_{i,m} & m \neq l = n; \\ -P_{i,n} \cdot P_{i,m} + 2P_{i,n} \cdot P_{i,m}^2 & m = l \neq n; \\ P_{i,n} - 3P_{i,n}^2 + 2P_{i,n}^3 & m = l = n \\ 2P_{i,n} \cdot P_{i,m} \cdot P_{i,l} & m \neq l \neq n; \end{cases}
\end{aligned}
\tag{10}
$$