Correction to the derivation of the LS classifier

Thomas Schön

January 30, 2013

Given $\{t_n, x_n\}_{n=1}^N$ as training data, the likelihood is

$$p(t_{1:N}; \widetilde{w}, \widetilde{x}_{1:N}) = \prod_{n=1}^{N} \mathcal{N}(t_n; \widetilde{w}^T \widetilde{x}_n, I). \tag{1}$$

Rather than maximizing the likelihood we will consider the equivalent problem of minimizing the negative log-likelihood, where

$$\begin{aligned} -\log p(t_{1:N}; \widetilde{w}, \widetilde{x}_{1:N}) &\propto \frac{1}{2} \sum_{n=1}^{N} (t_n - \widetilde{w}_n^T \widetilde{x}_n)^T (t_n - \widetilde{w}_n^T \widetilde{x}_n) \\ &= \frac{1}{2} \sum_{n=1}^{N} (t_n^T - \widetilde{x}_n^T \widetilde{w}_n)(t_n^T - \widetilde{x}_n^T \widetilde{w}_n)^T. \end{aligned} \tag{2}$$

Introducing the matrices

$$T = \begin{pmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{pmatrix} \in \mathbb{R}^{N \times K}, \qquad \widetilde{X} = \begin{pmatrix} \widetilde{x}_1^T \\ \widetilde{x}_2^T \\ \vdots \\ \widetilde{x}_N^T \end{pmatrix} \in \mathbb{R}^{N \times (D+1)} \tag{3}$$

we can now write (2) according to

$$\begin{aligned} -\log p(t_{1:N}; \widetilde{w}, \widetilde{x}_{1:N}) &\propto \frac{1}{2} \mathbf{Tr} (T - \widetilde{X}\widetilde{W})(T - \widetilde{X}\widetilde{W})^T \\ &= \frac{1}{2} \mathrm{Tr} (T - \widetilde{X}\widetilde{W})^T (T - \widetilde{X}\widetilde{W}), \end{aligned} \tag{4}$$

where we made use of the fact $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ in order to establish the last equality.