1 Partial least squares model

The partial least squares model is presented as a latent variable model in Kevin Murphy's 2012 book Machine Learning: A Probabilistic Perspective.

$$
\begin{aligned}
p(z) &= \mathcal{N}(z|0, I) && (1)\\
p(v|z, W, \mu, \sigma) &= \mathcal{N}(v|Wz + \mu, \sigma^2 I) && (2)
\end{aligned}
$$

where

$$
\begin{aligned}
W &= \begin{pmatrix} W_y & 0 \\ W_x & B_x \end{pmatrix} && (3)\\
z &= (z^s; z^x) && (4)\\
v &= (y; x) && (5)\\
\mu &= (\mu_y; \mu_x). && (6)
\end{aligned}
$$

Marginalising z gives

$$
\begin{aligned}
p(v|W, \mu, \sigma) &= \int \mathcal{N}(v|Wz + \mu, \sigma^2 I)\mathcal{N}(z|0, I)\, dz && (7)\\
&= \mathcal{N}(v|\mu, WW^T + \sigma^2 I) && (8)
\end{aligned}
$$

Conditioning on x gives

$$
p(y|x) = \mathcal{N}(y|m_{y|x}, S_{y|x}) \quad (9)
$$

where

$$
\begin{aligned}
C &= (B_x B_x^T + W_x W_x^T + \sigma^2 I)^{-1} && (10)\\
m_{y|x} &= \mu_y + W_y W_x^T C(x - \mu_x) && (11)\\
S_{y|x} &= \sigma^2 I + W_y W_y^T - W_y W_x^T C W_x W_y^T && (12)
\end{aligned}
$$

Suppose we now obtain N independent observations from the model

$$
v_n = (y_n; x_n), \quad 1 \le n \le N. \quad (13)
$$

We wish to estimate W, μ and σ.