# Exercise Sheet 3

Machine Learning 2, SS16

May 13, 2016

Sascha Huk, 321249; Viktor Jeney, 348969; Mario Tambos, ??????; Jan Tinapp, 0380549

## Exercise 1

(a) For data-span constructed $w_x = X\alpha_x$ and $w_y = Y\alpha_y$ the primal problem is:

$$\begin{aligned}
&\max_{\alpha_x,\alpha_y} {\alpha_x}^\top X^\top C_{xy} Y \alpha_y \\
\text{s.t. } &{\alpha_x}^\top X^\top C_{xx} X \alpha_x - 1 = 0 \\
&{\alpha_y}^\top Y^\top C_{yy} Y \alpha_y - 1 = 0
\end{aligned}$$

Lagrangian (the factor 1/2 is introduced just for convenience):

$$\mathcal{L} = \alpha_x^\top X^\top C_{xy} Y \alpha_y - \frac{1}{2}\lambda_x(\alpha_x^\top X^\top C_{xx} X\alpha_x - 1) - \frac{1}{2}\lambda_y(\alpha_y^\top Y^\top C_{yy} Y \alpha_y - 1)$$

Partial derivatives:

$$\frac{\partial \mathcal{L}}{\partial \alpha_x^\top} = X^\top C_{xy} Y \alpha_y - \lambda_x X^\top C_{xx} X \alpha_x \stackrel{!}{=} 0 \qquad , \qquad \frac{\partial \mathcal{L}}{\partial \alpha_y^\top} = Y^\top C_{yx} X \alpha_x - \lambda_y Y^\top C_{yy} Y \alpha_y \stackrel{!}{=} 0$$

We now multiply with $\alpha_x^\top$, $\alpha_y^\top$

$$\begin{aligned}
\alpha_x^\top X^\top C_{xy} Y \alpha_y &= \lambda_x \alpha_x^\top X^\top C_{xx} X \alpha_x \qquad , \qquad \alpha_y^\top Y^\top C_{yx} X \alpha_x = \lambda_y \alpha_y^\top Y^\top C_{yy} Y \alpha_y \\
\implies \alpha_x^\top X^\top C_{xy} Y \alpha_y &= \lambda_x \alpha_x^\top X^\top C_{xx} X \alpha_x \qquad , \qquad \alpha_x^\top X^\top C_{xy} Y \alpha_y = \lambda_y \alpha_y^\top Y^\top C_{yy} Y \alpha_y
\end{aligned}$$

From the auto-cov constraints follows

$$\alpha_x^\top X^\top C_{xy} Y \alpha_y = \lambda_x \underbrace{\alpha_x^\top X^\top C_{xx} X \alpha_x}_{=1} = \lambda_y \underbrace{\alpha_y^\top Y^\top C_{yy} Y \alpha_y}_{=1} \quad \implies \quad \lambda_x = \lambda_y$$

Now the derivatives can be rewritten as follows:

$$X^\top C_{xy} Y \alpha_y \stackrel{!}{=} \lambda_x X^\top C_{xx} X \alpha_x \qquad , \qquad Y^\top C_{yx} X \alpha_x \stackrel{!}{=} \lambda_x Y^\top C_{yy} Y \alpha_y$$

(b+c) The same in blockmatrix form:

$$\begin{aligned}
&\begin{bmatrix} 0 & X^\top C_{xy} Y \\ Y^\top C_{yx} X & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} \stackrel{!}{=} \lambda_x \begin{bmatrix} X^\top C_{xx} X & 0 \\ 0 & Y^\top C_{yy} Y \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} \\
\implies &\begin{bmatrix} X^\top C_{xx} X & 0 \\ 0 & Y^\top C_{yy} Y \end{bmatrix}^{-1} \begin{bmatrix} 0 & X^\top C_{xy} Y \\ Y^\top C_{yx} X & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} \stackrel{!}{=} \lambda_x I \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}
\end{aligned}$$

By finding the solutions ${\alpha_x}^*$ and ${\alpha_y}^*$, the dual variable $\lambda_x$ is identified as this is an eigenvalue problem. Each eigenvalue $\lambda_x$ corresponds to an eigenvector $[\alpha_x, \alpha_y]^\top$. Hence the lagrangian does not depend on $\lambda_x$ which means $\forall \lambda_x . \mathcal{L}({\alpha_x}^*, {\alpha_y}^*, \lambda_x) = \mathcal{L}({\alpha_x}^*, {\alpha_y}^*)$. For the dual problem we therefore find:

$$\min_{\lambda_x} \max_{\alpha_x,\alpha_y} \mathcal{L}(\alpha_x, \alpha_y, \lambda_x) = \min_{\lambda_x} \mathcal{L}({\alpha_x}^*, {\alpha_y}^*, \lambda_x) = \mathcal{L}({\alpha_x}^*, {\alpha_y}^*)$$