ML2 Sheet 2

1 Exercise 3

1.1 Exercise 3 (a)

The two *min* functions are going to return a large negative value if either $\omega_x^\top C_{xx}\omega_x$ or $\omega_y^\top C_{xx}\omega_y$ or both are large. Therefore, they will subtract away much of the impact of the first term $\omega_x^\top C_{xy}\omega_y$. This means that, dependent on the size of the positive scalar α, the last two terms will impose a "penalty" whenever a large ω_x or ω_y is used, which will mitigate the issue where the maximization is caused by any such large vector, similarly to the constrained CCA.

1.2 Exercise 3 (b)

$$\frac{\partial\phi_x}{\partial\theta_x} = \begin{pmatrix} \frac{\partial\phi_{x_1}}{\partial\theta_{x_1}} & \dots & \frac{\partial\phi_{x_n}}{\partial\theta_{x_1}} \\ & \ddots & \vdots \\ \frac{\partial\phi_{x_{h_1}}}{\partial\theta_{x_1}} & \dots & \frac{\partial\phi_{x_{h_1}}}{\partial\theta_{x_n}} \end{pmatrix}$$

1. a

$w_x \in \mathbb{R}^{d_1} \quad X \in \mathbb{R}^{d_1 \times N} \quad \alpha_x \in \mathbb{R}^N$

We can represent $w_x = X \cdot \alpha_x$ as a system of equations

$$\underbrace{\left.\begin{cases} X_{11} \cdot \alpha_1 + X_{12} \cdot \alpha_2 + \dots + X_{1N} \cdot \alpha_N = w_{x_1} \\ X_{21} \cdot \alpha_1 + X_{22} \cdot \alpha_2 + \dots + X_{2N} \cdot \alpha_N = w_{x_2} \\ X_{31} \cdot \alpha_1 + X_{32} \cdot \alpha_2 + \dots + X_{3N} \cdot \alpha_N = w_{x_3} \\ \vdots \qquad\qquad\qquad\qquad \vdots \end{cases}\right\} d_1}_{N}$$

Having N variables and d_1 equations would normally mean that no solutions exist because $N > d_1$, but considering the condition $N << d_i$ system becomes determined (even overdeterm.), meaning there exist a solution that lies in the span of the data.

1.b

$$\begin{bmatrix} 0 & AB \\ BA & 0 \end{bmatrix} \cdot \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} A^2 & 0 \\ 0 & B^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} \quad (1)$$

We can rewrite (1) as

$$AB\,\alpha_y = \rho A^2 \alpha_x$$
$$BA\,\alpha_x = \rho B^2 \alpha_y$$

By setting $w_x = X\alpha_x$, $w_y = Y\alpha_y$ we obtain

$$X^TX\,Y^TY\,\alpha_y = \rho X^TX\,X^TX\,\alpha_x$$
$$Y^TY\,X^TX\,\alpha_x = \rho Y^TY\,Y^TY\,\alpha_y$$

$$X^TC_{XY}w_y = \rho X^TC_{XX}\alpha_x$$
$$Y^TC_{YX}w_x = \rho Y^TC_{yy}\alpha_y$$

3.1

After dividing by X^T and Y^T we get

$$C_{xy}w_y = \rho C_{xx}\alpha_x$$
$$C_{yx}w_x = \rho C_{yy}\alpha_y$$

which is the original eigenvalue problem.

3.2

2. a

Original ~~eigenvalue~~ CCA problem

.1 Reformulate the problem in terms of dot products

First, set $w_x = X \alpha_x$ - Solution will lie in the
$w_y = Y \alpha_y$ space spanned by the data not number of dimensions

Lagrangean looks like that:

$$\mathcal{L} = \alpha_x^T X^T X Y^T Y \alpha_y - \lambda_1 (\alpha_x^T X^T X X^T X \alpha_x - 1) -$$

$$- \lambda_2 (\alpha_y^T Y^T Y Y^T Y \alpha_y - 1)$$

.2 Replace all dot products with <u>valid</u> kernels

$K_x = \phi_1(X) \phi_1(X)^T$, $K_y = \phi_2(Y) \phi_2(Y)^T$ where $\phi_1 : x \mapsto \phi_1(x)$
$\phi_2 : y \mapsto \phi_1(y)$

$$\mathcal{L} = \alpha_x^T K_x K_y \alpha_y - \lambda_1 (\alpha_x K_x^2 \alpha_x - 1) - \lambda_2 (\alpha_y^T K_y^2 \alpha_y - 1)$$

Original eigenvalue problem.

The course of action is the same as above

.1 $C_{\cancel{xy}} w_y = \lambda_1 C_{xx} w_x \rightarrow X Y^T Y \alpha_y = \lambda_1 X X^T X \alpha_x$ (1)

$C_{\cancel{yx}} w_x = \lambda_1 C_{yy} w_y \qquad Y X^T X \alpha_y = \lambda_1 \cancel{X} Y Y^T Y \alpha_y$ (2)

Multiplying both sides by X^T for (1) and Y^T for (2) we get

$$X^T X Y^T Y \alpha_y = \lambda_1 X^T X X^T X \alpha_x$$

$$Y^T Y X^T X \alpha_x = \lambda_1 Y^T Y Y^T Y \alpha_y$$

.2 Replace dot products with kernels

$$K_x K_y \alpha_y = \lambda_1 K_x^2 \alpha_x$$

$$K_y K_x \alpha_x = \lambda_1 K_y^2 \alpha_y$$

2.b kCCA computes new coordinate system that is optimal in maximizing correlation between two sets of variables. Each axis is nonlinear combination of variables in corresponding set of variables.

Index of comments

3.1 Likely not invertible

3.2 What is the original eigenvalue problem?