~~Virtualisation~~

Exercise 1

a) $w_x = X a_x$, $w_y = Y a_y$

CCA objective.

$$\underset{w_x, w_y}{\text{argmax}} \left(w_x^T X Y^T w_y \right) \quad \text{s.t} \quad \begin{array}{l} w_x^T X X^T w_x = 1 \\ w_y^T Y Y^T w_y = 1 \end{array}$$

Lagrangian

$$L = w_x^T C_{xy} w_y - \frac{1}{2} \alpha \left(w_x^T C_{xx} w_x - 1 \right) - \frac{1}{2} \beta \left(w_y^T C_{yy} w_y - 1 \right)$$

$$= (X a_x)^T C_{xy} Y a_y - \frac{1}{2} \left(\alpha (X a_x)^T C_{xx} (X a_x) - 1 \right) - \frac{1}{2} \beta \left(Y a_y C_{yy} Y a_y - 1 \right)$$

$$\frac{\partial L}{\partial a_x} = X^T C_{xy} Y a_y - \alpha X^T C_{xx} X a_x = 0$$

$$\frac{\partial L}{\partial a_y} = a_x^T X^T C_{xy} Y - \beta Y^T C_{yy} Y a_y = 0.$$

$$X^T C_{xy} Y a_y = \alpha X^T C_{xx} X a_x \qquad \times a_x^T \dots (1)$$

$$a_x^T X^T C_{YY} Y = \beta Y^T C_{yy} Y a_y \qquad a_y^T \dots (2)$$

From (1).

$$a_x^T X^T C_{xy} Y a_y = \alpha a_x^T X^T C_{xx} X a_x.$$

$$(w_x)^T C_{xx} w_y = \alpha w_x^T C_{xx} w_x$$

From (2)

$$a_x^T X^T C_{yy} Y a_y = \beta a_y^T Y^T C_{yy} Y a_y.$$

$$(w_x)^T C_{yy} w_y = \beta (w_y)^T C_{yy} w_y$$

From auto-covariance matrix

$$w_x^T C_{xx} w_x = w_y^T C_{yy} w_y = 1.$$

follows

$$\alpha = \beta.$$

1.b) CCA Objective

$$w_x^T X Y^T w_y = 1 \qquad \text{s.t} \qquad w_x^T X X^T w_x = 1 \Rightarrow a_x^T X^T X X^T X a_x = 1$$
$$\Rightarrow a_x^T A^2 a_x = 1$$
$$w_y^T Y Y^T w_y = 1$$
$$\Rightarrow a_y^T Y^T Y Y^T Y a_y = 1 \Rightarrow a_y^T B^2 a_y = 1$$

Applying Lagrangian

$$L = w_x^T X Y^T w_y - \frac{1}{2}\alpha(w_x^T X X^T w_x - 1) - \frac{1}{2}\beta(w_y^T Y Y^T w_y - 1)$$
$$= (X a_x)^T X Y^T (Y a_y) - \frac{1}{2}\alpha((X a_x)^T X X^T (X a_x) - 1) - \frac{1}{2}\beta((Y a_y)^T Y Y^T (Y a_y) - 1)$$
$$= a_x^T \underbrace{X^T X}_{A} \underbrace{Y^T Y}_{B} a_y - \frac{1}{2}\alpha(a_x^T \underbrace{X^T X}_{A} \underbrace{X^T X}_{A} a_x - 1) - \frac{1}{2}\beta(a_y^T \underbrace{Y^T Y}_{B} \underbrace{Y^T Y}_{B} a_y - 1)$$
$$= a_x^T A . B a_y - \frac{1}{2}\alpha(a_x^T A^2 . a_x - 1) - \frac{1}{2}\beta(a_y^T . B^2 . a_y - 1)$$

$$\frac{\partial L}{\partial a_x^T} = A.B a_y - \alpha A^2 a_x \qquad \frac{\partial L}{\partial a_y^T} = B.A.a_x - \beta B^2 a_y$$

Setting it to 0.

$$A.B.a_y = \alpha A^2 a_x \qquad \times a_x^T$$
$$B.A.a_x = \beta B^2 a_y \qquad \times a_y^T$$

$$a_x^T.A.B.a_y = \alpha a_x^T A^2 a_x$$
$$a_x^T A.B.a_y = \beta a_y^T B^2 a_y$$

From auto-covariance constraint

$$a_x^T A^2 a_x = a_y^T B^2 a_y = 1.$$

follows $\alpha = \beta$.

Given $\alpha = \beta$, partial derivative becomes

$$A.B.a_y = \alpha A^2 a_x$$
$$\cancel{A.2B2a}\ B.A.a_x = \alpha B^2 a_y$$

We can write these equation ~~in~~ matrix form as

$$\begin{bmatrix} 0 & A.B \\ B.A & 0 \end{bmatrix} \begin{bmatrix} a_x \\ a_y \end{bmatrix} = \alpha \begin{bmatrix} A^2 & 0 \\ 0 & B^2 \end{bmatrix} \begin{bmatrix} a_x \\ a_y \end{bmatrix}$$

which is a generalised eigen value problem.

Exercise 1. (c)

$$\begin{bmatrix} 0 & A\cdot B \\ B\cdot A & 0 \end{bmatrix} \cdot \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

$$= \lambda \begin{bmatrix} A^2 & 0 \\ 0 & B^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

$$M = \begin{bmatrix} 0 & A\cdot B \\ B\cdot A & 0 \end{bmatrix}$$

$$N = \begin{bmatrix} A^2 & 0 \\ 0 & B^2 \end{bmatrix}$$

$$M \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \lambda N \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

$$MA = \lambda NA$$

$$N^{-1} M A = \lambda A$$

$\cancel{\alpha_x, \alpha}$ $\begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$ is the eigenvector of $N^{-1}M$

Exercise 2

(a) From generalized eigen value problem we get

$A = K_x$, $B = K_y$

$$\begin{bmatrix} 0 & K_x \cdot K_y \\ K_y \cdot K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ 0 & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

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

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

Based on the results of the 1st exercise, we can derive kernelized version of original CCA problem by going from down to up:

$$\frac{\partial L}{\partial \alpha_x^T} = K_x K_y \alpha_y - \rho K_x^2 \alpha_x$$

$$\frac{\partial L}{\partial \alpha_y^T} = K_y K_x \alpha_x - \rho K_y^2 \alpha_y$$

$$L = \alpha_x^T K_x K_y \alpha_y - \frac{1}{2}(\alpha_x^T K_x^2 \alpha_x - 1) - \frac{1}{2}\rho(\alpha_y^T K_y^2 \alpha_y - 1)$$

$$\max \ \alpha_x^T K_x K_y \alpha_y$$

$$\text{s.t. } \alpha_x^T K_x^2 \alpha_x = 1$$

$$\alpha_y^T K_y^2 \alpha_y = 1$$

(b)

Kernelized: As solution by CCA has to lie in subspace spanned by data, sufficient representation is obtained by inner product of data points (linear kernels).
Through mapping to higher dimensional space overfitting (high number of features in contrast to small set of examples) has to be avoided (--> regularization).
If one of the kernel matrices is singular or ill-conditioned the problem becomes numerically intractable (--> add low positive number on the diagonal elements of each kernel matrix (in the variance equation))

Exercise 3:

(a) $\alpha \simeq$ Regularization, no constraint only if whole term equals 0 (i.e original constraints both equal 0)

(b) $$\max \quad w_x^T E[\phi_x \phi_y^T] w_y + \alpha [\min(0, 1 - w_x^T E[\phi_x \phi_x^T] w_x) + \min(0, 1 - w_y^T E[\phi_y \phi_y^T] w_y)$$

$$\frac{\partial}{\partial \theta_x} = w_x^T E\left[\frac{\partial \phi_x}{\partial \theta_x} \phi_y^T\right] w_y + \alpha \begin{cases} 0 & \text{if } 0 <= 1 - w_x^T C_{xx} w_x \\ -w_x^T \dfrac{\partial E[\phi_x \phi_x^T]}{\partial \theta_x} \cdot w_x & \text{if } 0 > 1 - w_x^T C_{xx} w_x \end{cases}$$