M532 **Spring 2018**

Problem Set 3. Due Tuesday, March 20

Part A: Theory and Foundations

Problem 1

Show that the optimization problem

$$\text{maximize } u^T v$$

subject to $u \in \mathcal{R}(X), v \in \mathcal{R}(Y)$ given the vectors u, v which subtend the smallest angles between the subspaces $\mathcal{R}(X)$ and $\mathcal{R}(Y)$.

Problem 2

Using the notation from the CCA lecture, show that

$$a = R_X^\dagger \phi, \ \ b = R_Y^\dagger \psi$$

are solutions for the transformation vectors a, b in CCA.

Problem 3

In class we showed that the solution to the angles between subspaces optimization problem led to

$$Q_X^T Q_Y = \Phi \Sigma \Psi^T$$

Show that the elements of Σ are the cosines of the angles between the vector pairs $u^{(i)}, v^{(i)}$ where

$$u^{(i)} = Q_X \phi^{(i)}, \quad v^{(i)} = Q_Y \psi^{(i)}.$$

Problem 4

The GSVD equation can be written

$$\alpha_j X^T X \psi^{(j)} = \beta_j Q Q^T \psi^{(j)}$$

Show that the solutions $\psi^{(j)}$ to the generalized singular value problem are orthogonal in both of the following senses:

$$(\psi^{(i)})^T X^T X \psi^{(j)} = \lambda_i \delta_{ij}$$

and

$$(\psi^{(i)})^T Q^T Q \psi^{(j)} = \lambda_i \delta_{ij}$$

What can you conclude about the maximum noise fraction basis from this?

Problem 5

Show that we can approximate the covariance matrix of the noise by

$$N^T N = \frac{1}{2} dX^T dX$$

Keep careful track of your assumptions.

Part B: Computing

Download the open book/closed book data set from canvas. Take X to be the $P \times 2$ matrix of open book exam results and Y to be the $P \times 3$ matrix of closed book results.

Problem 1.

a) Compute the canonical vectors $\{a_1, a_2\}$ and $\{b_1, b_2\}$ for X and Y respectively.

b) Plot $\alpha_i = a_1^T x^{(i)}$ and $\beta_i = b_1^T y^{(i)}$. What conclusions can you draw?

c) Compute the vectors $u^{(1)}, u^{(2)}$ and $v^{(1)}, v^{(2)}$; we showed this explicitly for $u^{(1)}, v^{(1)}$ in class. Argue why the same equations produce $u^{(2)}, v^{(2)}$.

d) Verify that the angle between $u^{(1)}$ and $v^{(1)}$ is $\arccos \sigma_1$.

Problem 2.

Download the data set MNFdata.mat from canvas. This is a 629×4 data matrix X. This problem concerns the implementation of the maximum noise fraction algorithm and the SVD on this data set.

a) Compute the maximum noise fraction basis ϕ where $\phi = X\psi$ by solving the following problems

   i) the eigenvector problem

$$(dX^T dX/2)^{-1} X^T X \psi = \lambda \psi$$

   ii) the generalized eigenvector problem

$$C\psi = D\lambda\psi$$

   where $D = dX^T dX/2$ and $C = X^T X$

   iii) the GSVD of $A = dX$ and $B = X$

b) Compare your results in part a).

c) Compute the SVD of this data matrix and compare the resulting basis vectors with MNF.

d) Project the data (columns of X) onto the first 2 MNF basis vectors and the first 2 SVD left singular vectors and compare.