STAT 185 - Midterm Coding

NAME

Instructions

This coding assignment covers Principal Components, Singular Value Decomposition, Canonical Correlation Analysis, and Nonnegative Matrix Factorization. You will need the file *midterm.Rdata* available on Canvas. The commands for loading the data assume this file, *Midterm_Coding.Rmd*, and the Rdata file are in the same folder.

The data in this assignment are inherently **messy** in that they may or may not satisfy the necessary assumptions for various techniques. However, by sythnesizing ideas from multiple techniques we have discussed, in some cases you may find alternative approaches to the desired answers.

For all problems, justify your answers with relevant graphs or numerical summaries and identify any algorithms used to develop your answers.

Problems

1. The variable *problem1* contains 100 independent, identically distributed random vectors $\vec{x}_1, \dots, \vec{x}_{100}$ in $\mathbb{R}^{50}$. Do these vectors live on an affine subspace of $\mathbb{R}^{50}$? If so, also give the dimension of the subspace and give the magnitude of the orthogonal projection of the vector *test* onto this subspace.

2. The variable *problem2* contains 100 independent, identically distributed random vectors $\vec{x}_1, \dots, \vec{x}_{100}$ in $\mathbb{R}^{50}$. Do these vectors live on an affine subspace of $\mathbb{R}^{50}$? If so, also give the dimension of the subspace and give the magnitude of the orthogonal projection of the vector *test* onto this subspace.

3. The variable *problem3.X* contains 300 independent, identically distributed random vectors $\vec{x}_1, \dots, \vec{x}_{300}$ in $\mathbb{R}^{25}$ and *problem3.Y* contains 300 corresponding vectors $\vec{y}_1, \dots, \vec{y}_{300}$ in $\mathbb{R}^{10}$. For each pair of vectors $\vec{x}_i, \vec{y}_i$ is there a linear combination of the components of $\vec{x}_i$, $\vec{a}^T\vec{x}_i$, which is equal to a linear combination of $\vec{y}_1$, $\vec{b}^T\vec{y}_i$. Note, we are only intereted in 'non-trivial' combinations, i.e. $\vec{a}^T\vec{x}_i \neq 0$ and/or $\vec{b}^T\vec{y}_i \neq 0$ for $i = 1, \dots, 300$. If not, explain. If so, find all such pairs $\vec{a}, \vec{b}$ for which this is the case and print them below.

4. The variable *problem4.X* contains 300 independent, identically distributed random vectors $\vec{x}_1, \dots, \vec{x}_{100}$ in $\mathbb{R}^{10}$ and *problem4.Y* contains 300 corresponding vectors $\vec{y}_1, \dots, \vec{y}_{300}$ in $\mathbb{R}^{25}$. For each pair of vectors $\vec{x}_i, \vec{y}_i$ is there a linear combination of the components of $\vec{x}_i$, $\vec{a}^T\vec{x}_i$, which is equal to a linear combination of $\vec{y}_1$, $\vec{b}^T\vec{y}_i$. Note, we are only intereted in 'non-trivial' combinations, i.e. $\vec{a}^T\vec{x}_i \neq 0$ and/or $\vec{b}^T\vec{y}_i \neq 0$ for $i = 1, \dots, 300$. If not, explain. If so, find all such pairs $\vec{a}, \vec{b}$ for which this is the case and print them below.

5. The variable *problem5* contains 50 independent, identically distributed random vectors $\vec{x}_1, \dots, \vec{x}_{50}$ in $\mathbb{R}^{15}$ in the conical hull of a set of vectors $\vec{h}_1, \dots, \vec{k}$ in $\mathbb{R}^{15}$. Identify k and give the coordinates of $\vec{x}_1$ in the basis $\vec{h}_1, \dots, \vec{h}_k$.

6. The variable *problem6* contains 50 independent, identically distributed random vectors $\vec{x}_1, \dots, \vec{x}_{50}$ in $\mathbb{R}^{15}$. Is there a $k < 15$ and vectors $\vec{h}_1, \dots, \vec{h}_k$ in $\mathbb{R}^{15}$ with nonnegative entries such that the random vectors $\vec{x}_1, \dots, \vec{x}_{50}$ live in the positive span of $\vec{h}_1, \dots, \vec{h}_k$? If not, explain. If so, find k and the coordinates of $\vec{x}_1$ in the basis $\vec{h}_1, \dots, \vec{h}_k$.