Machine Learning Exercise Sheet 11

Dimensionality Reduction & Matrix Factorization

Homework

PCA & SVD

Problem 1: Use the SVD shown below. Suppose a new user Leslie assigns rating 3 to Alien and

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

Figure 11.6: Ratings of movies by users

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}}_{M} = \underbrace{\begin{bmatrix} .14 & 0 \\ .42 & 0 \\ .56 & 0 \\ .70 & 0 \\ 0 & .60 \\ 0 & .75 \\ 0 & .30 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} .58 & .58 & .58 & 0 & 0 \\ 0 & 0 & 0 & .71 & .71 \end{bmatrix}}_{V^{\mathrm{T}}}$$

rating 4 to Titanic, giving us a representation of Leslie in the 'original space' of $[0, 3, 0, 0, 4]$. Find the representation of Leslie in concept space. What does that representation predict about how well Leslie would like the other movies appearing in our example data?

Problem 2: Consider the latent space distribution

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\mathbf{0}, \boldsymbol{I})$$

and a conditional distribution for the observed variable $\boldsymbol{x} \in \mathbb{R}^d$,

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \boldsymbol{\Phi})$$

where $\boldsymbol{\Phi}$ is an arbitrary symmetric, positive-definite noise covariance variable. Now suppose that we make a nonsingular linear transformation of the data variables $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is a non-singular $d \times d$ matrix. If $\boldsymbol{\mu}_{ML}$, $\boldsymbol{W}_{ML}$, and $\boldsymbol{\Phi}_{ML}$ represent the maximum likelihood solution corresponding to the original untransformed data, show that $\boldsymbol{A}\boldsymbol{\mu}_{ML}$, $\boldsymbol{A}\boldsymbol{W}_{ML}$, and $\boldsymbol{A}\boldsymbol{\Phi}_{ML}\boldsymbol{A}^T$ will represent the corresponding maximum likelihood solution for the transformed data set. Finally, show that the form of the model is preserved if $\boldsymbol{A}$ is orthogonal and $\boldsymbol{\Phi}$ is proportional to the unit matrix so $\boldsymbol{\Phi} = \sigma^2\boldsymbol{I}$ (i.e. probabilistic PCA). The transformed $\boldsymbol{\Phi}$ matrix remains proportional to the unit matrix, and hence probabilistic PCA is covariant under a rotation of the axes of data space, as is the case for conventional PCA.

Problem 3: Let the matrix $\boldsymbol{X} \in \mathbb{R}^{N\times D}$ represent N data points of dimension $D = 10$ (samples stored as rows). We applied PCA to $\boldsymbol{X}$. By using the $K = 5$ top principal components, we transformed/projected $\boldsymbol{X}$ into $\tilde{\boldsymbol{X}} \in \mathbb{R}^{N\times K}$. We computed that $\tilde{\boldsymbol{X}}$ preserves 70% of the variance of the original data $\boldsymbol{X}$.

Suppose now we apply PCA on the following matrices:

a) $\boldsymbol{Y}_1 = \boldsymbol{X}\boldsymbol{S}$ where $\boldsymbol{S} = \lambda\boldsymbol{I}$, with $\lambda \in \mathbb{R}$ and $\boldsymbol{I} \in \mathbb{R}^{D\times D}$ is the identity matrix

b) $\boldsymbol{Y}_2 = \boldsymbol{X}\boldsymbol{R}$ where $\boldsymbol{R} \in \mathbb{R}^{D\times D}$ and $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}$

c) $\boldsymbol{Y}_3 = \boldsymbol{X}\boldsymbol{P}$ where $\boldsymbol{P} = \operatorname{diag}(+5, -5, \ldots, +5, -5)$ is a $D \times D$ diagonal matrix

d) $\boldsymbol{Y}_4 = \boldsymbol{X}\boldsymbol{Q}$ where $\boldsymbol{Q} = \operatorname{diag}(1, 2, 3, \ldots, D-1, D)$ is a $D \times D$ diagonal matrix

e) $\boldsymbol{Y}_5 = \boldsymbol{X} + \boldsymbol{1}_N\boldsymbol{\mu}^T$ where $\boldsymbol{\mu} \in \mathbb{R}^D$ and $\boldsymbol{1}_N$ is an N-dimensional column vector of all ones

f) $\boldsymbol{Y}_6 = \boldsymbol{X}\boldsymbol{A}$ where $\boldsymbol{A} \in \mathbb{R}^{D\times D}$ and $\operatorname{rank}(\boldsymbol{A}) = 5$

and obtain the projected data $\tilde{\boldsymbol{Y}}_1, \ldots \tilde{\boldsymbol{Y}}_6 \in \mathbb{R}^{N\times K}$ using the principal components corresponding to the top $K = 5$ largest eigenvalues of the respective $\boldsymbol{Y}_i$.

What fraction of variance of each $\boldsymbol{Y}_i$ will be preserved by each respective $\tilde{\boldsymbol{Y}}_i$? *Justify your answer.*

The answer "cannot tell without additional information" is also valid if you provide a justification.

Problem 4: You are given $N = 4$ data points: $\{\boldsymbol{x}_i\}_{i=1}^4$, $\boldsymbol{x}_i \in \mathbb{R}^3$, represented with the matrix $\boldsymbol{X} \in \mathbb{R}^{4\times 3}$.

$$\boldsymbol{X} = \begin{bmatrix} 4 & 3 & 2 \\ 2 & 1 & -2 \\ 4 & -1 & 2 \\ -2 & 1 & 2 \end{bmatrix}$$

Hint: In this task the results of all (final and intermediate) computations happen to be integers.

a) Perform principal component analysis (PCA) of the data $\boldsymbol{X}$, i.e. find the principal components and their associated variances in the transformed coordinate system. Show your work.

b) Project the data to two dimensions, i.e. write down the transformed data matrix $\boldsymbol{Y} \in \mathbb{R}^{4\times 2}$ using the top-2 principal components you computed in (a). What fraction of variance of $\boldsymbol{X}$ is preserved by $\boldsymbol{Y}$?

c) Let $\boldsymbol{x}_5 \in \mathbb{R}^3$ be a new data point. Specify the vector $\boldsymbol{x}_5$ such that performing PCA on the data including the new data point $\{\boldsymbol{x}_i\}_{i=1}^5$ leads to exactly the same principal components as in (a).

Upload a single PDF file with your homework solution to Moodle by 19.01.2020, 11:59pm CET. We recommend to typeset your solution (using LaTeX or Word), but handwritten solutions are also accepted. If your handwritten solution is illegible, it won't be graded and you waive your right to dispute that.

Problem 5: Load the notebook `exercise_11_notebook.ipynb` from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to pdf and add it to the printout of your homework.

In-class Exercises

Probabilistic PCA

Problem 6: For pPCA, we consider the latent space distribution

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I})$$

and a conditional distribution for the observed variable $\boldsymbol{x} \in \mathbb{R}^d$,

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2\boldsymbol{I})$$

a) Verify that the covariance of the marginal distribution $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{C})$ is given by $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2\boldsymbol{I}$. What is the interpretation of this result?

b) Verify that the model is unidentifiable, i.e. that the matrix $\boldsymbol{W}$ is only defined up to a rotation $\boldsymbol{R}$! What is the interpretation of this result?

c) Derive an expression for the posterior of the latent variables $p(\boldsymbol{z}|\boldsymbol{x})$!