Machine Learning Exercise Sheet 10

Dimensionality Reduction & Matrix Factorization, Part 1

In-class Exercises

Problem 1: In this exercise, we use proof by induction to show that the linear projection onto an M-dimensional subspace that maximizes the variance of the projected data is defined by the M eigenvectors of the data covariance matrix S, given by

$$S = \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T \qquad \bar{\boldsymbol{x}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n$$

corresponding to the M largest eigenvalues. In Section 12.1 in Bishop this result was proven for the case of $M = 1$. Now suppose the result holds for some general value of M and show that it consequently holds for dimensionality $M + 1$. To do this, first set the derivative of the variance of the projected data with respect to a vector $\boldsymbol{u}_{M+1}$ defining the new direction in data space equal to zero. This should be done subject to the constraints that $\boldsymbol{u}_{M+1}$ be orthogonal to the existing vectors $\boldsymbol{u}_1, \ldots \boldsymbol{u}_M$, and also that it be normalized to unit length. Use Lagrange multipliers to enforce these constraints. Then make use of the orthonormality properties of the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ to show that the new vector $\boldsymbol{u}_{M+1}$ is an eigenvector of S. Finally, show that the variance is maximized if the eigenvector is chosen to be the one corresponding to eigenvector λ_{M+1} where the eigenvalues have been ordered in decreasing value.

Problem 2: Proof that minimizing the error is equivalent to maximizing the variance.

Homework

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 $\boldsymbol{P} = \mathrm{diag}(+5, -5, \ldots, +5, -5)$ is a $D \times D$ diagonal matrix

d) $\boldsymbol{Y}_4 = \boldsymbol{X}\boldsymbol{Q}$ where $\boldsymbol{Q} = \mathrm{diag}(1, 2, 3, \ldots, D-1, D)$ 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 $\mathrm{rank}(\boldsymbol{A}) = 5$

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).

SVD

Problem 5: Use the SVD shown below. Suppose a new user Leslie assigns rating 3 to Alien and 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?

Upload a single PDF file with your homework solution to Moodle by 12.01.2022, 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.

	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

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

Problem 6: You want to perform linear regression on a data set with features $\boldsymbol{X} \in \mathbb{R}^{N \times D}$ and targets $\boldsymbol{y} \in \mathbb{R}^N$. Assume that you have already computed the SVD of the feature matrix $\boldsymbol{X} = \boldsymbol{U\Sigma V}^T$. Additionally, assume that $\boldsymbol{X}$ has full rank.

Show how we can compute the optimal linear regression weights $\boldsymbol{w}^\star$ in $\mathcal{O}(ND)$ operations by using the result of the SVD.

Hint: Matrix operations have the following asymptotic complexity

- *Matrix multiplication* $\boldsymbol{AB}$ *for arbitrary* $\boldsymbol{A} \in \mathbb{R}^{P \times Q}$ *and* $\boldsymbol{B} \in \mathbb{R}^{Q \times R}$ *takes* $\mathcal{O}(PQR)$
- *Matrix multiplication* $\boldsymbol{AD}$ *for an arbitrary* $\boldsymbol{A} \in \mathbb{R}^{P \times Q}$ *and a diagonal* $\boldsymbol{D} \in \mathbb{R}^{Q \times Q}$ *takes* $\mathcal{O}(PQ)$
- *Matrix inversion* $\boldsymbol{C}^{-1}$ *for an arbitrary matrix* $\boldsymbol{C} \in \mathbb{R}^{M \times M}$ *takes* $\mathcal{O}(M^3)$
- *Matrix inversion* $\boldsymbol{D}^{-1}$ *for a diagonal matrix* $\boldsymbol{D} \in \mathbb{R}^{M \times M}$ *takes* $\mathcal{O}(M)$

Coding

Problem 7: Download the notebook `exercise_10_notebook.ipynb` from Moodle. Fill in the missing code and run the notebook. Convert the evaluated notebook to pdf and add it to the printout of your homework.