582744 Advanced Course in Machine Learning

Exercise 1 — Due Mar 22, 10:00 AM

Rules:

1. Your submission is composed of two parts: (a) A single PDF file containing your answers to all questions, including both the pen and paper questions and the written answers and various plots for the programming questions. (b) a single compressed file (zip or tar.gz) containing your code (and nothing else). If your code is in a single file, it can be sent also a plain source code.
2. The submission should be sent directly to BOTH Arto (`arto.klami@cs.helsinki.fi`) and Aditya (`aditya.jitta@cs.helsinki.fi`).
3. All material must be renamed to your student ID. Always mention your name and student ID also in the written report.
4. The subject field of your email should be [AML][your student ID][exercise 1].
5. Please typeset your work using appropriate software such as LaTeX. However, there is no need to typeset the pen and paper answers – you can also include a scanned hand-written version.
6. The pen and paper exercises can alternatively be returned in paper form during the Tuesday lecture.

This set of exercises is due on Tuesday Mar 22, before 10:00 AM.

1 Your background (3 pts)

Explain briefly your background and level of knowledge for the prerequisities. Note that your subjective evaluations here have no impact on the course grade, and will only be needed for planning the course and for statistical purposes.

(a) Which other courses on machine learning and statistical modeling you have taken?

(b) How familiar you are with probabilities and distributions? Estimate your knowledge on a scale 0-5 where 0 means you have never heard of them and 5 indicates you are writing your MSc thesis on probability theory.

(c) How strong you are in linear algebra? Again use the scale 0-5, where 0 means you do not know what a vector is and 5 means you think of all computation in terms of tensors because that is the most natural way for you.

(d) Why did you take this course? Are you planning to finish it?

2 Machine learning in public media (3 pts)

Machine Learning and artificial intelligence have recently received notable media coverage. For example, in 2011 IBM Watson won the Jeopardy a quiz show versus the best human players using advanced question-answering algorithms. For this exercise, you should search for (online) newspaper articles that cover recent advances or uses of machine learning. List three articles that cover different topics, and explain your understanding of the techniques used. Can you make a rough guess what was done in case little or no information was given regarding the techniques? Did machine learning play a notable role or could the same have been achieved by other means as well? How was machine learning described in the article?

For each article, provide a link to an online version and describe the contents and your impressions in a couple (perhaps five to ten) of sentences.

3 Discrete probabilities (6 pts)

The following table gives joint probabilities of two random variables X and Y.

	Y = -1	Y = 0	Y = 1	Y = 2
X = 0	0.06	0.25	0.02	0.02
X = 1	0.31	0.02	0.27	0.02
X = 2	0.018	0	0.006	0.006

Perform the following calculations with pen and paper, not by programming.

(a) What are the marginal probabilities $P(X)$ and $P(Y)$?

(b) Compute the expectations $\mathbb{E}(X)$ and $\mathbb{E}(Y)$. Compute also the covariance matrix $\text{Cov}(X, Y)$.

(c) Compute the entropies $H(X)$ and $H(Y)$, and the mutual information $I(X, Y)$.

(d) Are the variables X and Y independent?

Now spend a minute thinking about how these quantities would be implemented on a numeric programming language. Could you solve each of them with a one-liner?

4 Expectations (3 pts)

Show that the variance of a sum of two random variables is given by $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$. (Exercise 2.3 in MLaPP)

5 Multivariate normal distribution (3 pts)

The multivariate normal probability density function (pdf) is given by

$$p_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^D |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right),$$

where D is the dimensionality of the random variable $\mathbf{x}$, $\boldsymbol{\mu}$ is the mean vector and $\mathbf{\Sigma}$ is the $D \times D$ covariance matrix. It is often denoted by $N(\boldsymbol{\mu}, \mathbf{\Sigma})$. Let $\mathbf{x}$ be a two-dimensional random variable with distribution $N(0, \mathbf{I})$, where $\mathbf{I}$ is a 2×2 identity matrix.

Now assume that $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an invertible square matrix. What is the pdf of $\mathbf{y}$? Solve the problem using the "change of variables" technique (section 2.6.2) and write down the full derivation. Can you think of other derivations that would be easier?

In case you do now have the book, the change of variables principle is adequately described also in Wikipedia: https://en.wikipedia.org/wiki/Probability_density_function.

6 Eigen-value decomposition (programming exercise) (6 pts)

This exercise is primarily intended for verying that you are comfortable operating with numerical matrices in your favorite programming environment, and that you can create simple plots for visualizing the results. You are free to choose the environment within reason, but you can use this exercise to guide the selection: You should pick a language so that this kind of problems can be solved with only a couple of lines of code. One good choice is Python, for which you can consult http://cs231n.github.io/python-numpy-tutorial/ for a good tutorial on numeric computation. You can also check http://hyperpolyglot.org/numerical-analysis for side-by-side reference for numerical computation in various languages.

Principal component analysis (PCA) is a linear method used for dimensionality reduction. Given a high dimensional data matrix $\mathbf{X}$ of size $D \times N$ where x_i is the i^{th} column, we want to find L linear orthogonal basis vectors $w_j \in \mathbb{R}^D$ and corresponding scores $z_i \in \mathbb{R}^L$, such that we minimize the reconstruction error given below

$$||X - \widehat{X}||_F^2 = ||X - WZ||_F^2 = \sum_{i=1}^{N} \sum_{j=1}^{D} (x_{ij} - \widehat{x}_{ij})^2,$$

where each column of $\widehat{X}$ is given by $\widehat{x}_i = W z_i$ and Z is a $L \times N$ matrix. The solution is found by computing the first L eigenvectors of the covariance matrix of X (see section 12.2.1). Perform the following steps to experiment with PCA.

(a) Load in a a small data set $\mathbf{X}$ from http://www.cs.helsinki.fi/u/aklami/teaching/AML/exercise_1_data.csv. The data is given in a CSV-file where each row corresponds to one column of X. What is N and D here?

(b) Compute the covariance matrix of $\mathbf{X}$.

(c) Compute the eigenvectors and eigenvalues of that matrix.

(d) Plot the data projected onto the first two principal components.

(e) Plot the reconstruction error as a function of the dimensions kept, starting from $L = 2$ and going to $L = D$. How does this relate to the eigenvalues computed in step (c)?