Machine Learning

Prof. Dr. Bernt Schiele

Dr. Paul Swoboda & Dr. Gerard Pons-Moll

Exercise Sheet 0 - 17.10.2018

This exercise does not have to be submitted. Please prepare the sheet for your next exercise group as it will be discussed.

Exercise 1 - Spectrum of Symmetric Matrices

Any real, symmetric matrix $A \in \mathbb{R}^{n\times n}$ has the decomposition

$$A = U\Sigma U^T,$$

where $\Sigma \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A on the diagonal and U is an orthogonal matrix in $\mathbb{R}^{n\times n}$, that is $UU^T = U^T U = \mathbb{1}$, which contains the corresponding eigenvectors (more precisely: an orthogonal basis of the eigenspace of the corresponding eigenvalue).

a. **(2 points)** Derive the eigenvalues and eigenvectors of A^k (matrix product with itself) for $k \in \mathbb{N}$.

b. **(2 points)** Prove that

$$\frac{\langle x, Ax\rangle}{\langle x, x\rangle} \leq \lambda_{\max}(A),$$

where $\langle x, y\rangle = x^T y$ is the inner product in $\mathbb{R}^n$, $\lambda_{\max}(A)$ is the largest eigenvalue of A.

Solution:

a. **(1 point)** The eigenvalues are λ_i^k, where λ_i, $i = 1, \ldots, n$ are the eigenvalues of A.
(1 point) The eigenvectors of A^k are the same as that of A. This follows by induction using the definition of an eigenvector, $Au_i = \lambda_i u_i$.

b. **(1 point)** We have

$$\langle x, Ax\rangle = \left\langle x, U\Sigma U^T x\right\rangle = \left\langle U^T x, \Sigma U^T x\right\rangle$$
$$\langle x, x\rangle = \left\langle x, UU^T x\right\rangle = \left\langle U^T x, U^T x\right\rangle$$

where we have used the definition of the transpose of a matrix and the property of orthogonal matrices. As U is orthogonal, it has full rank and thus $y = U^T x$ is a valid variable transformation with $y = 0 \iff x = 0$. Thus

$$\max_{x\neq 0} \frac{\langle x, Ax\rangle}{\langle x, x\rangle} = \max_{y\neq 0} \frac{\langle y, \Sigma y\rangle}{\langle y, y\rangle} = \max_{y\neq 0} \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2}$$

(1 point) Using, that y_i^2 is non-negative we have

$$\sum_{i=1}^n \lambda_i y_i^2 \leq \lambda_{\max} \sum_{i=1}^n y_i^2,$$

where one has equality if $y_i = 0$ if $i \notin \{j \in \{1, \ldots, n\} \mid \lambda_j = \max_{k=1,\ldots,n} \lambda_k\}$.

Exercise 2 - Empirical Mean and Covariance

Given a set of n points $X = [x_1, \dots, x_n]$, where $x_n \in \mathbb{R}^d, X \in \mathbb{R}^{d \times n}$.

a. **(2 points)** Derive the minimizer c^* for function

$$f(c) = \sum_{i=1}^{n} \|x_i - c\|_2^2\,,$$

where $\|x\|_2$ denotes the Euclidean norm $\|c\|_2 = \sqrt{\sum_{j=1}^{d} c_j^2}$.

b. **(3 points)** Show that the empirical covariance matrix for X

$$\Sigma_X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T$$

is positive semi-definite, that is $w^T \Sigma_X w \geq 0$, for all $w \in \mathbb{R}^d$, $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$.
Hint: consider using the Cauchy-Schwarz inequality, $\langle u, v\rangle^2 \leq \|u\|^2 \|v\|^2$.

Solution

a. **(1 point)** We compute the gradient and the Hessian of $f(c) := \sum_{i=1}^{n} \|x_i - c\|_2^2$,

$$\nabla_f(c) = 2\sum_{i=1}^{n} c - x_i,$$
$$(H_f)(c) = 2n\mathbb{1}.$$

(1 point) The function f is convex (Hessian is positive definite) and thus every local minimum is a global minimum. The global minimum can be computed from $\nabla_f(c) = 0$ which yields

$$c^* = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

b. We use the definition

$$w^T\Sigma_X w = \frac{1}{n}\sum_{i=1}^{n} w^T(x_i - \mu)(x_i - \mu)^T w = \frac{1}{n}\sum_{i=1}^{n}((x_i - \mu)^T w)^T((x_i - \mu)^T w)$$

Let $y = (x_i - \mu)^T w \in \mathbb{R}$, we have $w^T\Sigma_X w = \frac{1}{n}\sum_{i=1}^{n} y_i^2 \geq 0$ Hence, $w^T\Sigma_X w$ is positive semi-definite.

Exercise 3 - Multivariate Gaussian

In the lecture we have seen the multivariate Gaussian $x \sim N(\mu, \Sigma)$ where the density function is defined as

$$f(x) = (2\pi)^{-\frac{d}{2}}|\Sigma|^{-\frac{1}{2}} exp(-\frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu))$$

Now we have n multivariate Gaussian random vectors $\{x_i\}_{i=1}^{n}$, where $x_i \sim N(\mathbf{0}, \Sigma_i)$,$x_i \in \mathbb{R}^d$.

a. **(2 points)** Consider the case where all the random variables are mutually independent, derive the density function for $\sum_{i=1}^{n} x_i$.

b. **(2 points)** Consider the case $n = 2$. Given the covariance matrix $cov(x_1, x_2) = C$, derive the density function for the joint vector $\binom{x_1}{x_2} \in \mathbb{R}^{2d}$.

Solution:

a. Given the random variables are mutual independent
   (1 point) The sum of n mutual independent Gaussian random variables has a Gaussian distribution with mean

$$E[\sum_{i=1}^{n} x_i] = \sum_{i=1}^{n} E[x_i] = \sum_{i=1}^{n} \mu_i = \mathbf{0}$$

   (1 point) and covariance

$$Cov[\sum_{i=1}^{n} x_i] = \sum_{i=1}^{n} Cov[x_i] = \sum_{i=1}^{n} \Sigma_i$$

b. **(1 point)** The joint vector has mean

$$E[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}] = (\begin{smallmatrix} E[x_1] \\ E[x_2] \end{smallmatrix}) = (\begin{smallmatrix} \mu_1 \\ \mu_2 \end{smallmatrix}) = \mathbf{0}$$

   (1 point) and covariance

$$Cov[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}] = (\begin{smallmatrix} Cov(x_1,x_1), Cov(x_1,x_2) \\ Cov(x_2,x_1), Cov(x_2,x2) \end{smallmatrix}) = (\begin{smallmatrix} \Sigma_1 & C \\ C^T & \Sigma_2 \end{smallmatrix})$$