Machine Learning Assignment # 2
Universität Bern

Due date: 10/10/2018

Late submissions will incur a penalty. Submit your answers in ILIAS (as a pdf or as a picture of your written notes if the handwriting is very clear). Submission instructions will be provided via email. You are not allowed to work with others.

Calculus review [Total 100 points]

Recall that the Jacobian of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix of partial derivatives

$$Df(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \dots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

where $x = [x_1 \; x_2 \; \dots \; x_n]^\top$, $f(x) = [f_1(x) \; f_2(x) \; \dots \; f_m(x)]^\top$ and $\frac{\partial f_i(x)}{\partial x_j}$ is the partial derivative of the i-th output with respect to the j-th input.

Answer the following questions (show all the steps of your working)

1. Consider the function $f : \mathbb{R}^m \to \mathbb{R}^n$ and $f(x) = x^\top A$ where $A \in \mathbb{R}^{m\times n}$. Show that $Df(x) = A^\top$. [10 points]
   Solution. Let $y^\top = x^\top A$, and a_j denote the j-th column of A. Then, $y_j = x^\top a_j$

$$\frac{\partial y_j}{\partial x} = \frac{\partial x^\top a_j}{\partial x} = \frac{\partial a_j^\top x}{\partial x} = a_j^\top$$

   Then,

$$Df(x) = A^\top$$

2. Consider the function $g : \mathbb{R}^n \to \mathbb{R}$ and $g(x) = (x^\top x)^{m+1}$, with m an integer larger than 0. Calculate $\nabla g(x)$. [15 points]
   Solution.

$$\frac{\partial (x^\top x)^{m+1}}{\partial x} = (m+1)(x^\top x)^m \frac{\partial x^\top x}{\partial x} = (m+1)(x^\top x)^m 2x = 2(m+1)(x^\top x)^m x$$

3. Given a square matrix $A \in \mathbb{R}^{m\times m}$, when is the equality $tr(A^2) = tr(A^T A)$ true? [10 points]
   Hint: $A^2 = AA$.
   Solution

   $tr(A^T A) = tr(A^2)$ is true, if and only if $AA = A^\top A$.
   $AA = A^\top A$ is true, only when $A^\top = A$.

4. Assume $A \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{m\times n}$, and $B \in \mathbb{R}^{m\times m}$. Show that

- $\nabla_X tr(AX^T) = A$ [10 points]

  Solution Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}$.

$$tr(AX^T) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ij}$$

  Therefore, $\nabla_X tr(AX^T) = \begin{bmatrix} \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{11}} & \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{12}} & \cdots & \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{1n}} \\ \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{21}} & \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{22}} & \cdots & \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{m1}} & \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{m2}} & \cdots & \frac{\partial \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ij}}{\partial x_{mn}} \end{bmatrix} = A$

- $\nabla_X tr(X^T X) = 2X$. [10 points]

  Solution

$$tr(X^T X) = \sum_{j=1}^{n}\sum_{i=1}^{m} X_{ij}X_{ij}$$

$$\frac{\partial(tr(X^T X))}{\partial X_{ij}} = 2X_{ij} \tag{1}$$

  Hence $\nabla_X tr(X^T X) = 2X$

5. Suppose $A \in R^{m\times n}$ and is full rank, and $b \in R^m$ is a vector such that $b \notin \mathcal{R}(A)$. [20 points]
   In this case we will not be able to find a vector $x \in R^n$, such that $Ax = b$.
   Find a vector x such that Ax is as close as possible to b, as measured by the square of the Euclidean norm:

$$\|Ax - b\|_2^2$$

   Hint: $\|x\|_2^2 = x^\top x$. *Review the least squares method.*
   Solution

$$\begin{aligned}\|Ax - b\|_2^2 &= (Ax - b)^\top (Ax - b) \\ &= x^\top A^\top Ax - 2b^\top Ax + b^\top b\end{aligned}$$

$$\nabla_x (x^\top A^\top Ax - 2b^\top Ax + b^\top b) = 2A^\top Ax - 2A^\top b$$

   Setting the gradient to zero and solving for x yields:

$$x = (A^\top A)^{-1} A^\top b$$

6. Solve the following equality constrained optimization problem: [25 points]

$$\max_{x\in R^n} x^\top Ax \qquad \text{subject to } \|x\|_2^2 = 1$$

   for a symmetric matrix $A \in S^n$.
   Hint: Use the Lagrangian and the Eigen decomposition.
   Solution
   We start by constructing the Lagrangian:

$$\mathcal{L}(x, \lambda) = x^\top Ax - \lambda x^\top x$$

$$\nabla_x \mathcal{L}(x, \lambda) = 2A^\top x - 2\lambda x.$$

Setting the gradient to 0 yields:

$$2A^\top x - 2\lambda x = 0$$

$$A^\top x = \lambda x$$

Since A is symmetric:

$$Ax = \lambda x$$

and we recall the constraint $\|x\|_2^2 = 1$. We can then plug back the solution in the original cost function and obtain

$$\max_{x \in R^n} \lambda x^\top x \qquad \text{subject to } \|x\|_2^2 = 1$$

which is equivalent to

$$\max_{x \in R^n} \lambda \qquad \text{subject to } \|x\|_2^2 = 1.$$

Hence, the solution is the eigenvector corresponding to the largest eigenvalue of A.