Machine Learning Assignment # 2
Universität Bern

Due date: 09/10/2019

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. When f is a scalar-valued function (*i.e.*, when $f : \mathbb{R}^n \to \mathbb{R}$), the Jacobian $Df(x)$ is a $1 \times n$ matrix, *i.e.,* it is a row vector. Its transpose is called the *gradient* of the function

$$\nabla f(x) = Df(x)^\top = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \tag{1}$$

Also, recall that the **chain rule** is a tool to calculate gradients of function compositions. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x and $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(x)$. Define the composition $h : \mathbb{R}^n \to \mathbb{R}^p$ by $h(z) = g(f(z))$. Then h is differentiable at x, with Jacobian

$$Dh(x) = Dg(z)\Big|_{z=f(x)} Df(x). \tag{2}$$

1. Consider the function $g : \mathbb{R}^n \to \mathbb{R}$ with $g(x) = x^\top x$. We can readily calculate the gradient $\nabla g(x) = 2x$ by noticing that

$$\forall j = 1, \dots, n \qquad \frac{\partial x^\top x}{\partial x_j} = \frac{\partial x_j^2}{\partial x_j} = 2x_j \to \nabla g(x) = 2x. \tag{3}$$

   Consider also the function $a : \mathbb{R}^n \to \mathbb{R}^m$ with $a(x) = Ax$, and $A \in \mathbb{R}^{m\times n}$. The Jacobian of $a(x)$ is $Da(x) = A$. Given this, answer the following questions by using the above definitions (show all the steps of your working)

   (a) Consider the function $h : \mathbb{R}^n \to \mathbb{R}$ and $h(x) = x^\top Qx$, where $Q \in \mathbb{R}^{n\times n}$ is a symmetric matrix. **[15 points]**
   Calculate $\nabla h(x)$ by using the product rule, the gradient of g in eq. (3), and the Jacobian of the linear function $a(x)$.

   (b) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$, where $f(x) = \|Ax - b\|^2$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. **[15 points]**
   Calculate $\nabla h(x)$ by using the chain rule in eq. (2), the gradient of g in eq. (3), and the Jacobian of the linear function $a(x)$.

   (c) Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose we have a matrix $A \in \mathbb{R}^{n\times m}$ and a vector $x \in \mathbb{R}^m$.
   Calculate $\nabla_x f(Ax)$ as a function of $\nabla_x f(x)$. **[10 points]**

(d) Show that **[10 points]**

$$\frac{\partial}{\partial X} \sum_{i=1}^{n} \lambda_i = I$$

where $X \in R^{n \times n}$ and has eigenvalues $\lambda_1 \ldots \lambda_n$.

(e) Show that **[10 points]**

$$\frac{\partial}{\partial X} \prod_{i=1}^{n} \lambda_i = \det(X) X^{-T}$$

where $X \in R^{n \times n}$ and has eigenvalues $\lambda_1 \ldots \lambda_n$.

2. 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 $\nabla_X tr(AX^T B) = BA$. **[10 points]**

3. Solve the following equality constrained optimization problem: **[30 points]**

$$\max_{x \in R^n} x^\top A x \qquad \text{subject to } b^\top x = 1$$

for a symmetric matrix $A \in \mathbb{S}^n$. Assume that A is invertible and $b \neq 0$.