AI 501 Machine Learning for AI, Spring 2021

Homework assignment 1

1. (10 points) Consider the following linear mapping

$$\Phi : \mathbb{R}^3 \to \mathbb{R}^4 \tag{1}$$

where

$$\Phi\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + 2x_2 - x_3 \\ 3x_2 - 4x_3 \\ 4x_1 + 4x_2 + 3x_3 \\ x_1 \end{bmatrix} \tag{2}$$

   (a) (3 points) Find the transformation matrix $\boldsymbol{A}_\Phi$.

   (b) (3 points) Determine $\mathrm{rank}(\boldsymbol{A}_\Phi)$.

   (c) (4 points) Compute the kernel and image of Φ. What are $\dim(\ker(\Phi))$ and $\dim(\mathrm{Im}(\Phi))$?

2. (10 points) Show that subadditivity and homogeneity of norm implies $\|\boldsymbol{x}\| \geq 0$ for all $\boldsymbol{x} \in \mathcal{X}$.

   - **Subadditivity** $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$.
   - **Homogeneity** $\|a \cdot \boldsymbol{x}\| = |a| \cdot \|\boldsymbol{x}\|$ for all $a \in \mathbb{F}$ and for all $\boldsymbol{x} \in \mathcal{X}$.

3. (20 points) Let $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\forall p, q \in [1, \infty)$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Hölder's inequality is

$$\sum_{i=1}^{n} |x_i y_i| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q \tag{3}$$

   Prove the following statements using Hölder's inequality (you may not use if if not needed).

   (a) (10 points) Let's assume $p \in [1, \infty)$ and $\boldsymbol{x} \in \mathbb{R}^n$. Then for any $r \in (0, p)$, $\|\boldsymbol{x}\|_p \leq \|\boldsymbol{x}\|_r$.

   (b) (10 points) Let's assume $p \in [1, \infty)$ and $\boldsymbol{x} \in \mathbb{R}^n$. Then for any $r \in (0, p)$, $\|\boldsymbol{x}\|_r \leq n^{\frac{1}{r} - \frac{1}{p}} \|\boldsymbol{x}\|_p$.

4. (20 points) If $\boldsymbol{x} \neq 0 \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, show that

$$\max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \sigma_1 \tag{4}$$

   where σ_1 is the largest singular value of $\boldsymbol{A}$.

5. (15 points) Find the singular value decomposition of

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \tag{5}$$

6. (15 points) Use the chain rule in order to compute the ***derivative*** $Df(\boldsymbol{x})$ of the following function

$$f(z) = \exp\left(-\frac{1}{2}z\right) \tag{6}$$

$$z = g(\boldsymbol{y}) = \boldsymbol{y}^\top \boldsymbol{S}^{-1}\boldsymbol{y} \tag{7}$$

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\mu} \tag{8}$$

where $z \in \mathbb{R}$, $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^D, \boldsymbol{S} \in \mathbb{R}^{D\times D}$. **You should describe your steps in detail;** note that you can use $\frac{\partial f}{\partial \boldsymbol{x}}$ instead of $Df(\boldsymbol{x})$ for a simplicity of notation (Hint : Use 5.107 in textbook).

7. (10 points) Compute the ***derivative*** $D\boldsymbol{f}(\boldsymbol{x})$ of the following function

$$\boldsymbol{f}(\boldsymbol{z}) = \sin(\boldsymbol{z}) \tag{9}$$

$$\boldsymbol{z} = A\boldsymbol{x} + \boldsymbol{b} \tag{10}$$

where $\sin(\cdot)$ is applied to every element of $\boldsymbol{z}$ and $\boldsymbol{A} \in \mathbb{R}^{E\times D}, \boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{b} \in \mathbb{R}^E$. **You should describe your steps in detail;** note that you can use $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}$ instead of $D\boldsymbol{f}(\boldsymbol{x})$ for a simplicity of notation.