Indian Institute of Technology Kharagpur
Class Test I: 2020-21

Date: 19 Jan. 2021 **Subject No.:** CS60010 **Subject:** Deep Learning

1. (a) (4 points) The trace of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as the sum of the its diagonal entries, or $\mathrm{tr}\mathbf{A} = \sum_{i=1}^{n} \mathbf{A}_{ii}$. Prove the following fact.

$$||\mathbf{A}||_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{A}_{ij}^2} = \sqrt{\mathrm{tr}(\mathbf{A}^T\mathbf{A})}$$

where, $||\mathbf{A}||_F$, is the Frobenius norm. Show all your steps.

Solution: Let $\mathbf{C} = \mathbf{A}^T\mathbf{A}$. Then the j^{th} diagonal element of $\mathbf{C}$, *i.e.*, $\mathbf{C}_{j,j}$ is given by, $\sum_{i=1}^{n} \mathbf{A}_{ij}\mathbf{A}_{ij} = \sum_{i=1}^{n} \mathbf{A}_{ij}^2$. Now, trace of $\mathbf{C}$, *i.e.*, $\mathrm{tr}(\mathbf{C})$ is the sum of all diagonal elements, *i.e.* $\mathrm{tr}(\mathbf{C}) = \sum_{j=1}^{n} \mathbf{C}_{jj} = \sum_{j=1}^{n}\sum_{i=1}^{n} \mathbf{A}_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{A}_{ij}^2$. This implies that, $||\mathbf{A}||_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{A}_{ij}^2} = \sqrt{\mathrm{tr}(\mathbf{C})} = \sqrt{\mathrm{tr}(\mathbf{A}^T\mathbf{A})}$

(b) (3 points) Show that for a matrix $\mathbf{A}$ and vector $\mathbf{x}$, $\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}^{-1}) = -\mathbf{A}^{-1}\big(\frac{\partial}{\partial \mathbf{x}}\mathbf{A}\big)\mathbf{A}^{-1}$.
Hint: Use the fact that for any two matrices $\mathbf{A}$ and $\mathbf{B}$, $\frac{\partial \mathbf{AB}}{\partial \mathbf{x}} = \frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial \mathbf{x}}$.
Solution:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{A}^{-1}) &= 0 \qquad [\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}] \\
\frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{A}^{-1} + \mathbf{A}\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{x}} &= 0 \\
\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{A}^{-1} + \frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{x}} &= 0 \\
\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{x}} &= -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{A}^{-1}
\end{aligned}$$

2. (a) (3 points) Suppose we have a cost function

$$J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\theta}^T\mathbf{x}^{(i)} + b\mathbf{y}^{(i)} + \frac{1}{2}\boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}$$

where $\boldsymbol{\theta} \in \mathbb{R}^d$ is the parameter vector $\mathbf{x}^{(i)} \in \mathbb{R}^d$, $y^{(i)} \in \mathbb{R}$, $\{\mathbf{x}^{(i)}, y^{(i)}\}$ are N training data points, $\mathbf{A} \in \mathbb{R}^{d\times d}$ is a symmetric matrix and $b \in \mathbb{R}$. We want to find parameters $\boldsymbol{\theta}$ using gradient descent. Find the vector of partial gradients of the cost function.

Solution: $\frac{\partial}{\partial \boldsymbol{\theta}} J(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{(i)} + 0 + \mathbf{A}\boldsymbol{\theta}$

(b) (1 point) Give the closed-form solution of θ from the above expression you found.

Solution: $\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{(i)} + \mathbf{A}\boldsymbol{\theta} = 0, \implies \boldsymbol{\theta} = -\frac{1}{N}\mathbf{A}^{-1} \sum_{i=1}^{N} \mathbf{x}^{(i)}$

(c) (4 points) Let λ and $\mathbf{x}$ are respectively the eigenvalue and eigenvector of a square matrix $\mathbf{A}$. Prove that $\mathbf{x}$ is also an eigenvector of $\mathbf{A}^k$ where k is a positive integer. Also prove that λ^k is the eigenvalue of $\mathbf{A}^k$.

Solution:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \lambda\mathbf{x} \\ \mathbf{A}\mathbf{A}\mathbf{x} &= \mathbf{A}\lambda\mathbf{x} \\ \mathbf{A^2}\mathbf{x} &= \lambda\mathbf{A}\mathbf{x} = \lambda\lambda\mathbf{x} = \lambda^2\mathbf{x} \end{aligned}$$

The eigenvalues of $\mathbf{A^2}$ are the squares of the eigenvalues of $\mathbf{A}$. The eigenvectors of $\mathbf{A^2}$ are the same as the eigenvectors of $\mathbf{A}$. Similarly, it ca be proved for a generic positive integer k also.