Machine Learning

**Assignment 1**

Author: Kasra Amani
Student No. 98101171

Instructor: **Prof. Ali Sharifi Zarchi**

Date Last Edited: October 19, 2022

**Problem 1**

**Problem 1.1**

**Theorem 1:** Let $y = Ax$ where $y$ and $x$ are n-element vectors and $A$ is an $n \times n$ matrix; then: $A = \frac{\partial y}{\partial x}$.
The proof is trivial when we look at the definition of $y$.

**Theorem 2:** Let the scalar $\alpha$ be as follows: $\alpha = y^T Ax$ where $A$, $x$ and $y$ have similar features to theorem 1, we can derive that $\frac{\partial \alpha}{\partial x} = y^T A$ and $\frac{\partial \alpha}{\partial y} = x^T A^T$.
Proof: $z$ (a transposition of a $n \times 1$ vector) is defined as: $z = y^T A$ (therefore $\alpha = zx$); utilizing theorem 1 we can derive that: $\frac{\partial \alpha}{\partial x} = z = y^T A$; also since $\alpha$ is an scalar, the following is true: $\alpha = \alpha^T = x^T A^T y$ and applying theorem 1 again gives us: $\frac{\partial \alpha}{\partial y} = x^T A^T$.

**Theorem 3:** Let $\alpha$ be of quadratic form $\alpha = x^T Ax$, then: $\frac{\partial \alpha}{\partial x} = x^T(A + A^T)$.
Proof: By the definition of $\alpha$ we have: $\alpha = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$; differentiating according to the $k$th element of $\mathbf{x}$ we get: $\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$ which holds true for all $k$ in $[1, n]$ we obtain: $\frac{\partial \alpha}{\partial x} = x^T A^T + x^T A = x^T(A^T + A)$.

## a

This theorem only holds if $A$ is symmetric; we define the scalar $\alpha$ as follows: $\alpha = x^T Ax$. Then by applying theorem 3 in a special case, we can derive $\frac{\partial \alpha}{\partial x} = 2x^T A$.

## b

Since $\alpha = x^T Ax$ is a scalar, its trace is $\alpha$ itself and from there on, it is the identical theorem to theorem 3.

**Problem 1.2**

## a

Consider the characteristic polynomial: $p(\lambda) = |\lambda I - A| = \lambda^n + c_{n-1}\lambda^{n-1} + ... + c_1\lambda + c_0$.
Since the eigenvalues are the roots of this polynomial, it can be rewritten as $(\lambda - \lambda_1)...(\lambda - \lambda_n)$.
Consider $p(0) = c_0$, this can be calculated in two ways:
**1.** $p(0) = (0 - \lambda_1)...(0 - \lambda_n) = (-1)^n \lambda_1...\lambda_n$
**2.** $p(0) = |0I - 1| = |-A| = (-1)^n|A|$
So we have shown that the product of eigenvalues is equal to the determinant.

## b

Similar to the previous part, we will use two methods to calculate $c_{n-1}$;
**First**, we expand $p(\lambda) = (\lambda - \lambda_1)...(\lambda - \lambda_n)$; this will give us $c_{n-1} = -(\lambda_1 + ... + \lambda_n)$.
**Second**, it can be calculated by obtaining the determinant $|\lambda I - A|$:

$$|\lambda I - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{vmatrix}$$

One major part of the determinant is the product of all elements on the main diagonal ($(\lambda - a_{11})...(\lambda - a_{nn})$). It is clear that the remaining part has products that contain at most $n-2$ elements of the main diagonal(and therefore $n-2$ lambdas($\lambda$)). The second part has no effect on the value of $c_{n-1}$ and thus, this value can be calculated by finding the coefficient of $\lambda^{n-1}$ in the first polynomial ($(\lambda - a_{11})...(\lambda - a_{nn})$) and we obtain $c_{n-1} = -(a_{11} + ... + a_{nn})$.
So we have $c_{n-1} = -(a_{11} + ... + a_{nn}) = -(\lambda_1 + ... + \lambda_n)$ and since the first part is $trace(A)$, we obtain that if we add all eigenvalues, the result is the trace of that matrix.

**Problem 1.3**

**Problem 1.4**

$||A||_1 = \max_j(\sum_{i=1}^{n} |a_{ij}|)$

$||A||_2 = \sqrt{\lambda_{max}(A^*A)}$

$||A||_\infty = \max_i(\sum_{j=1}^{n} |a_{ij}|)$

$$A = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix} \implies ||A||_1 = 8, ||A||_2 = \sqrt{51.089} = 7.147, ||A||_\infty = 11$$

**Problem 2**

**Problem 2.1**

**Problem 2.2**

**Problem 2.3**

Similar matrices: $A$ and $B$ are similar if the invertible matrix $P$ exists such that: $B = P^{-1}AP$ so essentially, similar matrices represent the same map under different bases where $P$ does the basis change and because of this, they have identical eigenvalues. Thus all we have to do, is show that $M$ and $N$ are similar where $N = P^{-1}MP$; consider the matrix $P^{-1}$ (which is invertible); we know that: $P^{-1^{-1}}NP^{-1} = PP^{-1}MPP^{-1} = M$ and we conclude that $M$ and $N$ are similar and thus, have the same set of eigenvalues.