Math 4570-Matrix Methods-Fall 2021

Instructor: He Wang

Test 2

Student Name: ____________________ ____________/**50**

Rules and Instructions for Exams:

1. Unless otherwise specified, to receive full credits you must show **all** necessary work. The grading is based on your work shown. Only a final result from computer will receive zero point.

2. You need to finish the exam yourself. Any discussions with the other people will be considered as **academic dishonesty**. **Cheating, Unauthorized Collaboration, and Facilitating Academic Dishonesty are not allowed.** You can read a description of each here http://www.northeastern.edu/osccr/academic-integrity-policy/

3. This is an open exam. You are allowed to look at textbooks, and use a computer.

4. You are **not** allowed to discuss with any other people.

5. You are **not** allowed to ask questions on any internet platform.

6. For programming questions, please following the specific instruction on the use of libraries.

Notation: $\vec{x} \in \mathbb{R}^n$ means a column vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

1. (10 points) Let $\mathbb{R}^5$ be the Euclidean space with dot product. Let V be a subspace spanned by

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

(1) Apply the Gram-Schmidt process to find the **orthonormal** basis of V.

Denote $\vec{b}_1 = \vec{u}_1$ and $\vec{b}_2 = \vec{u}_2$.

$\vec{v}_1 = \vec{b}_1 \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\vec{v}_2 = \vec{b}_2 - \frac{\langle \vec{b}_2, \vec{v}_1 \rangle}{\vec{v}_1, \vec{v}_1} \vec{v}_1 = \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 4 \\ 3 \\ 0 \end{bmatrix}$

So, the orthonormal basis is formed by

$\frac{\vec{v}_1}{||\vec{v}_1||} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\frac{\vec{v}_2}{||\vec{v}_2||} = \frac{1}{\sqrt{30}} \begin{bmatrix} -2 \\ -1 \\ 4 \\ 3 \\ 0 \end{bmatrix}$

(2) Find the **orthogonal complement** of V.

$V^\perp = \ker A$ where $A = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \end{bmatrix}$ or $A = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ -2 & -1 & 4 & 3 & 0 \end{bmatrix}$

Solve the equation $A\vec{x} = 0$ we get a basis for the **orthogonal complement** $V^\perp = \ker A$.

$$\begin{bmatrix} 3 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

(3) Find a formula to calculate the **shortest** distance from $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$ to V.

The shortest distance from $\vec{x}$ to V is

$||\vec{x}^\perp|| = ||\vec{x} - Prog_V \vec{x}||$

From (1) we have orthogonal basis for V.

So $Prog_V \vec{x} = \frac{\vec{x} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 + \frac{\vec{x} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2$.

So $||\vec{x}^\perp|| = ||\vec{x} - \frac{\vec{x} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 + \frac{\vec{x} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2||$

2. (10 points) For any two continuous functions $f(x)$ and $g(x)$, let the inner product

$$\langle f, g\rangle = \int_0^{\pi} f(x)g(x)dx.$$

(1) Find an **orthogonal** basis for the inner product space $P = \text{Span}\{1,\ 2x+3x^2\}$

Let $\vec{b}_1 = 1$ and $\vec{b}_2 = 2x + 3x^2$
By Gram-Schmidt process,
$\vec{v}_1 = \vec{b}_1 = 1$ and
$\vec{v}_2 = \vec{b}_2 - \dfrac{\langle \vec{b}_2, \vec{v}_1\rangle}{\vec{v}_1, \vec{v}_1}\vec{v}_1 = 2x + 3x^2 - \dfrac{\int_0^{\pi} 2x + 3x^2 dx}{\int_0^{\pi} 1 dx} = 2x + 3x^2 - (\pi - \pi^2)$

(2) Find the **least squares approximation** to the function $f(x) = \sin x$ by a quadratic function $a+bx+cx^2$ in the interval $[0, \pi]$.

(You may need: $\int_0^{\pi} \sin(x)dx = 2$; $\int_0^{\pi} x\sin(x)dx = \pi$; $\int_0^{\pi} x^2\sin(x)dx = \pi^2 - 4$; $\int_0^{\pi} x^3\sin(x)dx = \pi(\pi^2 - 6)$)

Let $W = \text{Span} 1, x, x^2$ and suppose $\vec{z} = a + bx + cz$. Then $(\vec{z} - f(x)) \perp W$. So we have

$$\begin{cases} \langle \vec{z} - f(x), 1\rangle = 0 \\ \langle \vec{z} - f(x), x\rangle = 0 \\ \langle \vec{z} - f(x), x^2\rangle = 0 \end{cases}$$

So,

$$\begin{cases} \int_0^{\pi} a + bx + cz - \sin x = 0 \\ \int_0^{\pi} (a + bx + cz - \sin x)x = 0 \\ \int_0^{\pi} (a + bx + cz - \sin x)x^2 = 0 \end{cases}$$

Calculate the integrals, we have

$$\begin{cases} \pi a + \frac{1}{2}\pi^2 b + \frac{1}{3}\pi^3 c - 2 = 0 \\ \frac{1}{2}\pi^2 a + \frac{1}{3}\pi^3 b + \frac{1}{4}\pi^4 c - \pi = 0 \\ \frac{1}{3}\pi^3 a + \frac{1}{4}\pi^4 b + \frac{1}{5}\pi^5 c - \pi^2 - 4 = 0 \end{cases}$$

Solve the linear system by rref, we have the least squares approximation to the function $f(x) = \sin x$ by a quadratic function $a + bx + cx^2$ in the interval $[0, \pi]$ is .

$$\vec{z} = -0.0505 + 1.3122x - 0.4177x^2$$

3. (10 points) Let $X \in \mathbb{R}^{n\times d}$ and $\vec{b} \in \mathbb{R}^n$ and let $J(\vec{\theta}) = \frac{1}{n}||X\vec{\theta} - \vec{b}||^2$. Here the norm $||\ \ ||$ is the standard l_2-norm defined by dot product. You can use **any** results in the lecture notes.
(1) Calculate the **gradient** of the function $J(\vec{\theta})$.

(2) Calculate **Hessian matrix** of $J(\vec{\theta})$.

(3) Write down the update formula for approximating $\mathbf{argmin}_\theta J(\vec{\theta})$ using **Gradient Decent**, using α for the learning rate.

(4) Write down the update formula for approximating $\mathbf{argmin}_\theta J(\vec{\theta})$ using **Newton's method**.

(5) Find the $\mathbf{argmin}_\theta J(\vec{\theta})$.

In matrix calculus section, we did most of the calculations
(1)
$$J(\vec{\theta}) = \frac{1}{n}||X\vec{\theta} - \vec{b}||^2 = \frac{1}{n}(X\vec{\theta} - \vec{b})^T(X\vec{\theta} - \vec{b}) = \frac{1}{n}(\vec{\theta}^T X^T X\vec{\theta} - 2\vec{b}^T X\vec{\theta} - \vec{b}^T\vec{b})$$
So, the gradient of $J(\vec{\theta})$ is
$$\nabla J(\vec{\theta}) = \frac{1}{n}(2X^T X\vec{\theta} - 2X^T\vec{b})$$
(2) The Hessian matrix of $J(\vec{\theta})$ is
$$H(J(\vec{\theta})) = 2X^T X$$
(3)
$$\vec{\theta}^{(t+1)} = \vec{\theta}^{(t)} - \alpha\nabla J = \theta^{(t)} + 2\frac{\alpha}{n}X^T(X\vec{\theta}^{(t)} - \vec{b})$$
(4)
$$\vec{\theta}^{(t+1)} = \vec{\theta}^{(t)} - H^{-1}\nabla J = \theta^{(t)} - (X^T X)^{-1}X^T(X\vec{\theta}^{(t)} - \vec{b}) = (X^T X)^{-1}X^T\vec{b}$$
if $rank(X) = d$
(5)
$\mathbf{argmin}_\theta(J)$ is $(X^T X)^{-1}X^T\vec{b}$ if $rank(X) = d$

4. (10 points) Consider the data

$x^{(i)}$	0	0.2	0.4	0.6	0.8	1	1.2	1.4
$y^{(i)}$	5.1	6.4	6.1	8.2	9.5	8.6	12	14.8

You may use Python (with **only** numpy library) to solve the matrix calculations.

(1) Use the Method of Least Squares to fit a **linear** model $f(x) = \theta_0 + \theta_1 x_1$ to this dataset.

The least squares solution is

$$\vec{\theta} = (X^T X)^{-1} X^T \vec{y}$$

$X = \begin{bmatrix} 1 & \vec{x}^{(1)} \\ 1 & \vec{x}^{(2)} \\ \vdots & \vdots \\ 1 & \vec{x}^{(n)} \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$. So, $\theta_0 = 4.475$ and $\theta_1 = 6.2321$

(2) Use the Method of Least Squares to fit a **quadratic** model $g(x) = \theta_0 + \theta_1 x_1 + \theta_2 x^2$ to this dataset.

$$\vec{\theta} = (X^T X)^{-1} X^T \vec{y}$$

Here $X = \begin{bmatrix} 1 & \vec{x}^{(1)} & (\vec{x}^{(1)})^2 \\ 1 & \vec{x}^{(2)} & (\vec{x}^{(12)})^2 \\ \vdots & \vdots & \vdots \\ 1 & \vec{x}^{(n)} & (\vec{x}^{(n)})^2 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$ So, $g(x) = 5.5208 + 1.0030x + 3.7351x^2$

(3) Calculate and compare the RSS cost $RSS(\theta) = ||X\vec{\theta} - \vec{b}||^2$ for the above **linear** fit and **quadratic** fit.

RSS cost $RSS(\theta)$ for linear model is 8.61
RSS cost $RSS(\theta)$ for quadratic model is 4.86

5. (10 points) Consider the classification problem consisting of a data set with two labels:

Label 0:

X_1	0.2	0.6	2	2.6	3.1	3.8
X_2	3.4	1.8	2	2.7	3.5	1.5

Label 1:

X_1	-0.7	-2.1	-2.5	-3	-3.9
X_2	-2.9	-2.8	-1.3	-2	-1.5

Use **logistic regression** $p(Y = 1|\vec{x}) = \dfrac{1}{1 + e^{-\theta^T \vec{x}}}$ to classify the data. (In this question, you can use any Python library including Scikit-learn.)

(1) Find the **logistic function** $h(\vec{x}) = \dfrac{1}{1 + e^{-\theta^T \vec{x}}}$.

$$p(y = 1|\vec{x}) = \frac{1}{(1 + \exp(0.2237 + 0.6839x_1 + 0.8666x_2)}$$

(2) Find the formula for the line forming the **decision boundary**.

The decision boundary is $0.2237 + 0.6839x_1 + 0.8666x_2 = 0$

(3) Find the probability $P(y = 0|\vec{x})$ and $P(y = 1|\vec{x})$ for a test point $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for the logistics model in the above question.

For Logistic regression: $p(y = 0|\vec{x}) = \dfrac{1}{(1 + \exp(-(0.2237 + 0.6839x_1 + 0.8666x_2))} = 0.556$ when $\vec{x} = (0, 0)$
$p(y = 0|\vec{x}) = 1 - 0.556 = 0.444$

(4). What is the predicted label for the point $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$?

label is 0.