Homework #1 — MTH 602

Assigned: Wednesday, September 10, 2025
Due: Monday, September 22, 2025 (print off and submit your report in class if possible)

Hardcopy reports should be submitted to Scott by Monday, September 22 at 5pm. If you cannot make it to class, you must find some other way to submit the report. **You can work in groups of 1, 2, or 3.**

Report: Put all the figures and answers asked for in the questions into a single, well-organized PDF document (prepared using LaTeX, a Jupyter notebook, or something else). **Please submit your report by uploading it to your git project and printing it off to hand in.** Ensure your submission is *reproducible*: I should be able to clone your repository and generate your reported results by using available code. Paper and pencil questions can be included in the PDF report or turned in separately

I. MATH REFRESHER: LINEAR ALGEBRA & VECTOR CALCULUS

1. (10 points) Consider the scalar function

$$\phi(x_1, x_2, x_3) = \frac{1}{\sqrt{x_1^2 + x_2^2 + x_3^2}}. \tag{1}$$

   (a) (5 points) Compute its gradient, $\vec{F} = \nabla\phi(x_1, x_2, x_3)$.

   (b) (5 points) Evaluate and sketch the vector field $\vec{F}$ at the following points: $(1, 0, 0)$, $(-1, 0, 0)$, $(0, 1, 0)$, $(0, -1, 0)$. Clearly indicate both the direction and magnitude (using the vector 2-norm) of each vector.

2. (10 points) Recall that the inner product between two vectors $\vec{x}, \vec{y} \in \mathbb{R}^N$ is

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{N} x_i y_i = \vec{x}^T \vec{y}, \tag{2}$$

   and that $\vec{x}$ and $\vec{y}$ are orthogonal if $\langle \vec{x}, \vec{y} \rangle = 0$. Let $Q \in \mathbb{R}^{N\times N}$ be an orthogonal matrix, meaning $Q^T Q = I$ (equivalently, $Q^T = Q^{-1}$). Show that the action of Q preserves both angles and lengths:

   (a) (5 points) Show that $\langle Q\vec{x}, Q\vec{y} \rangle = \langle \vec{x}, \vec{y} \rangle$.

   (b) (5 points) Deduce that $\|Q\vec{x}\|_2 = \|\vec{x}\|_2$.

Optional: Why is the fact that orthogonal matrices preserve lengths and angles important in machine learning methods like PCA? Write a short (2 to 3 sentence) explanation.

3. (10 points) Verify the following inequalities relating the L^2 and L^∞ vector norms of $\vec{x} \in \mathbb{R}^N$:

   (a) (5 points) $\|\vec{x}\|_\infty \le \|\vec{x}\|_2$.

   (b) (5 points) $\|\vec{x}\|_2 \le \sqrt{N}\, \|\vec{x}\|_\infty$.

4. (10 points) In class we derived the closed-form solution for the linear least-squares problem. Here you will fill in the details (for full credit, please justify all steps). Following the notation and definitions from class, let $A \in \mathbb{R}^{N\times(M+1)}$, $\vec{y} \in \mathbb{R}^N$, $\vec{\omega} \in \mathbb{R}^{M+1}$ and consider

$$\min_{\vec{\omega}} \|\vec{y} - A\vec{\omega}\|_2^2.$$

   (a) (5 points) Expand the squared error to show

$$\|\vec{y} - A\vec{\omega}\|_2^2 = \vec{\omega}^T A^T A\vec{\omega} - 2\vec{\omega}^T A^T \vec{y} + \vec{y}^T \vec{y}.$$

(b) (5 points) Compute the gradient with respect to $\vec{\omega}$ and show

$$\nabla_{\vec{\omega}} \|\vec{y} - A\vec{\omega}\|_2^2 = 2A^T A\vec{\omega} - 2A^T \vec{y}.$$

You may find the following facts useful:

$$\nabla_{\vec{\omega}} \left(\vec{\omega}^T \vec{y}\right) = \vec{y} \qquad \nabla_{\vec{\omega}} \left(\vec{\omega}^T B\vec{\omega}\right) = \left(B + B^T\right)\vec{\omega}$$

5. (20 points) The least-squares solution has a geometric interpretation: it is the orthogonal projection of $\vec{y} \in \mathbb{R}^N$ onto the column space (range) of A. Define the projection matrix

$$P = A(A^T A)^{-1} A^T.$$

Show the following:

(a) (5 points) P is a projection operator: $P(P\vec{y}) = P\vec{y}$.

(b) (5 points) If $\vec{y}_M = A\vec{\omega}$ is any vector in the range of A (that is $\vec{\omega}$ is any vector not necessarily the optimal least squares solution), then $P\vec{y}_M = \vec{y}_M$.

(c) (5 points) Let $\vec{r} = \vec{y} - \vec{y}_M$ be the residual of the least-squares solution. That is, we now use the optimal least squares solution vector $\vec{y}_M = A\vec{\omega}_*$. Show that $P\vec{r} = 0$.

(d) (5 points) Conclude that P is an orthogonal projection operator that (i) maps vectors $\vec{y} \in \mathbb{R}^N$ to the optimal least squares solution $\vec{y}_M = A\vec{\omega}_*$ and (ii) maps the residual vector $\vec{r}$ to zero.

Optional: In the projection view of least squares, the columns of A represent features. What special simplifications occur if the features are orthogonal? Why might this be beneficial for interpretability in machine learning models? Write a short (2 to 3 sentence) explanation.

II. CODING COMPETITION: POLYNOMIAL REGRESSION, CONDITIONING, AND OVERFITTING

You are given a noisy dataset $\{t_i, y_i\}_{i=1}^N$, provided as a CSV file (hw1.csv in the class github project). The data was generated from a polynomial of degree at most 9 (some coefficients may be zero), plus noise. Your task is to fit polynomial regression models of varying degree using least squares, and to evaluate both predictive performance and numerical stability. The group that predicts the coefficients to highest accuracy gets bonus points!

1. Write a function to construct the Vandermonde-ish (design) matrix: Your function should take in the polynomial degree M, the time samples $\vec{t}$ and return the matrix $A \in \mathbb{R}^{N\times(M+1)}$.

2. Fit the model: For a chosen degree M, compute the least squares solution

$$\vec{\omega} = (A^T A)^{-1} A^T \vec{y}.$$

and report the coefficients $\omega_0, \ldots, \omega_M$ and plot the fitted polynomial against the data. Do this for a few representative examples of $M = \{1, 3, 9\}$. (Warning: for poorly conditioned problems forming the inverse is problematic. Feel free to use a better approach than direct inversion.)

3. Training vs. testing error: Randomly split the dataset into training (80%) and testing (20%) data. Compute the mean squared error (MSE) on both sets for the same values of M you used in the previous step.

4. Condition number and stability: For each polynomial degree $M = \{1, 3, 9\}$ and different training subsample sizes $N = 10, 40, 100$, compute the condition number of the least-squares problem (Note: this is interpolation when $N = 10$ and $M = 9$). Plot the condition number as a function of M for several values of N. Briefly explain: what does a large condition number imply about the stability of the solution? From the viewpoint of conditioning, what appears to be a reasonable value of N and M?

5. Model selection and overfitting: Plot the training MSE, testing MSE, and the condition number as functions of M. Based on these plots, argue for a reasonable choice of degree M and training dataset size N that balances accuracy (low test error) and stability (moderate condition number, good generalization error).

A. Class Competition (bonus +10 points)

The dataset was generated from a polynomial of degree ≤ 9 with some coefficients potentially zero. Your challenge is to identify the underlying polynomial as accurately as possible. Submit your best estimate of the coefficients $\omega_0, \ldots, \omega_9$ (you may pad with zeros if you believe the true degree is lower). Accuracy will be measured against the hidden ground-truth coefficients using the ℓ_2 norm:

$$\|\vec{\omega}_{\text{your solution}} - \vec{\omega}_{\text{true}}\|_2.$$

The solution with the smallest error will earn 10 bonus points!!