MATH 21B, MARCH 7: LINEAR LEAST SQUARES, DATA FITTING, AND AN INTRODUCTION TO DETERMINANTS

(1) We want to find a quadratic $f(x) = c_2x^2 + c_1x + c_0$ such that $f(-1) = 1, f(0) = 0, f(1) = 2$, and $f(2) = 5$. Write down a matrix equation to solve for $\vec{x} = \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix}$.

Solution: This gives us the system of equations

$$c_2(-1)^2 + c_1(-1) + c_0 = 1$$
$$c_2(0)^2 + c_1(0) + c_0 = 0$$
$$c_2(1)^2 + c_1(1) + c_0 = 2$$
$$c_2(2)^2 + c_1(2) + c_0 = 5$$

which can be written using the matrix equation

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 5 \end{bmatrix}$$

Bonus: I'm going to demonstrate finding the least squares solution to this system, using the material in the next two pages of this sheet. We've already found that $A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$, so $A^T = \begin{bmatrix} 1 & 0 & 1 & 4 \\ -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}$. Multiplying both sides of $A\vec{x} = \vec{b}$ on the left by A^T gives the normal equation

$$\begin{bmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} 23 \\ 11 \\ 8 \end{bmatrix}$$

Solving by row-reduction gives us $c_0 = 3/10$, $c_1 = 2/5$, $c_2 = 1$. Thus, the best-fit quadratic is $\boxed{f(t) = t^2 + \frac{2}{5}t + \frac{3}{10}}$. You can check that $f(-1) = 0.9$, $f(0) = 0.3$, $f(1) = 1.7$, $f(2) = 5.1$, a pretty good fit! This means that $\begin{bmatrix} 0.9 \\ 0.3 \\ 1.7 \\ 5.1 \end{bmatrix}$ is the projection of $\begin{bmatrix} 1 \\ 0 \\ 2 \\ 5 \end{bmatrix}$ onto the image of A.

Least Square Solutions: Let $A\vec{x} = \vec{b}$ be an inconsistent system of linear equations. A vector $\vec{x}^*$ which makes $A\vec{x}^*$ as close as possible to $\vec{b}$ is called a *least square solution.* Such vectors $\vec{x}^*$ are solutions to the consistent system

$$A\vec{x} = \mathrm{proj}_{\mathrm{im}(A)}(\vec{b}) = \vec{b} - \vec{b}^{\perp}$$

(2) Let $\vec{b}^{\perp}$ be the part of $\vec{b}$ orthogonal to im(A).

(a) Explain why $A^T\vec{b}^{\perp} = 0$.

Solution: $\vec{b}^{\perp}$ is orthogonal to each column $\vec{v}_1, \ldots, \vec{v}_n$ of A, i.e. $\vec{v}_i \cdot \vec{b}^{\perp} = 0$ for $i = 1, \ldots, n$. These are the row vectors of A^T. So it's clear that $A^T\vec{b}^{\perp} = 0$.

(b) Using this fact, show that $\vec{x}^*$ satisfies the equation $A^TA\vec{x}^* = A^T\vec{b}$. (The equation $A^TA\vec{x} = A^T\vec{b}$ is called the *normal equation* of $A\vec{x} = \vec{b}$.)

Solution: $\vec{x}^*$ satisfies $A\vec{x}^* = \vec{b} - \vec{b}^{\perp}$. Apply A^T to both sides of this equation:

$$A^TA\vec{x}^* = A^T(\vec{b} - \vec{b}^{\perp}) = A^T\vec{b} - A^T\vec{b}^{\perp} = A^T\vec{b}$$

(c) Show that if A has linearly independent columns, then $\ker(A^TA) = 0$ and therefore A^TA is invertible. (Hint: If $A^TA\vec{w} = 0$ for some vector $\vec{w}\vec{v}$, then $A\vec{w}$ is orthogonal to all of the columns of A. But remember that $A\vec{w}$ is some linear combination of the columns of A.)

Solution: Suppose that $\vec{w}$ is a nonzero vector that's in the kernel of A^TA, i.e. $A^TA\vec{w} = 0$. This means that the vector $A\vec{w}$ is orthogonal to the rows of A^T, i.e., it is orthogonal to the columns $\vec{v}_1, \ldots, \vec{v}_n$ of A, and thus lies outside of the span of $\vec{v}_1, \ldots, \vec{v}_n$. However, $A\vec{w}$ is a linear combination $w_1\vec{v}_1 + \ldots + w_n\vec{v}_n$ of these columns! We have a contradiction, and therefore, no such vector $\vec{w}$ can exist.

Solution written a different way: If $A^TA\vec{w} = 0$, then $A\vec{w}$ is in the kernel of A^T. But $\ker(A^T) = \mathrm{im}(A)^{\perp}$. Therefore, $A\vec{w}$ is contained in both im(A) and im$(A)^{\perp}$, and thus must be the zero vector.

Unique Least Square Solution: If the columns of A are linearly independent (i.e., $\ker(A) = 0$) then the least square solution is unique, and is given by the formula

$$\vec{x}^* = (A^TA)^{-1}A^T\vec{b}$$

(3) Given a matrix A with linearly independent columns and a vector $\vec{b}$ which is not on im(A), use least squares on the inconsistent system $A\vec{x} = \vec{b}$ to calculate a formula for $\mathrm{proj}_{\mathrm{im}(A)}(\vec{b})$. (Note: if the columns of A were orthonormal, then we have the formula $\mathrm{proj}_{\mathrm{im}(A)}(\vec{b}) = AA^T\vec{b}$. The method above works even if the columns are *not* orthonormal.)

Solution: The least squares solution is $\vec{x}^* = (A^TA)^{-1}A^T\vec{b}$. Since this satisfies $A\vec{x}^* = \mathrm{proj}_{\mathrm{im}(A)}(\vec{b})$, we have the general formula

$$\boxed{\mathrm{proj}_{\mathrm{im}(A)}(\vec{b}) = A(A^TA)^{-1}A^T\vec{b}}$$

(4) We are going to use least squares to find the best linear fit $y = rx + s$ for the points $(1, 0), (2, 1)$, and $(3, 3)$.

(a) Write down an inconsistent system $A\vec{x} = \vec{b}$ which solves for the coefficients $\vec{x} = \begin{bmatrix} r \\ s \end{bmatrix}$.[1]

Solution: As in the first problem on this worksheet, we have the (potentially inconsistent) matrix equation:

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

(b) Compute A^TA and its inverse.

Solution: By standard matrix multiplication/inversion,

$$A^TA = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \qquad (A^TA)^{-1} = \begin{bmatrix} 1/2 & -1 \\ -1 & 7/3 \end{bmatrix}$$

(c) Compute $\vec{x}^* = (A^TA)^{-1}A^T\vec{b}$, the least square solution.

Solution:

$$\vec{x}^* = \begin{bmatrix} 1/2 & -1 \\ -1 & 7/3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/2 & -1 \\ -1 & 7/3 \end{bmatrix} \begin{bmatrix} 11 \\ 4 \end{bmatrix} = \begin{bmatrix} 3/2 \\ -5/3 \end{bmatrix}$$

(d) What is the best-fit line? Try plugging in $x = 1, 2, 3$.

Solution: The best-fit line is $\boxed{y = \frac{3}{2}x - \frac{5}{3}}$. This line passes through the points $(1, -1/6), (2, 4/3), (3, 17/6)$.

Determinant: The *determinant* of an $n \times n$ matrix A is defined as the sum

$$\det(A) = \sum_{\pi} (-1)^{|\pi|} A_{1\pi(1)} A_{2\pi(2)} \cdots A_{n\pi(n)}$$

where π varies over the $n!$ different permutations of the sequence $\{1, 2, \ldots, n\}$ and $|\pi|$ is the number of *up-crossings* in the pattern given by π. The determinant satisfies the following properties (some will only be discussed next class):

- $\det(A) = \det(A^T)$
- If A is upper triangular, then the determinant of A is the product of the diagonal entries.

---

[1]Notation confusion: The vector of coefficients, $\vec{x}$, is unrelated to the x-coordinate of the original question.

- Scaling a row or column by c scales $\det(A)$ by c.
- A is invertible $\iff \det(A) \neq 0$.
- Swapping two rows or two columns multiplies $\det(A)$ by -1.
- Adding a multiple of one row (or column) to another row (or column) does not change $\det(A)$.

Laplace Expansion: We can go along the entries in the first column of A and get an expression for the determinant

$$\det(A) = A_{11}\det(B_{11}) - A_{21}\det(B_{21}) + \ldots + (-1)^{n-1}A_{n1}\det(B_{n1})$$

where B_{i1} refers to the $(n-1)\times(n-1)$ matrix formed by removing the i-th row and first column of A. The same can be done with the entries in any other column (since we can swap columns and this introduces a sign) or any row (since we can transpose).

(5) Calculate the determinants.

(a) $\begin{bmatrix} 1 & 0 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{bmatrix}$.

Solution: $\det \begin{bmatrix} 1 & 0 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{bmatrix} = (1)(-1)(-2) + (0)(4)(2) + (3)(0)(1) - (1)(4)(1) - (0)(0)(-2) - (3)(-1)(2) = 2+0+0-4-0+6 = \boxed{4}$.

(b) $\begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 7 & 1 & 0 \\ 1 & 2 & 4 & 1 \\ 5 & 10 & 2 & 5 \end{bmatrix}$

Solution: We swap the first and second row (which multiplies the determinant by -1), and then use Laplace expansion along the second row.

$$-0\det\begin{bmatrix} 2 & 0 & -1 \\ 2 & 4 & 1 \\ 10 & 2 & 5 \end{bmatrix} + 7\det\begin{bmatrix} 1 & 0 & -1 \\ 1 & 4 & 1 \\ 5 & 2 & 5 \end{bmatrix} - 1\det\begin{bmatrix} 1 & 2 & -1 \\ 1 & 2 & 1 \\ 5 & 10 & 5 \end{bmatrix} + 0\det\begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 4 \\ 5 & 10 & 2 \end{bmatrix}$$

The first and last term are zero. The third determinant is zero because its second and third rows and multiples of each other. So our answer is

$$7\det\begin{bmatrix} 1 & 0 & -1 \\ 1 & 4 & 1 \\ 5 & 2 & 5 \end{bmatrix} = \boxed{252}$$

(c) $\begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ -1 & 5 & 3 \end{bmatrix}$

Solution: This has determinant $\boxed{0}$ because the first and second rows are multiples of one another.

(d) $\begin{bmatrix} 0 & 0 & 0 & 2 \\ 1 & 2 & 4 & 5 \\ 0 & 7 & 2 & 9 \\ 0 & 0 & 6 & 4 \end{bmatrix}$

Solution: Using Laplace expansion,

$$\det \begin{bmatrix} 0 & 0 & 0 & 2 \\ 1 & 2 & 4 & 5 \\ 0 & 7 & 2 & 9 \\ 0 & 0 & 6 & 4 \end{bmatrix} = -2 \det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 7 & 2 \\ 0 & 0 & 6 \end{bmatrix} = (-2)(1) \begin{bmatrix} 7 & 2 \\ 0 & 6 \end{bmatrix} = (-2)(1)(7)(6) = \boxed{-84}$$

(6) For what values of λ is the matrix $\begin{bmatrix} 3-\lambda & -2 & 6 \\ 1 & -\lambda & 10 \\ 0 & 0 & 7-\lambda \end{bmatrix}$ invertible?

Solution: Take the determinant by Laplace expanding along the bottom row:

$$\det \begin{bmatrix} 3-\lambda & -2 & 6 \\ 1 & -\lambda & 10 \\ 0 & 0 & 7-\lambda \end{bmatrix} = (7-\lambda) \det \begin{bmatrix} 3-\lambda & -2 \\ 1 & -\lambda \end{bmatrix} = (7-\lambda)((3-\lambda)(-\lambda)+2)$$

$$= (7-\lambda)\lambda^2 - 3\lambda + 2) = (7-\lambda)(\lambda-2)(\lambda-1)$$

This is nonzero (and thus the original matrix is invertible) as long as $\lambda \neq 7, 2, 1$.