LINEAR ALGEBRA AND VECTOR ANALYSIS

MATH 22B

Unit 10: Data Fitting

Lecture

10.1. A $n \times m$ matrix A defines a linear space $V = \text{im}(A) \subset \mathbb{R}^n$. If there is no kernel of A, then the image is a m-dimensional subspace of the n-dimensional space. Given a vector $b \in \mathbb{R}^n$, we can solve the equation $Ax = b$ only if b is in V. If n is larger than m, we in general can not find a solution, but we can try to find a "solution which is best".

10.2. The best possible solution x^* to this equation is the vector x^* for which Ax^* is closest to b. This means that $b - Ax^*$ is perpendicular to V, meaning that x^* is in the kernel of A^T. In other words, $A^T(b - Ax^*) = 0$ or $A^T b = A^T A x^*$. We see:

$$x^* = (A^T A)^{-1} A^T b.$$

10.3. The vector x^* is called the **least square solution** of $Ax = b$ because the $|Ax - b|^2$ is minimized. This is a sum of squares. As a bonus, we get a formula for the orthogonal projection P of $\mathbb{R}^n$ to the m-dimensional space V.

$P = A(A^T A)^{-1} A^T$ is the orthogonal projection onto $\text{im}(A)$.

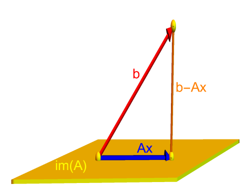

FIGURE 1. $Ax - b$ perpendicular to $\text{im}(A)$ means $Ax - b$ is in $\ker(A^T)$.

10.4. The least square principle is one of the most important paradigms in mathematics. It solves an optimization problem. The use is everywhere. If we write down a wish-list of things we want to have satisfied in the form of a system of linear equations $Ax = b$, we are in a situation where the least square solution is the best possible form. In virtually all problems of science, technology or engineering, we are exposed to a huge amount of data which we want to organize using few parameters. By projecting the data onto the image of a matrix, we get the **best possible solution** for the model at hand.

10.5. In order for the projection formula to work, we needed the formula $(\text{im}(A))^\perp = \ker(A^T)$. We also need to know when A^TA is invertible. This is the case if and only if A does not have a kernel:

Theorem: $\ker(A) = \ker(A^TA)$

Proof. We prove that in class. You should be able to reproduce this proof yourself. □

Examples

10.6. The maximum $f(x) = \max_A \log(\max_{ij} A^{-1}_{ij})$ over all Boolean $x \times x$ matrices A is measured with the computer and gives the data points $(x, y = f(x))$ given by $(3, \log(1))$, $(4, \log(2))$, $(5, \log(3))$, $(6, \log(5))$, $(7, \log(9))$, $(8, \log(18))$, $(9, \log(36))$. How do these data grow? We are not aware that this problem has been asked already so that we have to experiment first. What is the best linear fit? To find the best function $ax + c = y$, we plug in the values x_k, y_k in this equation, to get

$$\begin{array}{rcl} a3+c &=& \log(1) \\ a4+c &=& \log(2) \\ a5+c &=& \log(3) \\ a6+c &=& \log(5) \\ a7+c &=& \log(9) \\ a8+c &=& \log(18) \\ a9+c &=& \log(36) \end{array}, A = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \\ 7 & 1 \\ 8 & 1 \\ 9 & 1 \end{bmatrix}, b = \begin{bmatrix} \log(1) \\ \log(2) \\ \log(3) \\ \log(5) \\ \log(9) \\ \log(18) \\ \log(36) \end{bmatrix}.$$

We compute $(A^TA)^{-1} = \begin{bmatrix} \frac{1}{28} & -\frac{3}{14} \\ -\frac{3}{14} & \frac{10}{7} \end{bmatrix}$ and $A^Tb = [88.6775, 12.0723]^T$. The best parameters are $(a, c) = (-0.595901, 0.580129)$ which corresponds to the line $y = -0.595901x + 0.580129$.

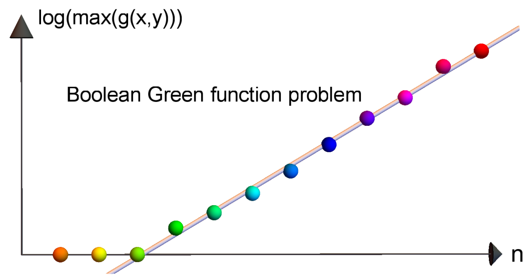

Figure 2. Linear data fitting of the Boolean Green function problem: $\max A^{-1}_{ij}$ among all invertible Boolean $n \times n$ matrices A.

10.7. Problem. Find a quadratic polynomial $p(x) = ux^2 + vx + w$ which best fits the four data points $(-1, 8), (0, 8), (1, 4), (2, 16)$. **Solution.** We write down the equations $ux^2 + vx + w = y$, for every data point (x, y). This gives us a system of four equations

$Ax = b$ with $A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$ $b = \begin{bmatrix} 8 \\ 8 \\ 4 \\ 16 \end{bmatrix}^T$. $A^TA = \begin{bmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{bmatrix}$. The solution is $x^* = (A^TA)^{-1}A^Tb = \begin{bmatrix} 3, -1, 5 \end{bmatrix}^T$.

10.8. Problem. Prove that $\text{im}(A) = \text{im}(AA^T)$.
Solution. The image of AA^T is contained in the image of A because we can write $v = AA^Tx$ as $v = Ay$ with $y = A^Tx$. On the other hand, if v is in the image of A, then $v = Ax$. If $x = y + z$, where y in the kernel of A and z orthogonal to the kernel of A, then $Ax = Az$. Because z is orthogonal to the kernel of A, it is in the image of A^T. Therefore, $z = A^Tu$ and $v = Az = AA^Tu$ is in the image of AA^T.

10.9. Here is a solution with a higher dimensional fitting problem: **Problem.** Which paraboloid $ax^2 + by^2 = z$ best fits the data

x	y	z
0	1	2
-1	0	4
1	-1	3

Solution. Find the least square solution for the system of equations for the unknowns a, b.

$$\begin{aligned} a0 + b\,1 &= 2 \\ a1 + b\,0 &= 4 \\ a1 + b\,1 &= 3 \,. \end{aligned}$$

In matrix form this can be written as $Ax = b$ with

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}.$$

We have $A^TA = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $A^Tb = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$. We get the least square solution with the formula

$$x = (A^TA)^{-1}A^Tb = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

The best fit is the function $\boxed{f(x, y) = 3x^2 + y^2}$. It has as a graph an elliptic paraboloid.

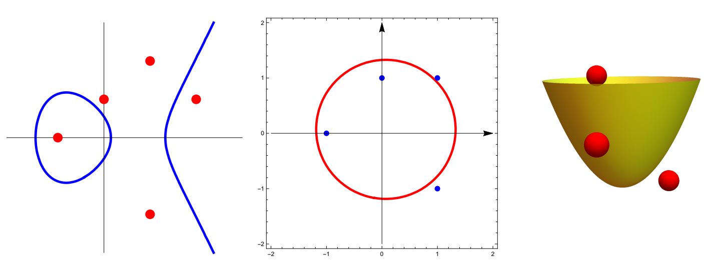

FIGURE 3. We can fit virtually anything. We see a fit with elliptic curves, circles or higher dimensional objects like quadrics.

Homework

This homework is due on Tuesday, 2/27/2019.

Problem 10.1: The first 7 prime numbers $2, 3, 5, 7, 11, 13$ define the data points $(1, 2), (2, 3), (3, 5), (5, 7), (6, 11), (7, 13)$ in the plane. Find the best line $y = ax + b$ which fits these data.

Problem 10.2: a) Find the least square solution x^* of the system $Ax = b$ with $A = \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 2 & -4 \end{bmatrix}$, and $b = \begin{bmatrix} 125 \\ 125 \\ 125 \end{bmatrix}$.
b) What is the matrix P which projects on the image of A?

Problem 10.3: A curve of the form

$$y^2 = x^3 + ax + b$$

is called an **elliptic curve** in Weierstrass form. Elliptic curves are important in cryptography. Use data fitting to find the best parameters (a, b) for an elliptic curve given the following points: $(x_1, y_1) = (1, 2), (x_2, y_2) = (-1, 0), (x_3, y_3) = (2, 1), (x_4, y_4) = (0, 1)$.

Problem 10.4: Find the circle $a(x^2 + y^2) + b(x + y) = 1$ which best fits the data

x	y
0	1
-1	0
1	-1
1	1

In other words, find the least square solution for the system of equations for the unknowns a, b which aims to have all 4 data points (x_i, y_i) on the circle.

Problem 10.5: Let us look at some extreme cases.
a) Analyze the best linear fit $f(x) = ax + b$ for the three data points $(1, 1), (1, 2), (1, 3)$.
b) To find the best linear $f(x) = a + bx$ for the four data points $(1, 1), (1, 1), (2, 2), (4, 7)$ with $f(x) = a + bx$, we try to take a short cut and pick the points $(1, 1), (2, 2), (4, 7)$ instead, as one data point obviously was redundant. Do we get the same solution?