Recitation 11

Tuesday October 19, 2022

1 Recap

1.1 Orthogonal Decomposition Revisited

Given a matrix $A \in \mathbb{R}^{m \times n}$. Any vector $b \in \mathbb{R}^m$ can be uniquely expressed as $b = x + y$ for which $x \in C(A)$ and $y \in N(A^\top)$. In particular, x and y are the orthogonal projections of b onto $C(A)$ and $N(A^\top)$ respectively.
When A is tall ($m \geq n$) and has linearly independent columns, we can write

$$x = A(A^\top A)^{-1}A^\top b \qquad \text{and} \qquad y = (I - A(A^\top A)^{-1}A^\top)b.$$

In other words, the projection matrix onto $C(A)$ is $P = A(A^\top A)^{-1}A^\top$ and onto $N(A^\top) = Q = I - P$.
It is a good exercise to verify that such decomposition makes sense; that is to explain why $(A^\top A)^{-1}$ exists, $x \in C(A)$, $y \in N(A^\top)$, and $b = x + y$.

1.2 Inverses

Given a tall matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and linearly independent columns; i.e. $\text{rank}(A) = n$. Then A has a *left* inverse $(A^\top A)^{-1}A^\top$.

Given a wide matrix $A \in \mathbb{R}^{m \times n}$ with $m \leq n$ and linearly independent rows; i.e. $\text{rank}(A) = m$. Then A has a *right* inverse $A^\top(AA^\top)^{-1}$.

1.3 Overdetermined System

Consider the linear system $Ax = b$, where we have more equations than variables; i.e. A is tall with more rows than columns. The system may not have a solution that satisfies all equations.

Least Squares Approximate Solution: Assume linearly independent columns

1. Orthogonal projection: Project b onto the column space of A, i.e. $\tilde{b} = proj_{C(A)}b = A(A^\top A)^{-1}A^\top b$. Then solve $Ax = \tilde{b} \Rightarrow x = (A^\top A)^{-1}A^\top b$. This vector minimizes the norm of the residual $r = Ax - b$.

2. Left Inverse: Consider the (specific) left inverse $B = (A^\top A)^{-1}A^\top$. Then approximate $x = Bb = (A^\top A)^{-1}A^\top b$.

3. Optimization: Want to find x that minimizes $\|Ax - b\|^2$. Setting gradient to zero gives $x = (A^\top A)^{-1}A^\top b$.

All these approaches give the same result – the least-squares approximate solution is $x_\star = (A^\top A)^{-1}A^\top b$.

1.4 Underdetermined System

More variables than equations; i.e. A has more columns than rows. The system has infinitely many solutions, and we need to pick a specific one.

Minimum Norm Solution: Assuming linearly independent rows, we pick the "smallest" solution, i.e. we minimize $\|x\|^2$ subject to the constraint $Ax = b$. The solution of minimum norm is $x_\star = A^\top (AA^\top)^{-1} b$.

1.5 Regularization

Our goal remains the same: to solve the system $Ax = b$; however, the solution we want now is the one that *minimizes* $T(x) := \|Ax - b\|^2 + \lambda \|x\|^2$ where $\lambda > 0$ is a regularization parameter. The unique optimal solution is given by $x^* = (A^\top A + \lambda I)^{-1} A^\top b$. It can be shown that the inverse is well-defined for all $\lambda > 0$.

2 Exercises

You may use Julia as a computational tool.

1. Consider the function values

$$f(-2) = 0, \quad f(-1) = 0, \quad f(0) = 1, \quad f(1) = 0, \quad f(2) = 0.$$

   (a) Find the straight line $f(t) = C + Dt$ that is closest (in the least squares sense) to these values.

   (b) Find the parabola $f(t) = C + Dt + Et^2$ that is closest (in the least squares sense) to these values. *Hint: Write down the system of equations $A\mathbf{x} = \mathbf{b}$ in three unknowns $x = (C, D, E)$ for the parabola $f(t)$ to go through the points.*

   (c) Find the closest 4th degree polynomial for these points. What is the least squares error?

2. Two points in $\mathbb{R}^3$ have (x, y, z) coordinates as follows.

$$a = (1, 0, 0), \quad b = (0, 1, 1),$$

   (a) Find the plane $z = C + Dx + Ey$ that gives the best fit to the two points a and b that minimizes $C^2 + D^2 + E^2$.

   (b) What is the least squares error?

   (c) Predict the value of z when $(x, y) = (2, -1)$.

3. Tim has a large number of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. He plots the N data points on the xy-plane and finds out that they do not look linear but rather are exponential – that is $y_i \approx ab^{x_i}$ for some constants a, b. Describe a procedure to help Tim determine a reasonable choice of a, b given the values of x_i's and y_i's.

3 Solutions

1. (a) We want to solve an *overdetermined* system $A\mathbf{x} = b$ for which

$$A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ with variables } \mathbf{x} = \begin{bmatrix} C \\ D \end{bmatrix}.$$

The least square answer is given by

$$\begin{bmatrix} C \\ D \end{bmatrix} = \mathbf{x} = (A^\top A)^{-1} A^\top b = \begin{bmatrix} 1/5 \\ 0 \end{bmatrix}$$

which means the closest line is $f(t) = 1/5$.

(b) We want to solve an *overdetermined* system $A\mathbf{x} = b$ for which

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ with variables } \mathbf{x} = \begin{bmatrix} C \\ D \\ E \end{bmatrix}.$$

The least square answer is given by

$$\begin{bmatrix} C \\ D \\ E \end{bmatrix} = \mathbf{x} = (A^\top A)^{-1} A^\top b = \begin{bmatrix} 17/35 \\ 0 \\ -1/7 \end{bmatrix}$$

which means the closest parabola is $f(t) = \frac{17}{35} - \frac{t^2}{7}$.

(c) Suppose that we want to solve for $f(t) = C + Dt + Et^2 + Ft^3 + Gt^4$. We want to solve an *overdetermined* system $A\mathbf{x} = b$ for which

$$A = \begin{bmatrix} 1 & -2 & 4 & -8 & 16 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ with variables } \mathbf{x} = \begin{bmatrix} C \\ D \\ E \\ F \\ G \end{bmatrix}.$$

We note that A is square and invertible, which means that we can solve x exactly and uniquely. By solving $\mathbf{x} = A^{-1}b$, we derive $\mathbf{x} = [1\ 0\ -5/4\ 0\ 1/4]^\top$ which gives $f(t) = 1 - \frac{5}{4}t^2 + \frac{1}{4}t^4$.

Since the system can be solve to the exact, we must have $Ax = b$ which means that the residual $r = Ax - b$ is zero. The least square error is thus $\|r\|^2 = 0$.

(d) Construct the system of equations $A\mathbf{x} = b$ in a similar fashion to previous parts. When the degree is 5, we have 5 equations with 6 variables which make the system *underdetermined*. This means we have infinitely many answers and the smallest answer is given by $\mathbf{x} = A^\top (AA^\top)^{-1} b = [1\ 0\ -5/4\ 0\ 1/4\ 0]^\top$ which gives $f(t) = 1 - \frac{5}{4}t^2 + \frac{1}{4}t^4$.

(e) With degree at least 4, we get the exact fit. On the other hand, the lower the degree is, the more general the best-fit line is. Note that there is no absolute best model/degree for data fitting. In this case, one might argue that linear fitting is the best because every point but $f(1) = 1$ yields value 0 which may lead us into thinking that $f(1) = 1$ is an *outlier*. Others may argue that it is absolutely needed to fit every point (or almost every point) onto the line so they tend to choose higher degree fitting. This; however, may lead to the *overfitting* problem.

2. (a) We have

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and $b = (0, 1)$ we wish to solve $Ax = b$ which minimizes $\|x\|^2$. The solution is $A^T(AA^T)^{-1}b = (1/3, -1/3, 2/3)$

(b) least squares error is 0 because the system is underdetermined.

(c) 2(-1/3) -1(2/3) = -4/3

3. Although the relationship $y_i \approx ab^{x_i}$ is not linear, what we can do is to rearrange it as $\log y_i \approx \log a + x_i \cdot \log b$. This creates a linear relationship between x_i versus $\log y_i$. In other words, we want to solve an *overdetermined* system $A\mathbf{x} = \mathbf{b}$ for which

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_N \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \log y_1 \\ \log y_2 \\ \dots \\ \log y_N \end{bmatrix} \text{ with variables } \mathbf{x} = \begin{bmatrix} \log a \\ \log b \end{bmatrix}.$$

We can solve $\mathbf{x} = (A^\top A)^{-1} A^\top \mathbf{b}$ and use it to calculate a and b as wished.