Math 290-2 Class 5

Wednesday 16th January 2019

The kernel of the transpose

Let A be an $m \times n$ whose columns are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and let $\vec{x}$ be a vector in $\mathbb{R}^n$. Then

$$A^T\vec{x} = \begin{pmatrix} \cdots & \vec{v}_1^T & \cdots \\ \cdots & \vec{v}_2^T & \cdots \\ & \vdots & \\ \cdots & \vec{v}_n^T & \cdots \end{pmatrix} \vec{x} = \begin{pmatrix} v_1^T\vec{x} \\ v_2^T\vec{x} \\ \vdots \\ \vec{v}_n^T\vec{x} \end{pmatrix} = \begin{pmatrix} \vec{v}_1 \cdot \vec{x} \\ \vec{v}_2 \cdot \vec{x} \\ \vdots \\ \vec{v}_n \cdot \vec{x} \end{pmatrix}$$

This means that $\vec{x}$ is in $\ker(A^T)$ if and only if $\vec{x}$ is perpendicular to all of the columns of A. Since the columns of A span its image, it follows that $\boxed{\operatorname{im}(A)^\perp = \ker(A^T)}$.

Least squares and data fitting

In a world without randomness, flipping a coin twice would yield exactly one head, flipping it ten times would yield five heads, and flipping it a hundred times would yield fifty heads. In this world, as the number of flips varies, we would have

$$\begin{pmatrix} \text{number of flips} \\ \text{number of heads} \end{pmatrix} = \begin{pmatrix} k \\ k/2 \end{pmatrix} = \frac{k}{2}\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

In the real world, however, this is not what happens. A trial run might in fact yield:

number of flips	2	10	100
number of heads	0	6	47

There is no line fitting this data: if $y = kx$ were such a line, then we'd have $2k = 0$, $10k = 6$ and $100k = 47$, but this system is inconsistent.

A *least squares* solution of a linear system $A\vec{x} = \vec{b}$ is a vector $\vec{x}^*$ such that $A\vec{x}^*$ is as close to $\vec{b}$ as possible—that is, such that $\|\vec{b} - A\vec{x}^*\| \leqslant \|\vec{b} - A\vec{x}\|$ for all $\vec{x}$ in $\mathbb{R}^n$.

Any least squares solution $\vec{x}^*$ to the system $A\vec{x} = \vec{b}$ must satisfy $A\vec{x}^* = \operatorname{proj}_V(\vec{b})$, where $V = \operatorname{im}(A)$. But then $\vec{b} - A\vec{x}^*$ is in $\operatorname{im}(A)^\perp = \ker(A^T)$, so that $A^T(\vec{b} - A\vec{x}^*) = \vec{0}$, and so $\boxed{A^TA\vec{x}^* = A^T\vec{b}}$. This is called the **normal equation** of the system $A\vec{x} = \vec{b}$. Fun facts:

- The normal equation of a linear system is always consistent.
- If $\ker(A) = \{\vec{0}\}$ then A^TA is invertible, so there is a *unique* least squares solution $\vec{x}^*$ and it is given by $\boxed{\vec{x}^* = (A^TA)^{-1}A^T\vec{b}}$.
- Given a subspace V of $\mathbb{R}^n$ with basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$, the matrix A whose columns are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ has kernel $\{\vec{0}\}$. For any $\vec{b}$ in $\mathbb{R}^n$, the unique least squares solution $\vec{x}^*$ satisfies $A\vec{x}^* = \operatorname{proj}_V(\vec{b})$, and so $\operatorname{proj}_V(\vec{b}) = A\vec{x}^* = A(A^TA)^{-1}A^T\vec{b}$. Thus the matrix for orthogonal projection onto V is $A(A^TA)^{-1}A^T$. [When the vectors $\vec{v}_1, \ldots, \vec{v}_k$ are orthonormal, this simplifies to AA^T.]

1. Find a basis for the orthogonal complement of the subspace of $\mathbb{R}^5$ spanned by the vectors

$$\begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ 4 \\ 1 \\ 2 \\ 3 \end{pmatrix}$$

2. Find a least squares solution to the system $\begin{cases} x + y = 59 \\ x - y = 1 \\ x - 2y = -29 \end{cases}$.

3. A herd of guinea pigs is released on a previously cavy-free island in the South Atlantic ocean. The island is free of predators and has an abundance of hay, so their population is believed to grow exponentially, but it is unknown exactly how fast. Scientists keep a close eye on the herd and count the number of guinea pigs on the island each month for three months. Their findings are summarised in the following table.

time (in months) since release	0	1	2	3
number of guinea pigs	6	25	95	382

Find a function f in a single variable t that approximates how many guinea pigs will be on the island at time t months after their release.

[Hint: we should have $\log f(t) = a + bt$ for some real numbers a and b.]

4. Find a line of best fit for the data $(a_1,b_1),(a_2,b_2),\dots,(a_n,b_n)$, where $a_i \neq a_j$ for some $i \neq j$.