6.5+6.7: Least Squares

Tuesday, October 18

Warmup

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, express $\mathbf{u} \cdot \mathbf{v}$ in terms of matrix algebra (i.e. without using the "dot" symbol).
ANSWER: $\mathbf{u}^T\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$.

If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, give an interpretation in words of the matrix-vector products $A\mathbf{x}$ and $A^T\mathbf{x}$.
ANSWER: $A\mathbf{x}$ is a linear combination of the columns of A given by the coefficients of $\mathbf{x}$. $A^T\mathbf{x}$ is a series of dot products between the columns of A and $\mathbf{x}$, taken independently.

Least Squares

Suppose that $A = QR$ where R is invertible and Q is orthogonal. Show that $(A^TA)^{-1}A^T\mathbf{b} = R^{-1}Q^T\mathbf{b}$.
ANSWER: $(A^TA)^{-1}A^T\mathbf{b} = (R^TQ^TQR)^{-1}R^TQ^T\mathbf{b} = (R^TR)^{-1}R^TQ^T\mathbf{b} = R^{-1}R^{-T}R^TQ^T\mathbf{b} = R^{-1}Q^T\mathbf{b}$.

Show that $A(A^TA)^{-1}A^T\mathbf{b} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}} = \operatorname{proj}_A \mathbf{b}$.
ANSWER: $A(A^TA)^{-1}A^T\mathbf{b} = QR(R^{-1}Q^T\mathbf{b}) = QQ^T\mathbf{b} = \hat{\mathbf{b}}$.

Find an expression for the distance from $\mathbf{b}$ to the span of A in terms of $\mathbf{b}$, and A or $\mathbf{b}$ and Q.
ANSWER: $\|\mathbf{b} - \hat{\mathbf{b}}\| = \|\mathbf{b} - QQ^T\mathbf{b}\| = \|(I - QQ^T)\mathbf{b}\|$.
Also, $\|\mathbf{b} - \hat{\mathbf{b}}\| = \|\mathbf{b} - A(A^TA)^{-1}A^T\mathbf{b}\|$.

More Least Squares

Say we want to find the best least-squares linear approximation $y = mx + b$ for the four points (-6,-1), (-2,2), (1,1), and (7,6). This amounts to solving the least-squares problem

$$\begin{bmatrix} b \\ m \end{bmatrix} = \arg\min_{\mathbf{x}} \left\| \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix} \right\|^2 .$$

Find the least-squares solution using either the normal equations or a QR decomposition, and plot the line and points.
ANSWER: Let's just do it using the normal equations. A is definitely full-rank since the two columns are linearly independent, so $A^T A$ is invertible:

$$\begin{aligned} \begin{bmatrix} b \\ m \end{bmatrix} &= (A^T A)^{-1} A^T \mathbf{b} \\ &= \begin{bmatrix} 4 & 0 \\ 0 & 90 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 45 \end{bmatrix} \\ &= \begin{bmatrix} 3/2 \\ 1/2 \end{bmatrix} . \end{aligned}$$

The optimal line is therefore $y = \frac{1}{2}x + \frac{3}{2}$. Interestingly, this equation was particularly easy to solve because the matrix $A^T A$ was diagonal. This means that the columns of A were orthogonal, which (since one of the columns is the all-ones vector) is equivalent to the data points x_i having mean zero.

Inner Product Spaces

Define an inner product on $\mathbb{P}_2$ by $\langle p, q \rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$. Find the inner products between the vectors $1, t$, and t^2. Find an orthogonal basis for $\mathbb{P}_2$ with respect to this inner product.
ANSWER: $\langle 1, t \rangle = \langle t, t^2 \rangle = 0$. $\langle 1, t^2 \rangle = 2$.
Since $\mathbf{1}$ and t^2 are both orthononal to t any linear combination of them will also be orthongal to t, so to find an orthogonal basis we can pick $\mathbf{1}$, t, and $t^2 + c \cdot \mathbf{1}$ for some constant c such that $0 = \langle \mathbf{1}, t^2 + c \cdot \mathbf{1} \rangle = \langle \mathbf{1}, t^2 \rangle + c\|\mathbf{1}\|^2 = 2 + 3c$. Thus $c = -2/3$ will work.