Data fitting

Aim lecture: Sometimes we cannot solve the lin eqn $A\mathbf{v} = \mathbf{w}$, but nevertheless wish to find the best possible soln $\mathbf{v}$. We use the theory of orthogonal projections to accomplish this.

Motivating example Suppose we know that two physical variables y, t are linearly related say by $y = \alpha + \beta t$ for some unkown $\alpha, \beta \in \mathbb{R}$. We wish to determine α, β experimentally & plot data points $(t_1, y_1), \dots, (t_n, y_n)$ for distinct $t_1, \dots, t_n \in \mathbb{R}$.

If all variables are perfectly measured then α, β can be computed by solving the

Ideal equation for linear relation

The soln to

$$\begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

gives the linear fn $y = \alpha + \beta t$ relating y & t.

Problem Most likely, there will be experimental error so you should take more than 2 data points, & they will not lie on a line so the ideal eqn above has no solution. The question is what is the best soln.

Best approximation

The following gives a new characterisation of orthogonal projections.

Prop-Defn

Let $A \in M_{nm}(\mathbb{R})$ & $X = \text{im } A \leq \mathbb{R}^n$. Let $\mathbf{w} \in \mathbb{R}^n$.

1. As $\mathbf{x}$ varies over X, $\mathbf{x} = \text{proj}_X \mathbf{w}$ is the unique vector which minimises $\|\mathbf{w} - \mathbf{x}\|$. Consequently, $\text{proj}_X \mathbf{w}$ is also called the *best approximation* to $\mathbf{w}$ in X.
2. $\mathbf{x} = \text{proj}_X \mathbf{w}$ is also the unique vector in X which satisfies $A^T \mathbf{x} = A^T \mathbf{w}$.

Proof. 1) is just Pythogoras thm (ex).

DRAW PICTURE

2) $\mathbb{R}^n = X \oplus X^\perp$ so $\mathbf{x} = \text{proj}_X \mathbf{w}$ is the unique vector in X such that $\mathbf{w} - \mathbf{x} \in (\text{im } A)^\perp = (\text{im}\,(A^T)^T)^\perp$. The propn on orthogonal complements of kernels in lecture 36 $\implies (\text{im } A)^\perp = \ker A^T$. Hence $\mathbf{x} = \text{proj}_X \mathbf{w}$ is the unique vector in X such that $A^T(\mathbf{w} - \mathbf{x}) = \mathbf{0}$. This gives 2).

Least squares solution

Cor-Defn

Let $A \in M_{nm}(\mathbb{R})$ & $\mathbf{w} \in \mathbb{R}^n$. A *least squares* solution to $A\mathbf{v} = \mathbf{w}$ is any $\mathbf{v} \in \mathbb{R}^m$ which minimises the error term $\|A\mathbf{v} - \mathbf{w}\|$. Equivalently, it is any $\mathbf{v} \in \mathbb{R}^m$ such that

1. $A\mathbf{v}$ is a best approximation to $\mathbf{w}$ in im A.
2. it is a solution to the *normal equation* $A^T A\mathbf{v} = A^T\mathbf{w}$.

If rank$A = m$ (i.e. the columns are lin indep) then $A^T A$ is invertible so the least squares soln is given uniquely as $\mathbf{v} = (A^T A)^{-1} A^T \mathbf{w}$. In particular, we obtain the projection formula

$$\text{proj}_{\text{im } A}\mathbf{w} = A(A^T A)^{-1} A^T \mathbf{w}.$$

Proof. 1) & 2) follow from the previous propn since $A\mathbf{v}$ varies over im A as $\mathbf{v}$ varies over $\mathbb{R}^m$.

Suppose now that $m = \text{rank}A = \text{rank}A^T A$ by propn lecture 36, so $A^T A \in M_{mm}(\mathbb{R})$ is invertible. The projection formula follows from the fact that $\text{proj}_{\text{im } A}\mathbf{w} = A\mathbf{v}$ where $\mathbf{v}$ is any least squares soln.

Example

E.g. Find the (least squares) line of best fit $y = \alpha + \beta t$ to the data points $(t, y) = (1, 5), (2, 3), (3, 3), (4, 0)$.

A We wish to find the least squares soln to the ideal eqn on page 1.

Example cont'd

Curves of best fit

One can further ask what is the a parabola of best fit $y = \alpha + \beta t + \gamma t^2$ to a set of data points & so forth. The general setup is as follows.

Setup Consider data points $(t_1, y_1), \dots, (t_n, y_n)$ for distinct $t_1, \dots, t_n \in \mathbb{R}$. Let $\phi_1(t), \dots, \phi_m(t)$ be $\mathbb{R}$-valued fns of t.

Curve of best fit

Suppose $y(t)$ has the form $y(t) = \sum_{i=1}^{m} \beta_i \phi_i(t)$ for some parameters $\mathbf{v} = (\beta_1, \dots, \beta_m)^T \in \mathbb{R}^m$. Let $\mathbf{w} = (y_1, \dots, y_n)^T$. The curve of best fit (wrt linear combns of $\phi_1, \dots, \phi_m$) is given by the least squares soln to $A\mathbf{v} = \mathbf{w}$ where $A = (\phi_j(t_i))_{ij} \in M_{nm}(\mathbb{R})$.

E.g. Find the quadratic function form $y = \beta_1 t + \beta_2 t^2$ which best fits $(-1, 1), (1, 1), (2, 1)$.
We wish to best solve

Example cont'd

An application of QR-factorisation

QR-factorisation is frequently used in numerical linear algebra (i.e. numerical algorithms that computers can implement to solve matrix problems).

Numerical Problem Suppose that $A \in M_{mn}(\mathbb{R})$ is a matrix whose columns are almost linearly dependent. For example

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1+\varepsilon \end{pmatrix}$$

for some small $\varepsilon > 0$. Then $A^T A$ will be close to non-invertible in the sense that $|\det(A^T A)|$ is small. By Cramer's rule, small round-off errors in computing $A^T A$ may lead to large errors in solving the normal eqn $A^T A\mathbf{v} = A^T \mathbf{w}$. In our example above

$$A^T A = \begin{pmatrix} 3 & 3+\varepsilon \\ 3+\varepsilon & 3+2\varepsilon+\varepsilon^2 \end{pmatrix}$$

has determinant $2\varepsilon^2$. Here if your computer rounded off ε^2 to 0, the determinant changes to $-\varepsilon^2$ so the computer error in solving the normal eqns is roughly a factor of 2!

Application cont'd

Question How can you get around this?

Lemma

Let $Q = (\mathbf{w}_1 \dots \mathbf{w}_n) \in M_{mn}(\mathbb{C})$ *be an* $m \times n$*-matrix with columns* $\mathbf{w}_1, \dots, \mathbf{w}_n \neq \mathbf{0}$*.*

1. *The columns are orthogonal iff* Q^*Q *is diagonal.*
2. *In this case,* Q^*Q *is the diagonal matrix* $D = (\|\mathbf{w}_1\|^2) \oplus \dots \oplus (\|\mathbf{w}_n\|^2)$ *so* Q *has a "left inverse"* $Q' = D^{-1}Q^*$ *i.e.* $Q'Q = I_n$*.*
3. $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ *is orthonormal iff* $Q^*Q = I_m$*.*

Proof. We think of Q as a $1 \times n$-matrix with entries $\mathbf{w}_1, \dots, \mathbf{w}_n \in M_{m1}(\mathbb{C})$. We may also view Q^* as an $n \times 1$-matrix with entries $\mathbf{w}_1^*, \dots, \mathbf{w}_n^* \in M_{1m}(\mathbb{C})$. We then multiply these matrices of matrices to see the (i,j)-th entry of Q^*Q is $\mathbf{w}_i^*\mathbf{w}_j = (\mathbf{w}_i|\mathbf{w}_j)$. All results now follow.

A QR-factorisations sometimes helps as follows. Suppose $A = QR$ is the QR-factorisation. Then the normal eqn is

$$R^TQ^TQR\mathbf{v} = R^TQ^T\mathbf{w} \iff R^TR\mathbf{v} = R^TQ^T\mathbf{w}$$

Also, R^T is invertible so this reduces to the better behaved $R\mathbf{v} = Q^T\mathbf{w}$ since $\det(R) = \sqrt{\det(A^TA)}$.