ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

IFOR
Institute for Operations Research
Department of Mathematics

Robert Weismantel

Week 9: An explicit formula for projections, application of projections to data fitting and orthonormal bases

Recall from the previous lecture:

Let S be a subspace in $\mathbb{R}^m$ generated by $a_1, \ldots, a_n \in S$, i.e.,

$$S = \text{ span}(a_1, \ldots, a_n) = C(A) = \{A\lambda \mid \lambda \in \mathbb{R}^n\}$$

where

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}.$$

Lemma

The projection of a vector $b \in \mathbb{R}^m$ to the subspace $S = C(A)$ can be written as

$$\text{proj}_S(b) = A\hat{x}, \text{ where } \hat{x} \text{ satisfies the normal equations } A^T A\hat{x} = A^T b.$$

Can we derive an explicit formula for projections?

Lemma

For a matrix $A \in \mathbb{R}^{m \times n}$ we have that $A^T A$ is invertible if and only if A has linearly independent columns.

Proof.

$A^T A$ is invertible if and only if $N(A^T A) = \{0\}$. We know that $N(A^T A) = N(A)$ and hence, the result follows.

Theorem

Let S be a subspace in $\mathbb{R}^m$ and A a matrix whose columns are a basis of S. The projection of $b \in \mathbb{R}^m$ to S is given by

$$\text{proj}_S(b) = Pb,$$

where $P = A\left(A^\top A\right)^{-1} A^\top$ is the projection matrix.

A few coments about projections I

Consider the projection matrix $P = A\left(A^\top A\right)^{-1} A^\top$.

- The matrix A (and $A^\top$) are not necessarily square, and so they don't have inverses.
- Hence, the expression $A\left(A^\top A\right)^{-1} A^\top$ **cannot** be simplified by expanding $\left(A^\top A\right)^{-1}$.
- This would yield $I = P$ and would only make sense if S was all of $\mathbb{R}^m$. This case is less interesting as it means that A is invertible.
- P can be viewed as a mapping: for a given vector b its projection is given by $\mathrm{proj}_S(b) = Pb$.

A few coments about projections II

Consider the projection matrix $P = A\left(A^\top A\right)^{-1} A^\top$.

- If $b \in \mathbb{R}^m$, then $\mathrm{proj}_S(\mathrm{proj}_s(b)) = \mathrm{proj}_S(b)$ by definition. This requires us to have that $PPb = Pb$, i.e., we should have $P^2 = P$. Indeed

$$P^2 = \left(A\left(A^\top A\right)^{-1} A^\top\right)^2 = A\left(A^\top A\right)^{-1} A^\top A\left(A^\top A\right)^{-1} A^\top = P.$$

- Let $S^\perp$ be the orthogonal complement of S and P the projection matrix onto the subspace S, i.e., $\mathrm{proj}_S(b) = Pb$. Then $I - P$ is the projection matrix that maps $b \in \mathbb{R}^m$ to $\mathrm{proj}_{S^\perp}(b)$. This follows since $b = e + \mathrm{proj}_S(b) = e + Pb$ where $e \in S^\perp$. Hence,

$$(I - P)b = b - Pb = e = \mathrm{proj}_{S^\perp}(b).$$

- Note that – as it should be – we have that $(I - P)^2 = I - 2P + P^2 = I - P$.

Another name for projections: least squares

Definition

For $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ a least squares solution solves

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|^2.$$

The link to projections

Consider the subspace $C(A) = \{Ax \mid x \in \mathbb{R}^n\}$. Then,

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|^2 = \min_{p\in C(A)} \|b - p\|^2 = \|b - \mathrm{proj}_{C(A)}(b)\|^2.$$

Remark

- A least squares solution is given by $\mathrm{proj}_{C(A)}(b) = A\hat{x}$, where $A^T A\hat{x} = A^T b$.
- If A has linearly independent columns, then $A^T A$ is invertible. Hence, for the least squares solution we have the explicit formula

An application of least squares

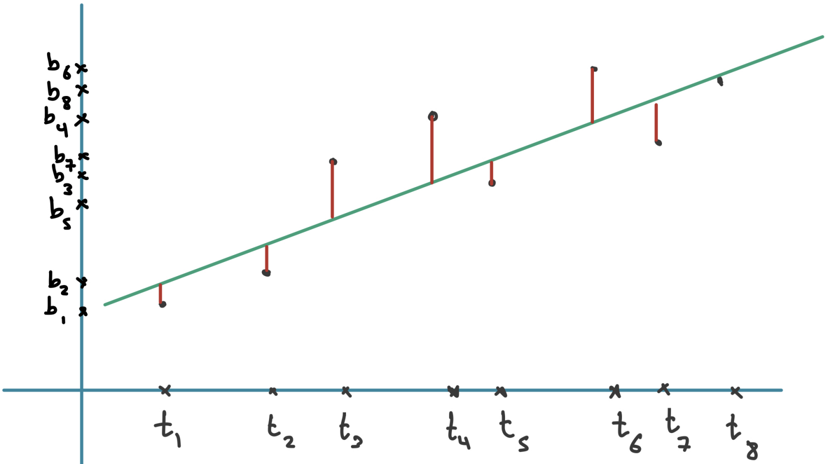

Figure: Fitting a line to points

Linear regression

- is the task to fit a line through data points.
- Consider data points

$$(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m),$$

  representing some attribute b over time t.
- If the relation between t and b is explained by a linear relationship then it makes sense to search for constants $\alpha_0 \in \mathbb{R}$ and $\alpha_1 \in \mathbb{R}$ such that

$$b_k \approx \alpha_0 + \alpha_1 t_k.$$

Find α_0 and α_1 minimizing the sum of squares of the errors

$$\min_{\alpha_0, \alpha_1} \sum^{m} (b_k - [\alpha_0 + \alpha_1 t_k])^2.$$

An explicit formula under one assumption.

In matrix-vector notation

$$\min_{\alpha_0,\alpha_1} \left\| b - A \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \right\|^2, \tag{1}$$

where

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \\ b_m \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_{m-1} \\ 1 & t_m \end{bmatrix}.$$

If A has independent columns, the solution is

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = (A^\top A)^{-1} A^\top b = \begin{bmatrix} m & \sum_{k=1}^m t_k \\ \sum_{k=1}^m t_k & \sum_{k=1}^m t_k^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^m b_k \\ \sum_{k=1}^m t_k b_k \end{bmatrix}$$

We can assume that A has independent columns

Lemma

The columns of the $m \times 2$ matrix A defined before are linearly dependent if and only if $t_i = t_j$ for all $i \neq j$.

Proof.

- Suppose that there are two indices $i \neq j$ such that $t_i \neq t_j$. Let $\mathbf{1}$ be the all ones-vector in $\mathbb{R}^m$ and t the vector with components $t_1, \ldots, t_m$.
- Consider the system in variables λ, μ

$$\lambda \mathbf{1} + \mu t = 0.$$

- Since $t_i \neq t_j$ we can subtract row j from row i to obtain

$$\lambda 0 + \mu (t_i - t_j) = 0 \iff \mu = 0 \text{ since } t_i - t_j \neq 0.$$

- This implies that $\lambda = 0$ and hence A has full column rank

Fitting a parabola: work it out!

- If we believe the relationship between t_k and b_k is quadratic we could attempt to fit a parabola:

$$b_k \approx \alpha_0 + \alpha_1 t_k + \alpha_2 t_k^2.$$

- This is a linear function in α_0, α_1, and α_2. Similarly as with linear regression, it is natural to attempt to minimze

$$\min_{\alpha_0,\alpha_1,\alpha_2} \left\| b - A \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \right\|^2. \tag{2}$$

- $$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \\ b_m \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \\ 1 & t_{m-1} & t_{m-1}^2 \\ 1 & t_m & t_m^2 \end{bmatrix}.$$

Orthonormal vectors

Definition (Orthonormal vectors)

$q_1, \ldots, q_n \in \mathbb{R}^m$ are orthonormal if they are orthogonal and have norm 1, i.e.,

$$q_i^T q_j = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases},$$

where δ_{ij} is the Kronecker delta.

Remark and example

- If Q is the matrix whose columns are the vectors q_i's, then the condition that the vectors are orthonormal can be rewritten as $Q^\top Q = I$.
- Q may not be a square matrix, and so it is not necessarily the case that $QQ^\top = I$.
- A classical example of an orthonormal set of vectors is the canonical basis, $e^1, \ldots, e^n \in \mathbb{R}^n$ where e^i is the vector with a 1 in

Orthogonal matrices

Definition (Orthogonal matrix)

A square matrix $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix when $Q^\top Q = I$. In this case, $QQ^\top = I$, $Q^{-1} = Q^\top$, and the columns of Q form an orthonormal basis for $\mathbb{R}^n$.

Example

The 2×2 matrix Q that corresponds to rotating, counterclockwise, the plane by θ,

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is an orthogonal matrix. Indeed,

$$R_\theta^T R_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = I.$$

Permutation matrices are orthogonal matrices

Definition

A permutation is a bijective map

$$\pi : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}, \text{ i.e., } \pi(i) \neq \pi(j) \text{ for } i \neq j.$$

The permutation matrix $A \in \mathbb{R}^{n \times n}$ associated with π has entries $A_{ij} = 1$ if $\pi(i) = j$ and $A_{ij} = 0$, otherwise.

Example

Permutation matrices are another example of orthogonal matrices. Indeed, A^T is the permutation matrix associated with the permutation σ defined as $\sigma(j) = i$ whenever $\pi(i) = j$. Hence, $A^T A = I$, i.e., A is an orthogonal matrix.

Challenge

For every permutation matrix A there exists a positive integer k such that $A^k = I$.

A first observation about orthogonal matrices

Proposition

Orthogonal matrices preserve norm and inner product of vectors. In other words, if $Q \in \mathbb{R}^{n\times n}$ is orthogonal then, for all $x, y \in \mathbb{R}^n$

$$\|Qx\| = \|x\| \text{ and } (Qx)^\top(Qy) = x^\top y$$

Proof of the second inequality.

$$\text{For } x, y \in \mathbb{R}^n, \quad (Qx)^\top(Qy) = x^\top Q^\top Qy = x^\top Iy = x^\top y.$$

Proof of the first equality.

For $x \in \mathbb{R}^n$ we have that $\|Qx\| \geq 0$ and $\|x\| \geq 0$. Then

$$\|Qx\|^2 = (Qx)^\top(Qx) = x^\top x = \|x\|^2 \quad \Rightarrow \quad \|Qx\| = \|x\|.$$

Projections with Orthonormal Basis

The message here

An access to an orthonormal basis simplifies calculations for projections.

Observation

Let S be a subspace of $\mathbb{R}^m$ and $q_1, \ldots, q_n$ an orthonormal basis for S. Let

$$Q = \left[\begin{array}{ccccc} q_1 & , & \cdots & , & q_n \end{array} \right] \in \mathbb{R}^{m \times n}.$$

The projection matrix that projects to S is given by $QQ^\top$ and the least squares solution attaining $\min_{x \in \mathbb{R}^n} \|Qx - b\|^2$ is given by $\hat{x} = Q^\top b$.

Remark

When Q is square then $QQ^\top$ is simply the identity corresponding to projecting to $\mathbb{R}^n$. For $x \in \mathbb{R}^n$ it writes it a linear combination of the orthonormal basis.