Robert Weismantel

Summary Lecture 17: A formula for projections and data fitting

An explicit formula for projections:

Lemma 5.2.4

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.

Theorem 5.2.5

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 called the projection matrix.

Notice,

- $A\left(A^\top A\right)^{-1} A^\top$ **cannot** be simplified by expanding $\left(A^\top A\right)^{-1}$.
- P defines a map: Given vector b its projection to S is $\text{proj}_S(b) = Pb$.

A first application: least squares for data fitting

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

$$\hat{x} = (A^T A)^{-1} A^T b.$$

Least squares for linear regression

Linear regression

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

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

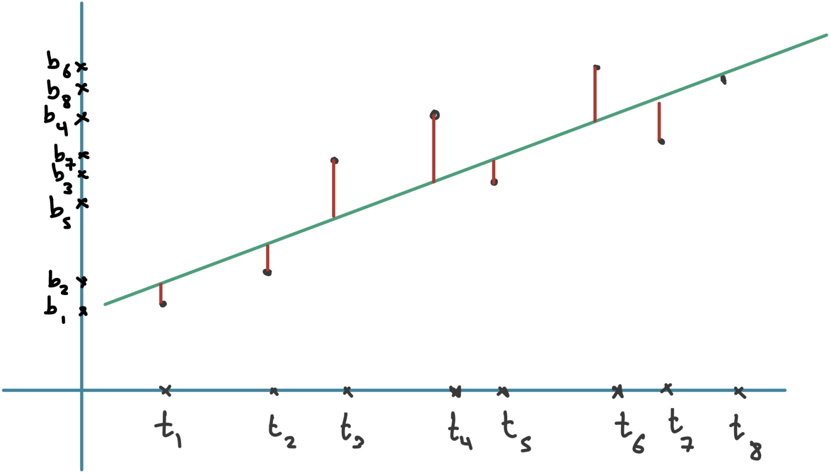
  representing some attribute b over time t.
- If the relation between t and b is explained by a linear relationship then it makes sense to search for constants $\alpha_0 \in \mathbb{R}$ and $\alpha_1 \in \mathbb{R}$ such that

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

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

$$\min_{\alpha_0, \alpha_1} \sum_{k=1}^{m} (b_k - [\alpha_0 + \alpha_1 t_k])^2.$$

Fitting a line to points



An explicit formula under one assumption.

In matrix-vector notation

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

where

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

We can assume that A has independent columns to get

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