Basics of Linear Models

Steve Avsec

Illinois Institute of Technology

January 22, 2024

Overview

1. Some Setup
2. Linear Regression
3. A Detour

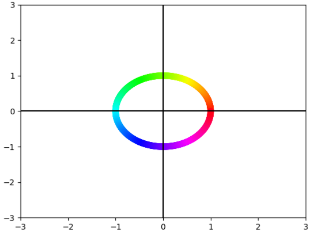
System from Last Time

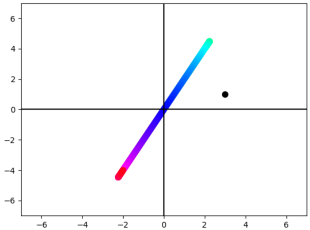
Look at $Ax = b$ where

$$A = \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix}$$

and

$$b = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$





Orthogonal projection

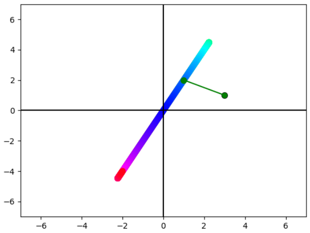


Figure: Orthogonal Projection onto Column Space

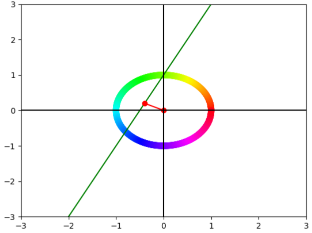


Figure: Orthogonal Projection onto Preimage

The Calculation

Singular Value Decomposition of A:

$$A = usv^t$$

where u, v are orthogonal, and s is a diagonal with 5 in the upper right and 0 in the lower left corners

The Calculation

Singular Value Decomposition of A:

$$A = usv^t$$

where u, v are orthogonal, and s is a diagonal with 5 in the upper right and 0 in the lower left corners

The "pseudoinverse" of A:

$$A^\dagger = vs^\dagger u^t$$

where $s^\dagger$ has $1/5$ in the upper right and 0 in the lower left.

The Calculation

Singular Value Decomposition of A:

$$A = usv^t$$

where u, v are orthogonal, and s is a diagonal with 5 in the upper right and 0 in the lower left corners

The "pseudoinverse" of A:

$$A^\dagger = vs^\dagger u^t$$

where $s^\dagger$ has $1/5$ in the upper right and 0 in the lower left.

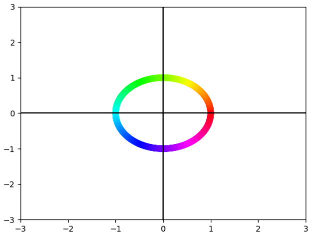
$$A^\dagger b = \left[\begin{array}{c} -0.4 \\ 0.2 \end{array} \right]$$

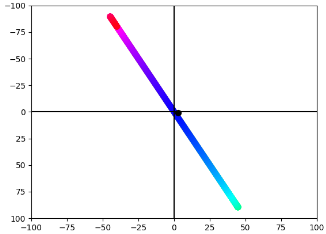
Another System

$$B = \begin{bmatrix} -40.0004 & 19.9992 \\ -79.9998 & 40.0004 \end{bmatrix}$$

and

$$b = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$





Condition Numbers

The singular values of B are 100 and 0.001.

Condition Numbers

The singular values of B are 100 and 0.001.

The *condition number* of a matrix is the ratio between its largest and smallest singular values.

Condition Numbers

The singular values of B are 100 and 0.001.

The *condition number* of a matrix is the ratio between its largest and smallest singular values.

A matrix is *poorly conditioned* if its condition number is large (e.g. larger than machine precision)

The Basic System

Suppose we have some data $(x_1, y_1), \ldots, (x_N, y_N)$ (all scalars).

The Basic System

Suppose we have some data $(x_1, y_1), \dots, (x_N, y_N)$ (all scalars).

Suppose we suspect that the $\{x_j\}$ is linearly related to the $\{y_j\}$, so there are reasonable numbers m and b such that

$$y_j = mx_j + b + \varepsilon_j$$

The Basic System

Suppose we have some data $(x_1, y_1), \ldots, (x_N, y_N)$ (all scalars).

Suppose we suspect that the $\{x_j\}$ is linearly related to the $\{y_j\}$, so there are reasonable numbers m and b such that

$$y_j = mx_j + b + \varepsilon_j$$

The $\{\varepsilon_j\}$ are the errors of our model which is a longer discussion.

The Basic System Cont.

Now we can rewrite this system as $Xr = y$ where

$$X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_N & 1 \end{bmatrix}$$

and

$$r = \begin{bmatrix} m \\ b \end{bmatrix}$$

The Basic System Cont.

Now we can rewrite this system as $Xr = y$ where

$$X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_N & 1 \end{bmatrix}$$

and

$$r = \begin{bmatrix} m \\ b \end{bmatrix}$$

Taking the pseudo-inverse,

$$r = (X^t X)^{-1} X^t y$$

Going Up in Dimension

Suppose we have some data $(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)$ (now the $\mathbf{x_j}$ are p-dimensional vectors).

Going Up in Dimension

Suppose we have some data $(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)$ (now the $\mathbf{x_j}$ are p-dimensional vectors).

Now our system $Xc = y$ looks like

$$X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ & \vdots & \vdots & \\ 1 & x_{N,1} & \cdots & x_{N,p} \end{bmatrix}$$

Up A Dimension Cont.

$$c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{bmatrix}$$

Up A Dimension Cont.

$$c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{bmatrix}$$

The *same* solution works

$$c = (X^t X)^{-1} X^t y$$

Up A Dimension Cont.

$$c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{bmatrix}$$

The *same* solution works

$$c = (X^t X)^{-1} X^t y$$

BUT there is a much greater chance that the matrix X will be poorly conditioned (curse of dimensionality).

A Reminder

A *matrix* is a *representation* of a linear transformation $T : V \to W$ given bases of V and W.

A Reminder

A *matrix* is a *representation* of a linear transformation $T : V \to W$ given bases of V and W.

A complicated example: Let $V = W = P(x)$ where $P(x)$ is the vector space of all polynomials (with real numbers as coefficients). Define T by

$$T(p)(x) = \int_{-\infty}^{\infty} p(y) e^{-\frac{(x-y)^2}{2}} dy$$

Some Functions

Let $f_j : \mathbb{R}^p \to \mathbb{R}$ be the "coordinate" functions:

$$f_j(\mathbf{x}) = x_j$$

and f_0 be the constant function 1 (i.e. $f_0(\mathbf{x}) = 1$).

Some Functions

Let $f_j : \mathbb{R}^p \to \mathbb{R}$ be the "coordinate" functions:

$$f_j(\mathbf{x}) = x_j$$

and f_0 be the constant function 1 (i.e. $f_0(\mathbf{x}) = 1$).

Now we can reframe: Which linear combination

$$c_0 f_0 + c_1 f_1 + \cdots c_p f_p$$

best fits our data?

Some Extensions

We can use any other set of functions that satisfy our problem:

Some Extensions

We can use any other set of functions that satisfy our problem:

- $$f_\alpha(\mathbf{x}) = \mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_p^{\alpha_p}$$
- $$f_m(x) = \sin(mx) \text{and} g_n(x) = \cos(nx)$$
- Let $K : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ be a function with the following properties:
  - Symmetry: $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{y})$
  - Positive-definiteness: For all $\{\mathbf{x}_j\} \in \mathbb{R}^p$ and all $\{c_j\} \in \mathbb{R}$:

$$\sum_{i,j} c_i c_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

Some Examples

- $K(\mathbf{x},\mathbf{y}) = \langle x, y\rangle$
- $$K(\mathbf{x},\mathbf{y}) = e^{-L\|\mathbf{x}-\mathbf{y}\|_2^2} \text{for} L > 0$$
- $$K(\mathbf{x},\mathbf{y}) = e^{-L\|\mathbf{x}-\mathbf{y}\|_2} \text{for} L > 0$$