Linear Regression

CS6140

Predrag Radivojac

Khoury College of Computer Sciences

Northeastern University

Spring 2021

Regression

Given: a set of observations $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, $(\boldsymbol{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y}$

Objective: find best approximator $f(\boldsymbol{x}) \in \mathcal{Y}$, where $f \in \mathcal{F}$

Example: $\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \mathbb{R}$

- take $f(\boldsymbol{x}) = \alpha + x_1 x_2^\beta$ ← nonlinear regression
- find α and β from data

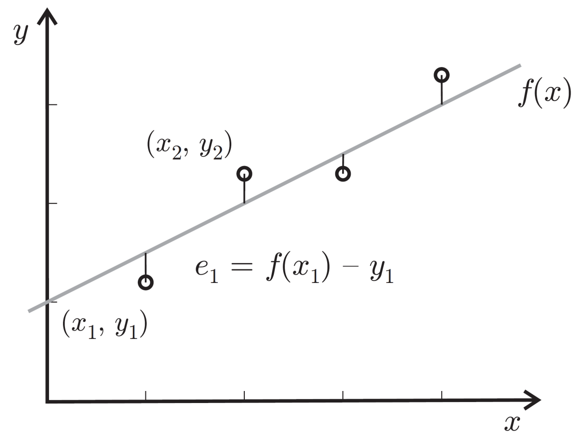
- take $f(\boldsymbol{x}) = w_0 + w_1 x_1 + w_2 x_2$ ← linear regression
- find w_0, w_1 and w_2 from data

LINEAR REGRESSION

Given: a set of observations $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$

Objective: find best linear approximator $f(\boldsymbol{x}) = w_0 + \sum_{j=1}^d w_j x_j$

$\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}$



BASIC FORMULATION

Given: a set of observations $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$

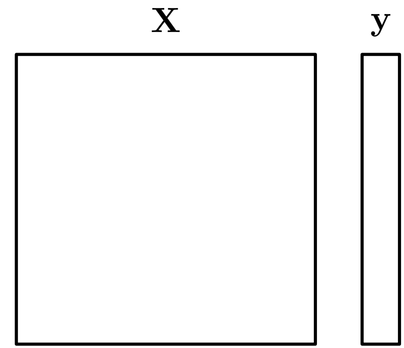
Goal: minimize sum of squares $\sum_{i=1}^{n}(f(\boldsymbol{x}_i) - y_i)^2$

Rewrite: minimize sum of squares $\sum_{i=1}^{n}(w_0 + \sum_{j=1}^{d} w_j x_{ij} - y_i)^2$

Derive: optimal coefficients $(w_0, w_1, \ldots, w_d)$

BASIC FORMULATION: VECTOR NOTATION

$$\boldsymbol{w} = (w_0, w_1, \ldots, w_d) \qquad \rightarrow \qquad \mathbf{w} = [w_0 \;\; w_1 \;\; \ldots \;\; w_d]^T$$
$$\boldsymbol{x} = (x_0 = 1, x_1, \ldots, x_d) \qquad\qquad \mathbf{x} = [x_0 = 1 \;\; x_1 \;\; \ldots \;\; x_d]^T$$



Reformulate:

Given: matrix $\mathbf{X}$ and vector $\mathbf{y}$

Goal: minimize $(\mathbf{Xw} - \mathbf{y})^T(\mathbf{Xw} - \mathbf{y})$

SUMMARY

Given: a set of observations $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$

Goal: minimize sum of squares $\sum_{i=1}^n (w_0 + \sum_{j=1}^d w_j x_{ij} - y_i)^2$

Use vector notation

Given: matrix $\mathbf{X}$ and vector $\mathbf{y}$

Goal: minimize $(\mathbf{Xw} - \mathbf{y})^T (\mathbf{Xw} - \mathbf{y})$

Solve

$$\mathbf{w}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

MAXIMUM LIKELIHOOD ESTIMATION

Assume: $Y = f(\boldsymbol{X}|\boldsymbol{\omega}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and f is a linear combination of $\boldsymbol{X}$ and $\boldsymbol{\omega}$

This gives: $Y|\boldsymbol{x}, \boldsymbol{\omega} \sim \mathcal{N}(\sum_{j=0}^{d} \omega_j x_j, \sigma^2)$ $\qquad p(y|\boldsymbol{x}, \boldsymbol{\omega}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \sum_{j=0}^{d} \omega_j x_j)^2}{2\sigma^2}}$

Maximum likelihood estimation

Assume: $Y = f(\boldsymbol{X}|\boldsymbol{\omega}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and f is a linear combination of $\boldsymbol{X}$ and $\boldsymbol{\omega}$

This gives: $Y|\boldsymbol{x}, \boldsymbol{\omega} \sim \mathcal{N}(\sum_{j=0}^{d} \omega_j x_j, \sigma^2)$ $\quad p(y|\boldsymbol{x}, \boldsymbol{\omega}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \sum_{j=0}^{d} \omega_j x_j)^2}{2\sigma^2}}$

Assume: data set $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$ drawn i.i.d.

Likelihood: $p(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{w}) = \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(y_i - \sum_{j=0}^{d} w_j x_{ij})^2}{2\sigma^2}}$

Log-likelihood: $\log p(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{w}) = \sum_{i=1}^{n} \log p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = n \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\sum_{i=1}^{n}(y_i - \sum_{j=0}^{d} w_j x_{ij})^2}{2\sigma^2}$

SUMMARY

Assume: $Y = f(\boldsymbol{X}|\boldsymbol{\omega}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and f is a linear combination of $\boldsymbol{X}$ and $\boldsymbol{\omega}$

Likelihood: $\prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(y_i - \sum_{j=0}^{d} w_j x_{ij})^2}{2\sigma^2}}$

Maximize likelihood = minimize sum of squared errors

$$\mathbf{w}_{\mathrm{ML}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

ALGEBRAIC VIEW

Consider: system $\mathbf{Ax} = \mathbf{b}$

Example:

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ x_1 + 3x_2 &= 5 \end{aligned}$$

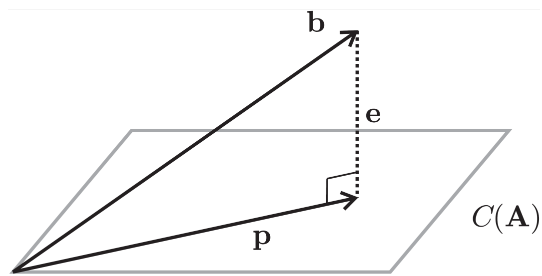
$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 3 \end{bmatrix} x_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

Algebraic View

Given: matrix $\mathbf{A}$ and vector $\mathbf{b}$

Goal: find $\mathbf{x}$ to minimize $||\mathbf{Ax} - \mathbf{b}||_2$



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