Linear Regression

Aarti Singh

Machine Learning 10-315
Feb 7, 2022

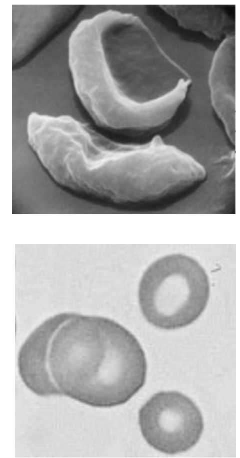
Supervised Learning Tasks

Classification

Sports
Science
News

X = Document

Y = Topic

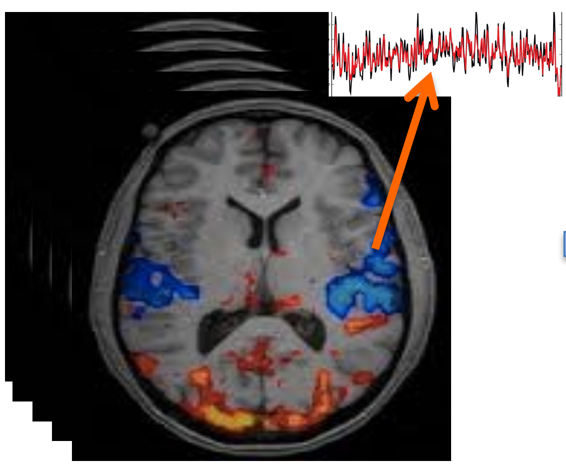


Anemic cell
Healthy cell

X = Cell Image

Y = Diagnosis

Regression

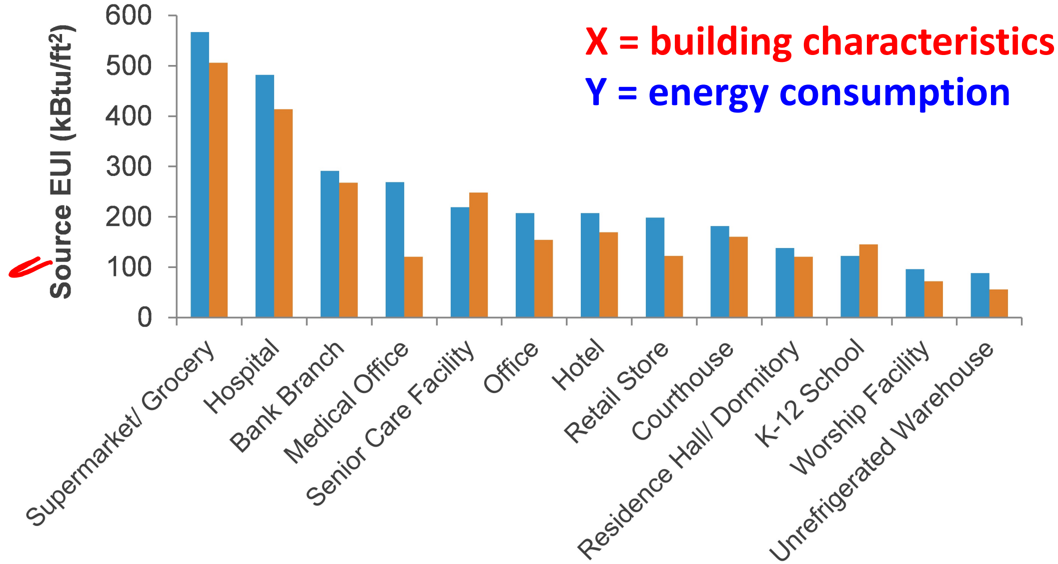


Y = Age of a subject

X = Brain Scan

Regression Tasks

Estimating Energy Usage



energystar.gov

Estimating Contamination

Mean Squared Error (MSE) Minimization

$(f(X) - Y)^2$

Optimal predictor: $f^* = \arg\min_f \mathbb{E}[(f(X) - Y)^2]$ $= E[Y|X]$

Empirical Minimizer: $\widehat{f}_n = \arg\min_{\boxed{f \in \mathcal{F}}} \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - Y_i)^2$

Empirical mean

Law of Large Numbers:

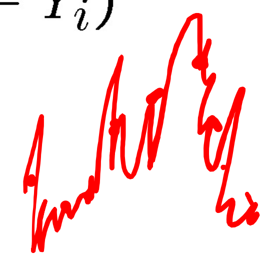
$$\boxed{\frac{1}{n}\sum_{i=1}^{n}[\text{loss}(Y_i, f(X_i))] \xrightarrow{n \to \infty} \mathbb{E}_{XY}[\text{loss}(Y, f(X))]}$$

Restrict class of predictors

Optimal predictor: $$f^* = \arg\min_f \mathbb{E}[(f(X) - Y)^2]$$

Empirical Minimizer: $$\hat{f}_n = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Y_i)^2$$

Class of predictors



- Why?

$\mathcal{F}$ - Class of Linear functions

- Class of Polynomial functions
- Class of nonlinear functions

Linear Regression

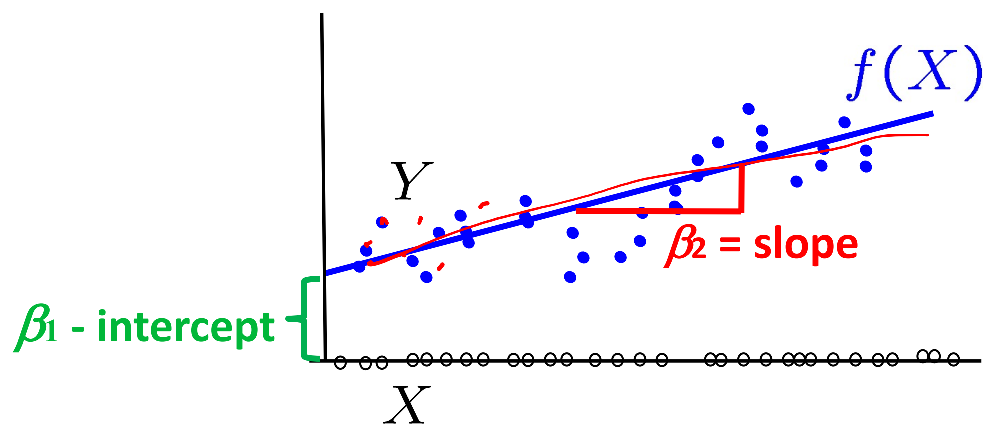
$f(x_i) = \sum_j \beta_j x_i^{(j)}$

$$\hat{f}_n^L = \arg\min_{f\in\mathcal{F}_L} \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - Y_i)^2$$ Least Squares Estimator

$\mathcal{F}_L$ - Class of Linear functions

Uni-variate case:

$$f(X) = \beta_1 + \beta_2 X$$



Multi-variate case:

$$\boxed{f(X)} = f(X^{(1)}, \ldots, X^{(p)}) = \beta_1 \cancelto{1}{X^{(1)}} + \beta_2 X^{(2)} + \cdots + \beta_p X^{(p)}$$

$$= X\beta \qquad \text{where} \quad X = [X^{(1)} \ldots X^{(p)}], \quad \beta = [\beta_1 \ldots \beta_p]^T$$

Linear Regression (Matrix-vector form)

$$\widehat{f}_n^L = \arg\min_{f \in \mathcal{F}_L} \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Y_i)^2$$

$$f(X_i) = X_i\beta = \begin{bmatrix} 1 & X_i^{(2)} & \cdots & X_i^{(p)} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$\widehat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (X_i\beta - Y_i)^2$$

$$\widehat{f}_n^L(X) = X\widehat{\beta}$$

test point

$$= \arg\min_{\beta} \frac{1}{n} (\boxed{\mathbf{A}\beta - \mathbf{Y}})^T (\mathbf{A}\beta - \mathbf{Y})$$

$$A\beta = \begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{bmatrix}_{n \times 1}$$

$$\underset{n \times p}{\mathbf{A}} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_1^{(1)} & \dots & X_1^{(p)} \\ \vdots & \ddots & \vdots \\ X_n^{(1)} & \dots & X_n^{(p)} \end{bmatrix} \qquad \underset{n \times 1}{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix}$$

n data points

p features

Linear Regression

$$\widehat{\beta} \;= \arg\min_{\beta} \frac{1}{n}(\mathbf{A}\beta - \mathbf{Y})^T(\mathbf{A}\beta - \mathbf{Y}) = \arg\min_{\beta} J(\beta)$$

$$J(\beta) = \;(\mathbf{A}\beta - \mathbf{Y})^T(\mathbf{A}\beta - \mathbf{Y})$$

- Poll

Is the objective convex in β?

A) Convex, quadratic in β
B) Non-convex, A may not be positive semi definite
C) Depends on conditioning (ratio of max:min eigenvalues) of A^TA
D) Convex, A^TA is positive semi definite