Simple case: $\{(x_i, y_i)\}_{i=1}^{N}$ $\qquad y_i \approx m x_i + b$

Multivariate Case: features of our data (K such)

samples: each sample consists of a measurement of the K features (N samples)

target variable that we are trying to predict

mpg $\approx$ displacement of engine, weight of car, acceleration of car.

$$y \simeq m_1 \cdot \text{disp} + m_2 \,\text{weight} + m_3 \cdot \text{accel} + b$$

"TIDY DATA CONVENTION"

$$\begin{array}{cc} \text{N rows} & \\ \text{each is a sample} & \begin{pmatrix} \quad & \quad & \quad \\ & & \\ & & \end{pmatrix} \\ & \text{K columns} \\ & \text{each is a feature} \end{array} \qquad \begin{array}{c} \text{N entries} \\ \begin{pmatrix} \ \\ \ \\ \ \end{pmatrix} \\ Y \end{array}$$

$$\begin{array}{cc} & \begin{array}{ccc} \text{disp} & \text{weight} & \text{accel} \end{array} \\ \begin{array}{c} \text{Car model}_1 \\ \text{Car model}_2 \\ \vdots \\ \text{Car model}_N \end{array} & \begin{pmatrix} \cdot & \cdot & \cdot \\ & \text{results of our data} & \\ & & \\ & & \end{pmatrix} \begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \end{array} \\ & X \end{array} \qquad \begin{array}{c} \begin{pmatrix} y_1 \\ \\ \\ y_N \end{pmatrix} \text{mpg} \\ Y \end{array}$$

$y_i \approx m_1 x_{i1} + m_2 x_{i2} + m_3 x_{i3} + b$

$$\underset{\substack{\text{observed} \\ \text{"target" values} \\ N \times 1 \text{ vector}}}{Y} \approx \underset{\substack{\text{tidy} \\ \text{data} \\ \text{matrix} \\ N \times K}}{\underline{X}} \underset{\substack{K \times 1 \\ \text{entries "are the slopes"}}}{M} + \underline{b} = \underset{\substack{N \times 1 \\ \text{matrix}}}{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}$$

add an extra column to our data matrix

$$X = \begin{pmatrix} & x_{ij} & & 1 \\ & & & 1 \\ & & & 1 \\ & & & 1 \\ & & & 1 \end{pmatrix} \qquad \underset{\substack{\#\text{of} \\ \text{samples}}}{N} \times \underset{\#\text{features}}{(K+1)}$$

$$Y \approx XM \qquad M \text{ is } K+1 \times 1 \text{ column vector}$$

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} & 1 \\ & & & & \end{pmatrix} \begin{pmatrix} m_1 \\ \vdots \\ m_{K+1} \end{pmatrix} = m_1 x_{11} + m_2 x_{12} + \cdots + m_K x_K + m_{K+1}$$

$$\|Y - XM\|^2$$

$$XM = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} & 1 \\ x_{21} & x_{22} & \cdots & x_{2K} & 1 \\ & & \vdots & & \end{pmatrix} \begin{pmatrix} m_1 \\ \vdots \\ m_{K+1} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}m_1 + x_{12}m_2 + \cdots + x_{1K}m_K + m_{K+1} \\ x_{21}m_1 + x_{22}m_2 + \cdots + x_{2K}m_K + m_{K+1} \\ \vdots \end{pmatrix}$$

$$\|Y - XM\|^2 = \sum_{i=1}^{N} \left(y_i - x_{i1}m_1 - x_{i2}m_2 - \cdots - x_{iK}m_K - m_{K+1}\right)^2$$

$$= MSE(m_1, \ldots, m_{K+1})$$

Multivariate optimization problem:

Find $m_1, \ldots, m_{k+1}$ so that $\|Y - XM\|^2$ is minimized.