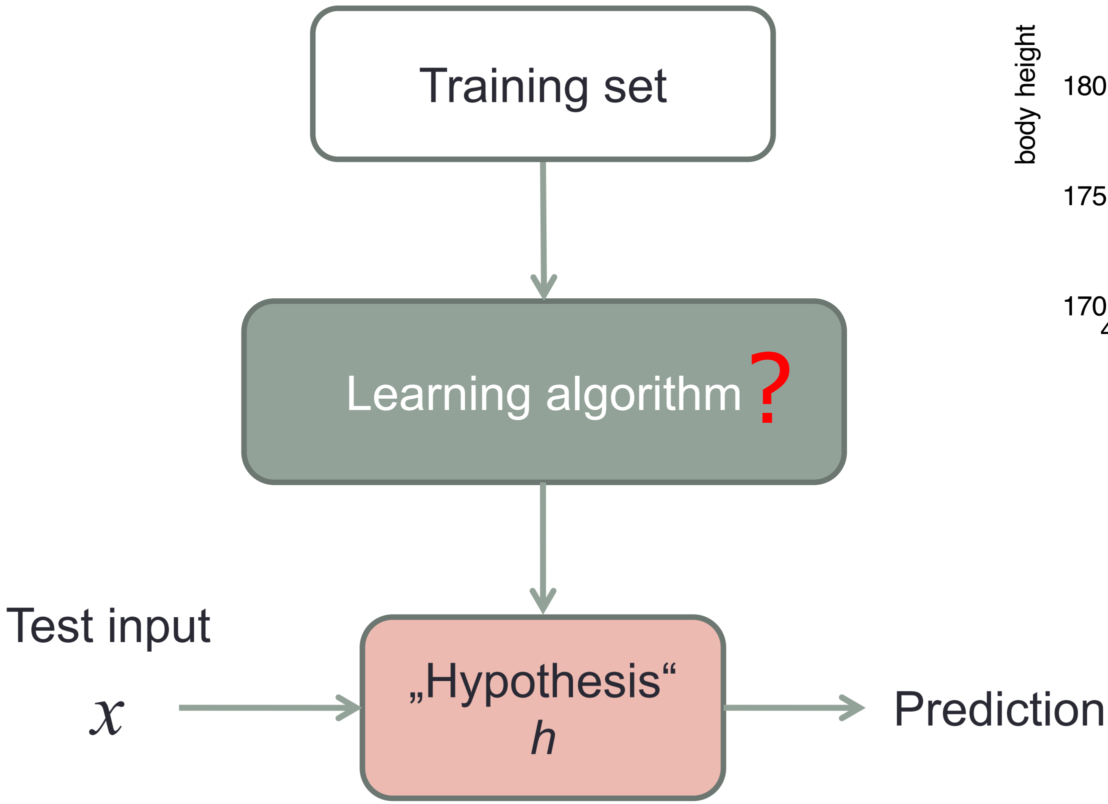
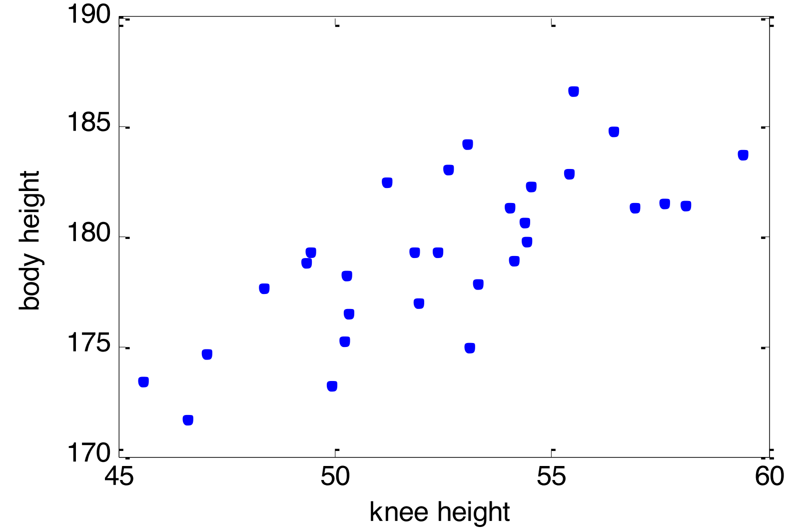
Linear regression with one input

$$\langle x^{(1)}, y^{(1)} \rangle \; \ldots \; \langle x^{(m)}, y^{(m)} \rangle$$





Hypothesis

$$h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 \cdot x$$

Parameters ?

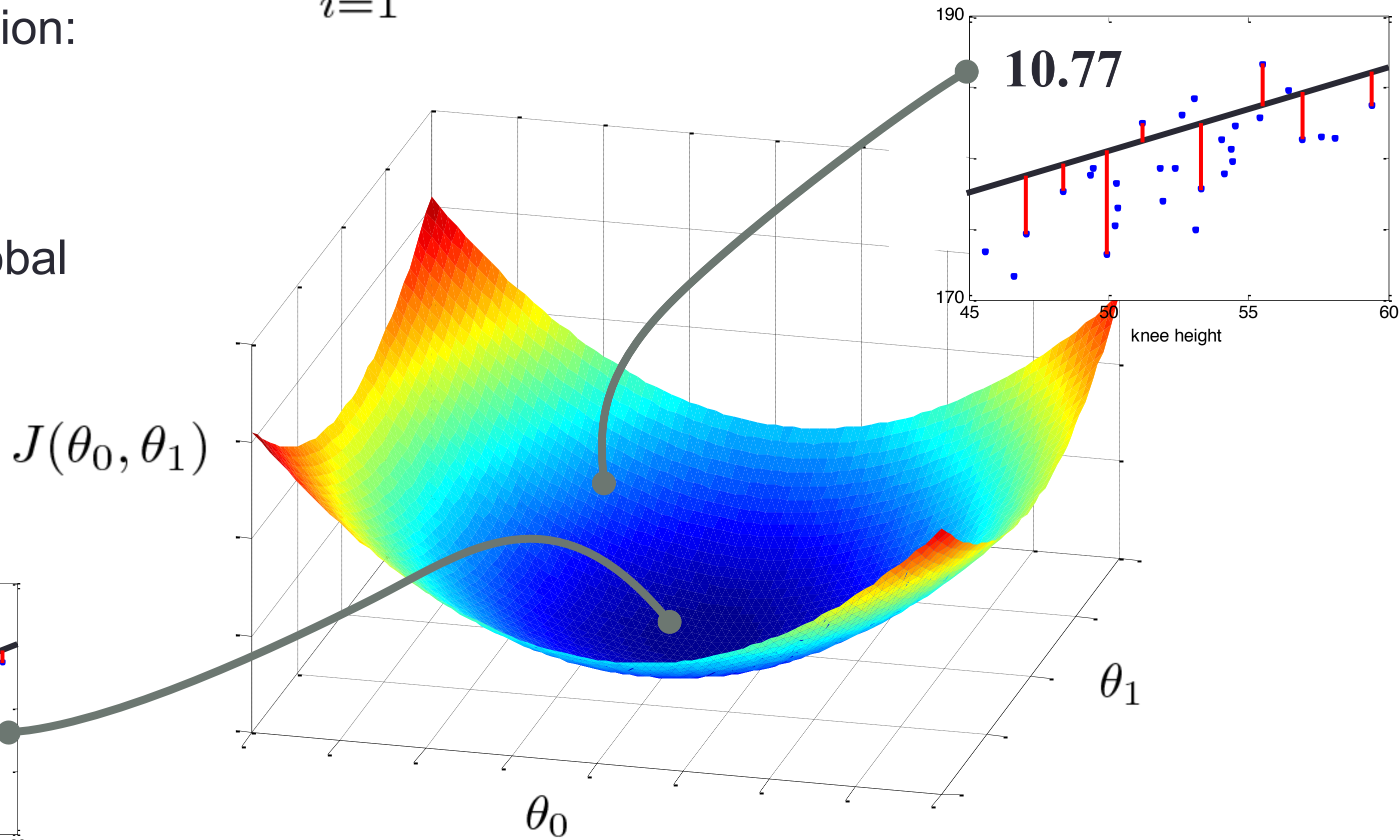
$$\boldsymbol{\theta} = (\theta_0, \theta_1)$$

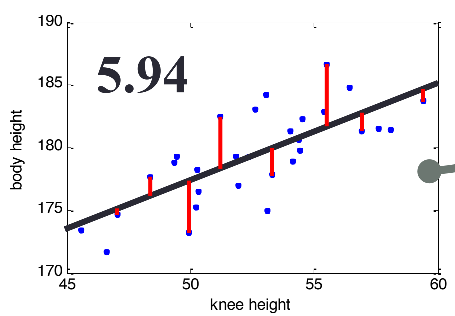
Cost function illustrated

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_{\boldsymbol{\theta}} \left(x^{(i)} \right) - y^{(i)} \right)^2$$

Properties of cost function:

- Quadratic function
- „Bowl"-shaped
- Unique local and global minimum (under „regular" conditions)





Linear hypothesis

- Hypothesis (one input):

$$h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 \cdot x$$

- Hypothesis (multiple input features):

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n$$

Example: h(x) = 50 + 0.5*kneeheight + 0.3*armspan + 0.1*age

- More compact notation:

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\theta}$$

$$\boldsymbol{x} = \begin{pmatrix} \boxed{x_0} \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$$

Introduce $\boxed{x_0 = 1}$

Why? Notation convenience!

Matrix and vector notation

x_0	Knee Height x_1	Arm span x_2	Age x_3	Height y
1	50	166	32	171
1	56	172	17	175
1	52	174	62	168

$$\boldsymbol{X} = \begin{pmatrix} \boxed{1 & 50 & 166 & 32} \\ 1 & 56 & 172 & 17 \\ 1 & 52 & 174 & 62 \end{pmatrix}$$

$$\boldsymbol{y} = \begin{pmatrix} 171 \\ 175 \\ 168 \end{pmatrix}$$

$$\boldsymbol{x}^{(i)} = \begin{pmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{pmatrix} \qquad \boldsymbol{X} = \begin{pmatrix} \boxed{\text{———} \left(\boldsymbol{x}^{(1)}\right)^T \text{———}} \\ \text{———} \left(\boldsymbol{x}^{(2)}\right)^T \text{———} \\ \vdots \\ \text{———} \left(\boldsymbol{x}^{(m)}\right)^T \text{———} \end{pmatrix} \qquad \boldsymbol{y} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

features of i'th training example

(n+1) × 1

design matrix

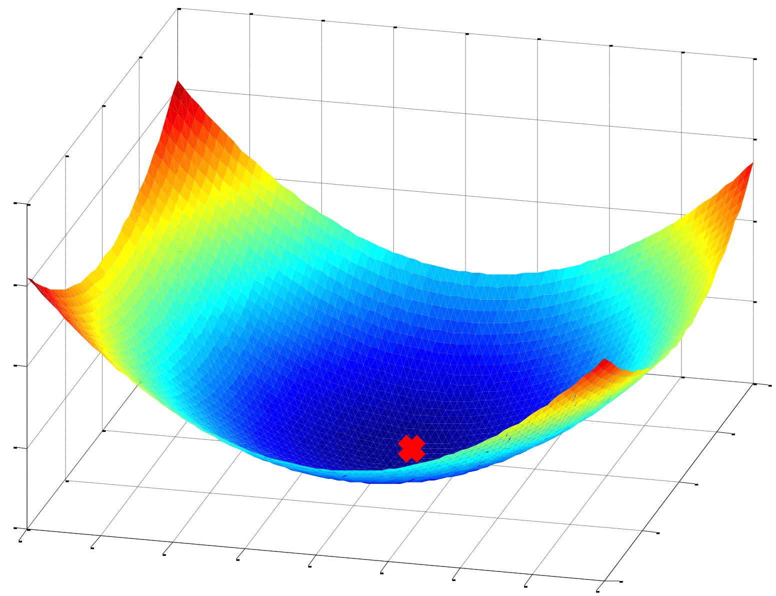
m × (n+1)

output/target vector

m × 1

Analytical solution

- Set all partial derivatives of cost function $J(\boldsymbol{\theta})$ = 0



- Solving system of linear equations yields:

$$\boldsymbol{\theta}^* = \underbrace{\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T}_{\text{Moore-Penrose Pseudoinverse of } \boldsymbol{X}}\boldsymbol{y}$$

$\boldsymbol{X}$ … design matrix

$\boldsymbol{y}$ … output/target vector

- *Note: This analytical solution requires that columns of* $\boldsymbol{X}$ *are linearly independent („regular“ conditions)*

Non-linear (quadratic) fit

x	y
0.01	-0.27
-1.22	2.63
0.17	-0.13
…	…

$$\boxed{\phi_0 = 1 \qquad \phi_1 = x \qquad \phi_2 = x^2}$$

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0.01 & 0.01^2 \\ 1 & -1.22 & (-1.22)^2 \\ 1 & 0.17 & (0.17)^2 \\ & \vdots & \end{pmatrix} \qquad \boldsymbol{y} = \begin{pmatrix} -0.27 \\ 2.63 \\ -0.13 \\ \vdots \end{pmatrix}$$

design matrix with non-linear features

Hypothesis: $h_{\boldsymbol{\theta}}(\boldsymbol{\phi}) = \theta_0 + \theta_1 \cdot \phi_1 + \theta_2 \cdot \phi_2$

Optimal parameters: $\boldsymbol{\theta}^* = \left(\mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \boldsymbol{y}$