IMAG/e

TU/e Technische Universiteit Eindhoven University of Technology

Linear regression

Cian Scannell

(slides from Mitko Veta)

Goals for today:

Introduce deep neural networks and their building blocks.

Define the linear regression method, which can be considered the most basic building block of neural networks.

Establish a connection between the knowledge obtained in the Registration topic with the CAD topic.

The Economist

SPECIAL REPORT

Artificial intelligence: Technology

From not working to neural networking

The artificial-intelligence boom is based on an old idea, but with a modern twist

The New York Times Magazine

The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

BY GIDEON LEWIS-KRAUS DEC. 14, 2016

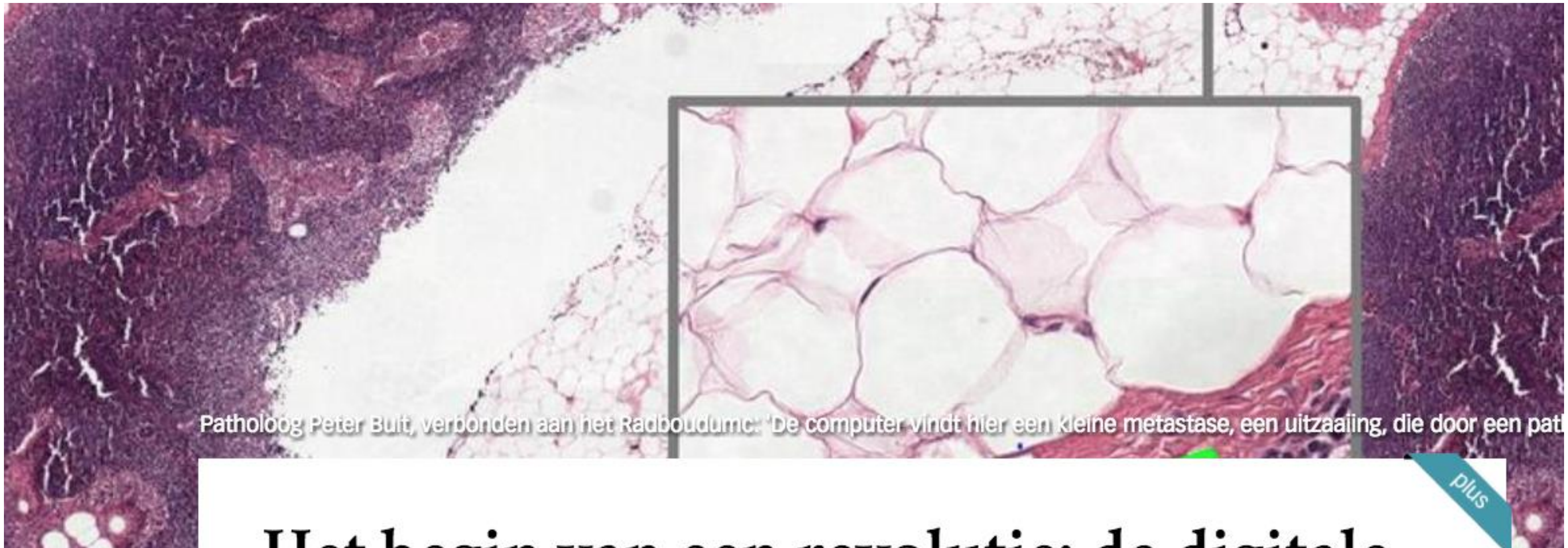
deVolkskrant

'AlphaGo kan Lee Sedol dan wel hebben verslagen, mij kan het niet verslaan.' ©

Zelfs 's werelds beste Go-speler is nu overtuigd: de mens is inferieur

Er was nog een sprankje hoop. Zo lang de hoogst geklasseerde Go-speler ter wereld, Ke Jie, volhield dat een computer nooit van hem zou winnen, zou de mens slimmer blijven dan de machine. Nu is dat toch gebeurd, op tamelijk bizarre wijze.

Door: Laurens Verhagen 5 januari 2017, 20:29



plus

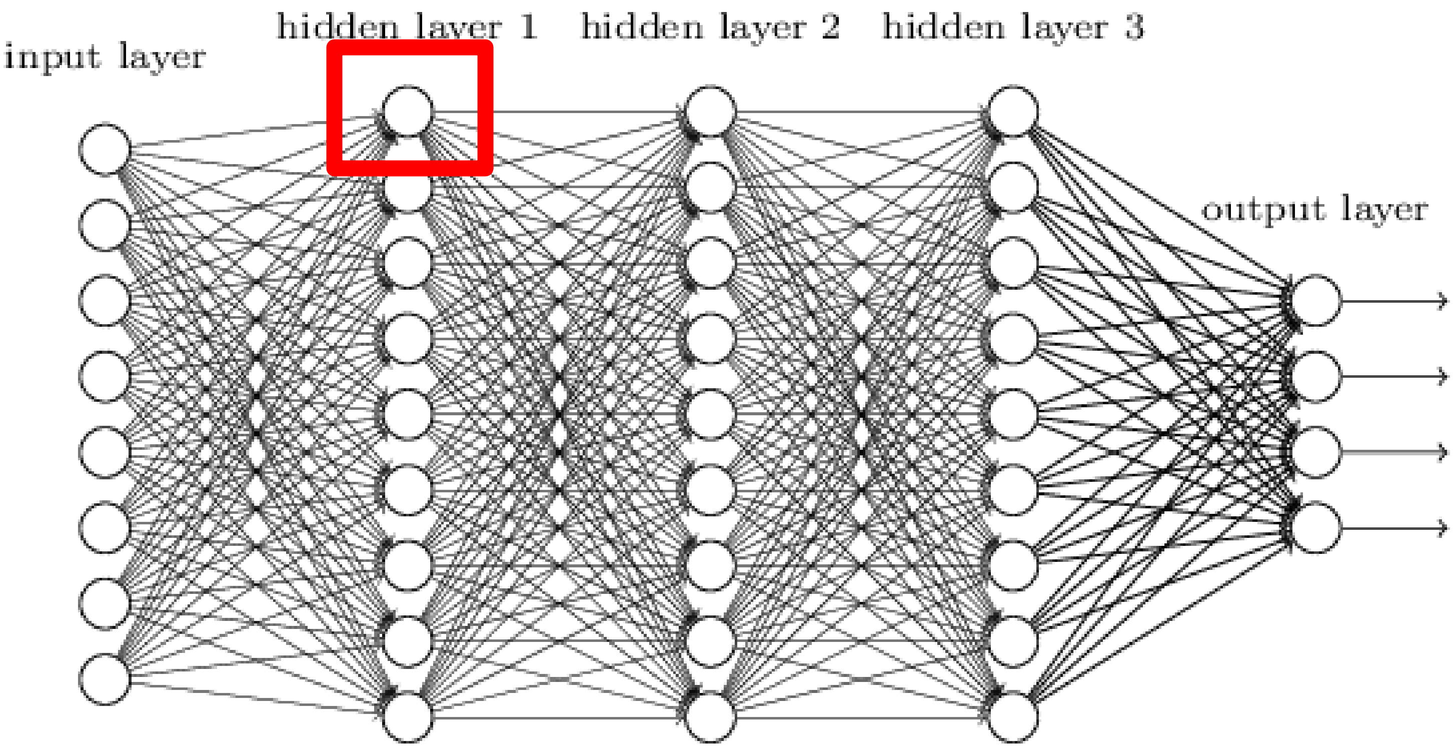
Het begin van een revolutie: de digitale patholoog

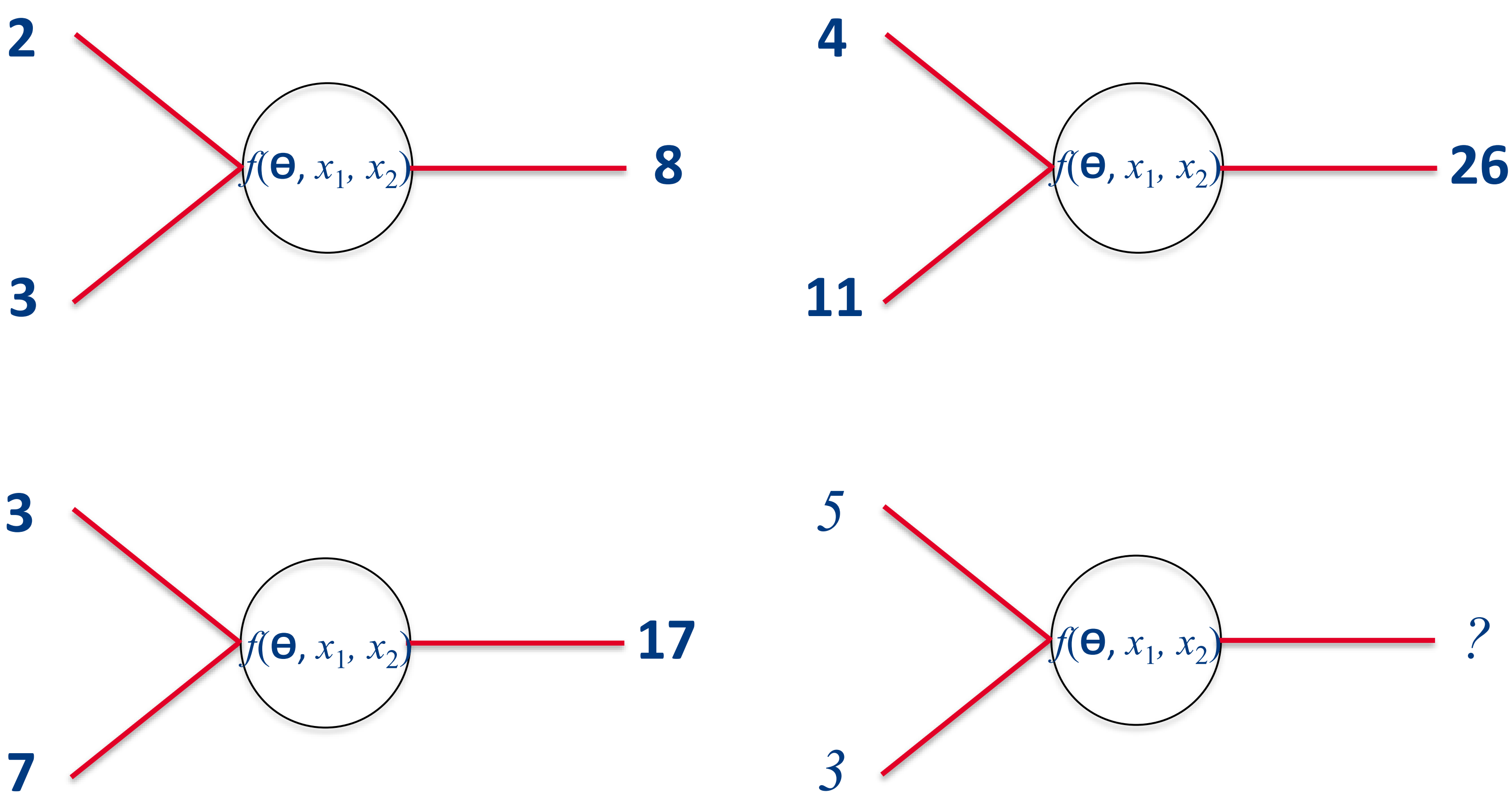
Kankeronderzoek

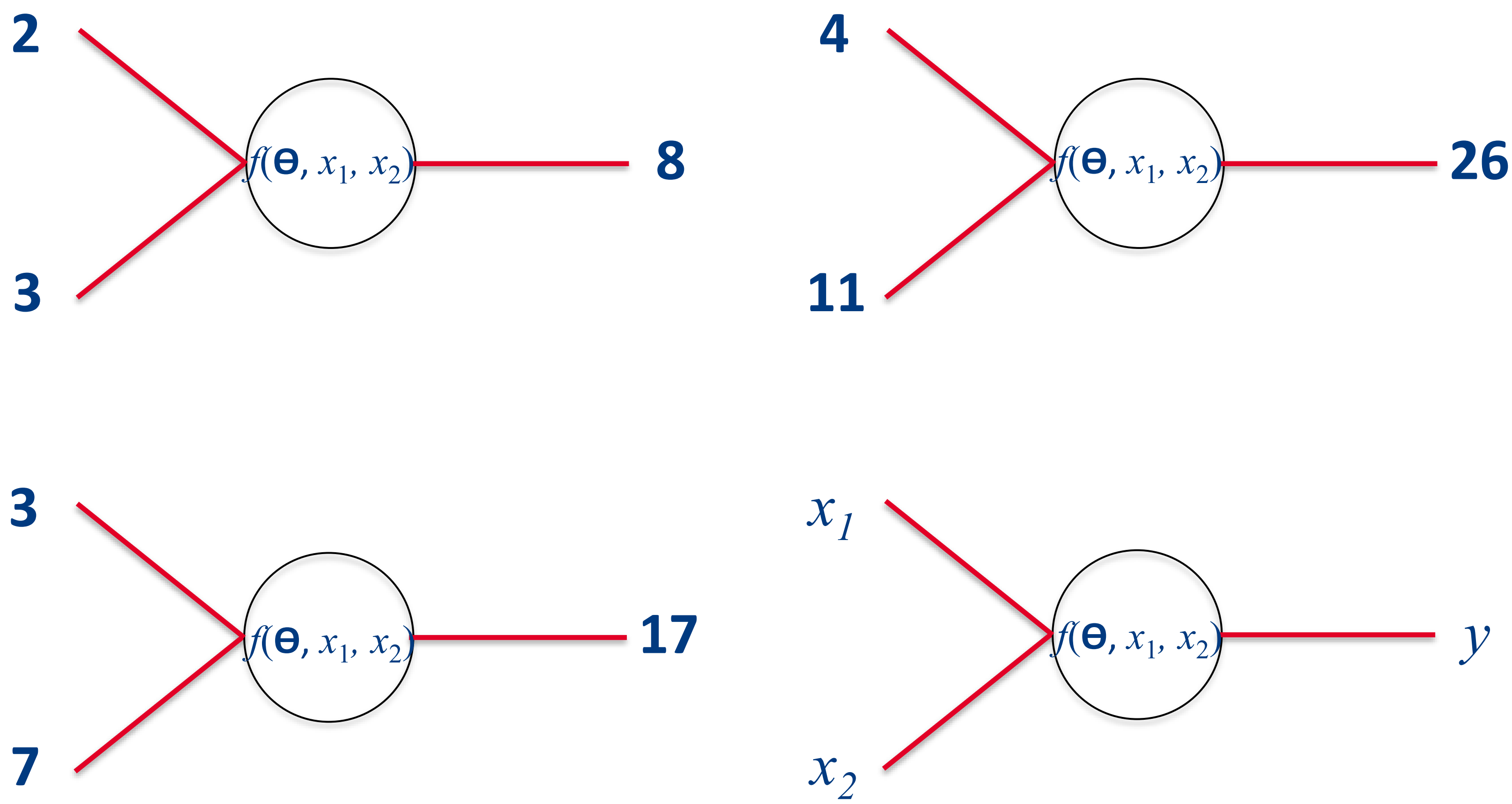
ARTIKEL Pathologen zeggen hun microscoop vaarwel. Computers beloven een revolutie bij opsporing van tumoren.

Door: Gerard Reijn 25 mei 2016, 02:00

A look ahead: Neural networks



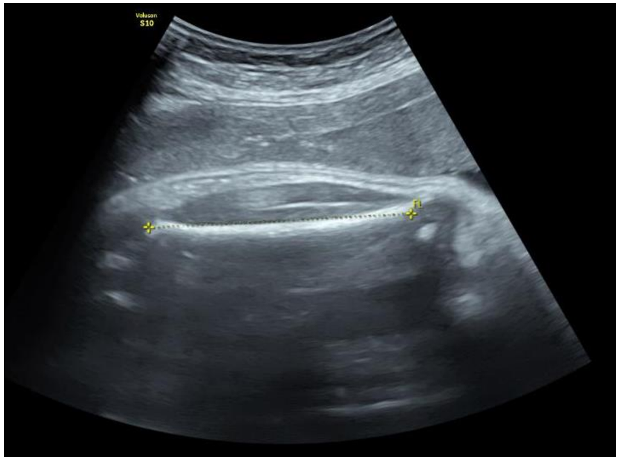

2
3
$f(\theta, x_1, x_2)$
8
4
11
$f(\theta, x_1, x_2)$
26
3
7
$f(\theta, x_1, x_2)$
17
5
3
$f(\theta, x_1, x_2)$
?

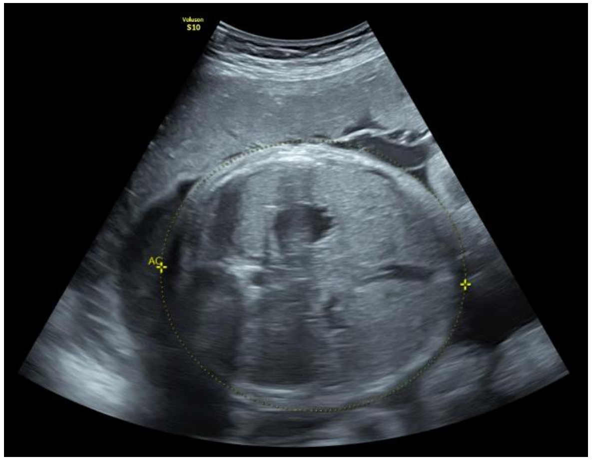


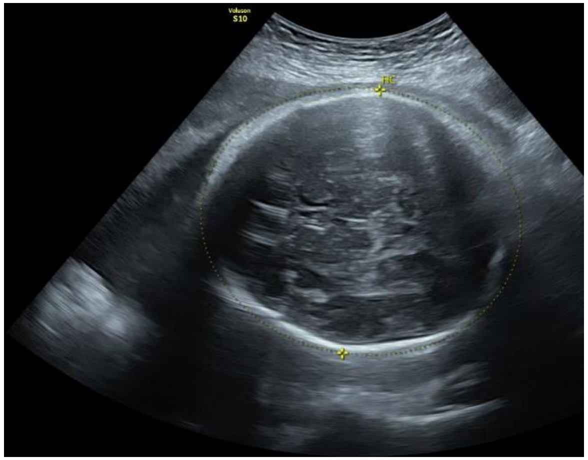
$$f(\boldsymbol{\theta}, x_1, x_2) = \theta_1 x_1 + \theta_2 x_2 + \theta_0$$

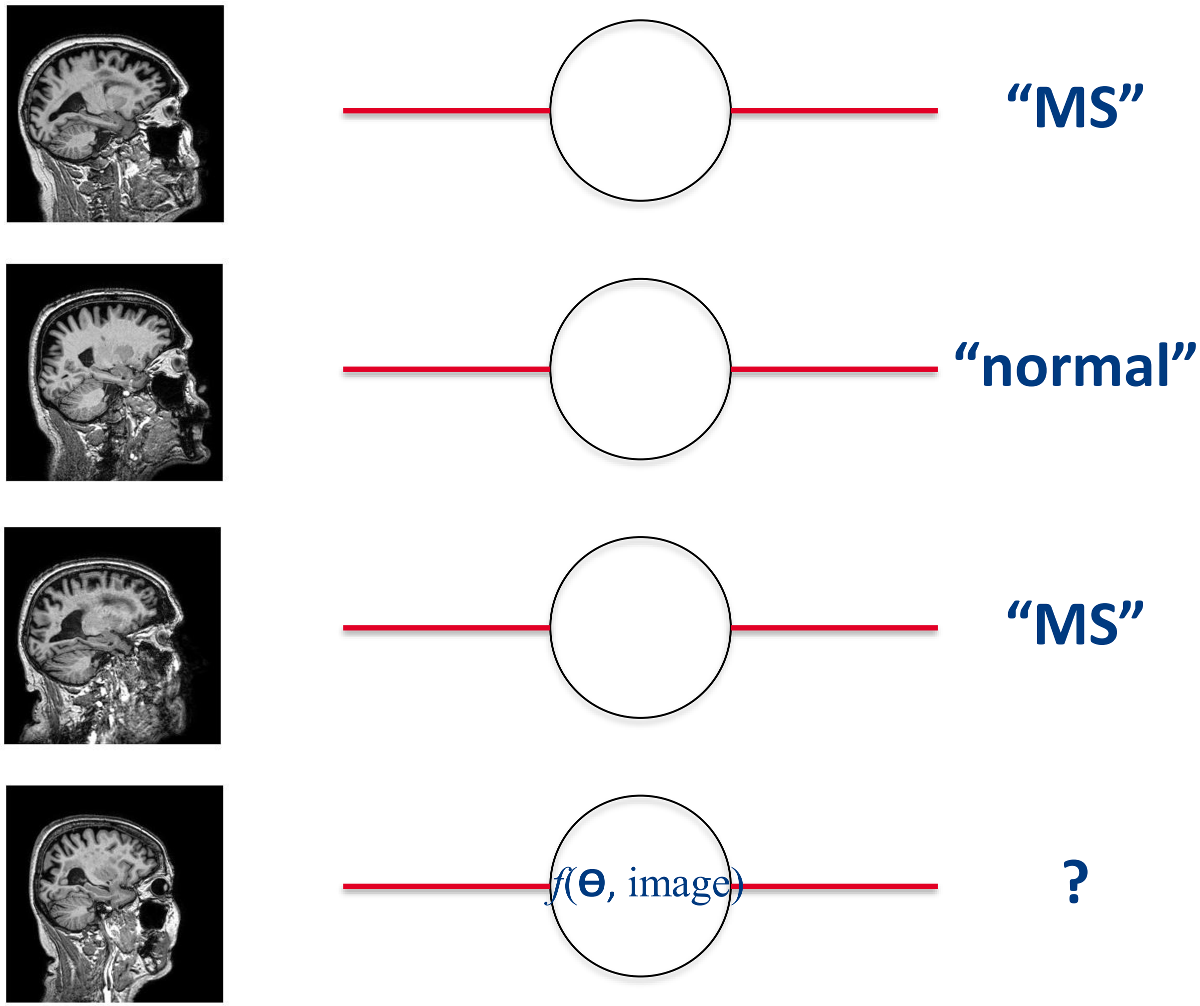
Fetal birth weight estimate model:

birth weight = $\Theta_1 \cdot$femur length + $\Theta_2 \cdot$abdominal circumfence + $\Theta_2 \cdot$head circumfence + Θ_0








"MS"
"normal"
"MS"
$f(\theta, \text{image})$
?

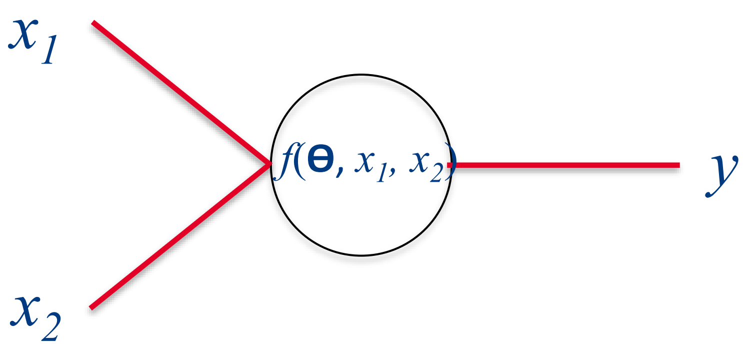
How to write a computer program that will solve the problem?

Step 0: Assume a model for the solution

Step 1: Write the error of the modelled solution given the available data

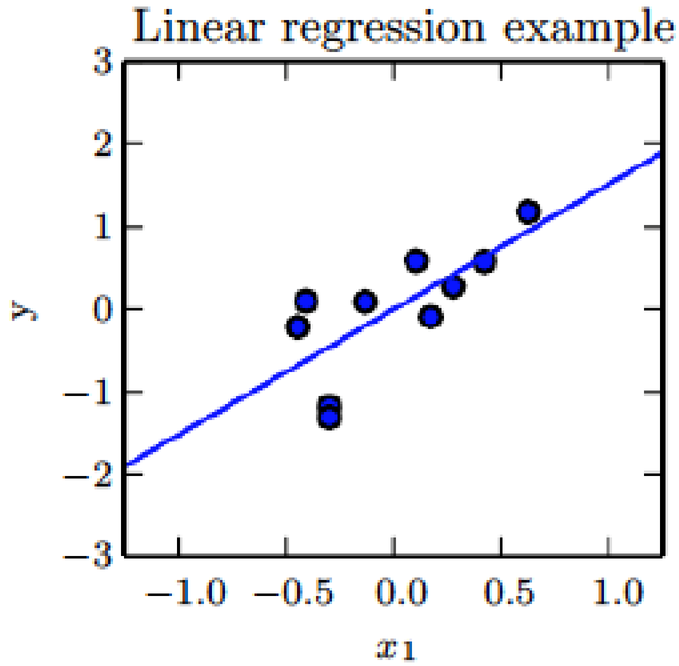
Step 2: Find a solution for the model parameters that minimizes the error and generalizes well to unseen data.

Step 0: linear model



$$f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$

Step 0: linear model

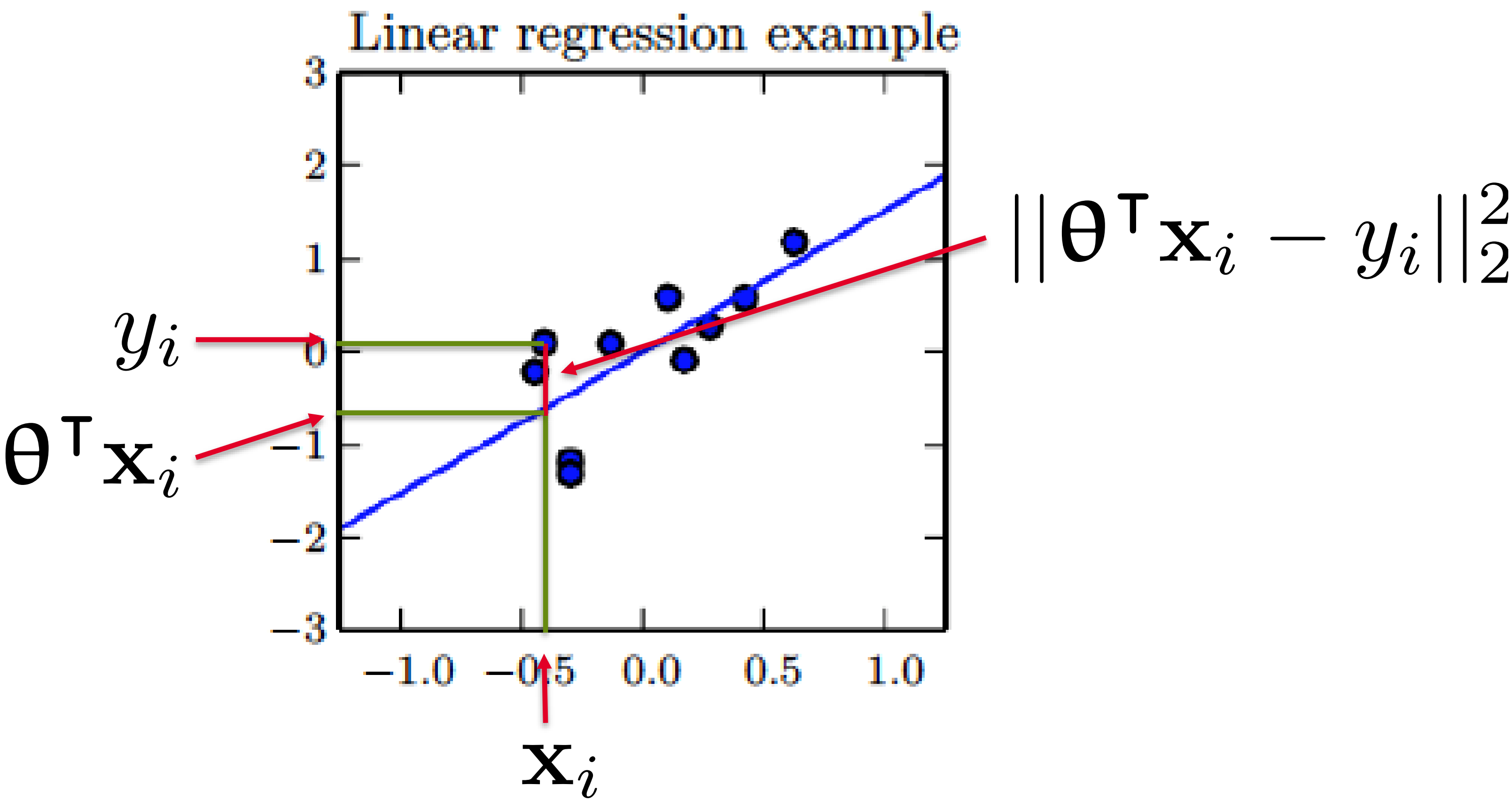


Step 0: linear model

$$f(w_1, w_2, b, x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$

$$f(\boldsymbol{\theta}, \mathbf{x}) = \boldsymbol{\theta}^\top \mathbf{x} \quad \boldsymbol{\theta} = \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

Step 1: error


Linear regression example
3
2
1
0
−1
−2
−3
−1.0
−0.5
0.0
0.5
1.0
y_i
$\theta^\top \mathbf{x}_i$
$\mathbf{x}_i$
$||\theta^\top \mathbf{x}_i - y_i||_2^2$

Step 1: error

$$J(\theta) = \sum_i ||\theta^\intercal \mathbf{x}_i - y_i||_2^2$$

$$J(\theta) = ||\mathbf{X}\theta - \mathbf{y}||_2^2 \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\intercal \\ \mathbf{x}_2^\intercal \\ \vdots \\ \mathbf{x}_N^\intercal \end{bmatrix}$$

Step 1: error

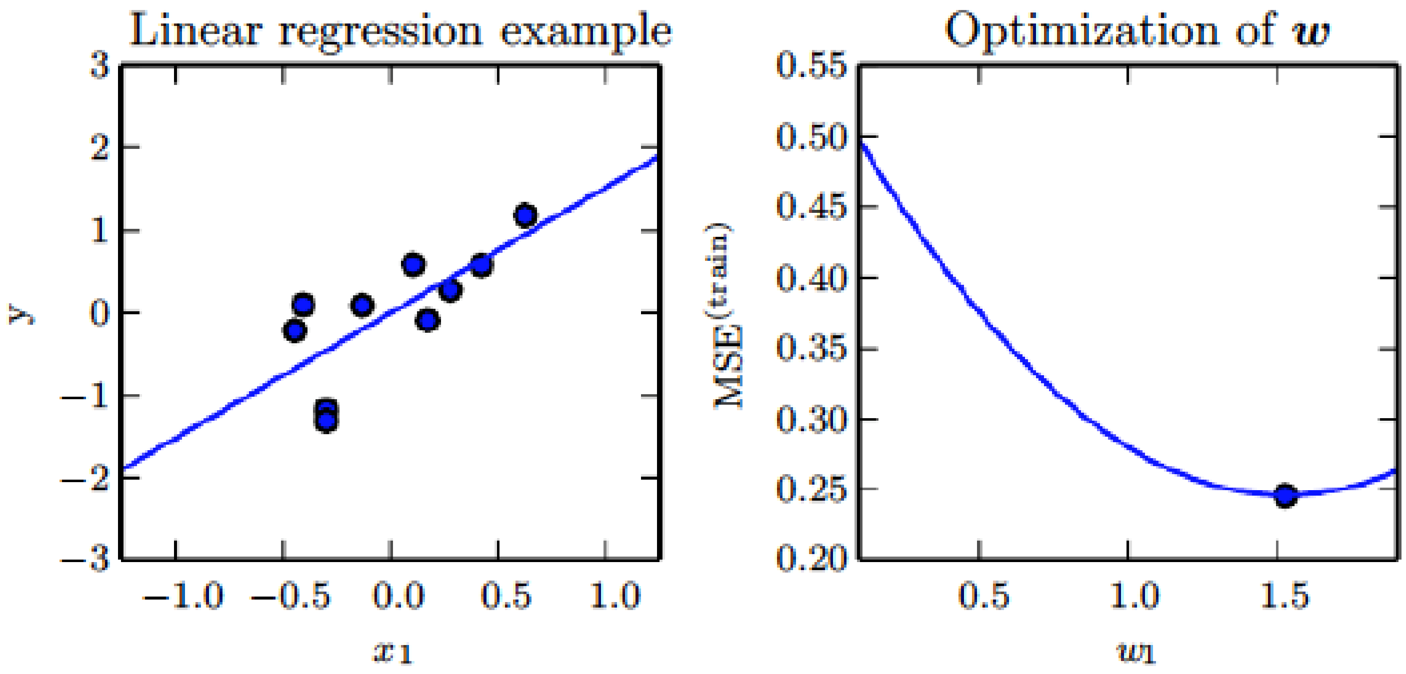


Figure from Deep Learning, Goodfellow et al.

Step 2: optimization

$$J(\boldsymbol{\theta}) = ||\mathbf{X}\boldsymbol{\theta} - \mathbf{y}||_2^2$$

In the point-based registration practical you have already implemented the solution for this optimization problem:

$$E^2 = ||\mathbf{A}\mathbf{w} - \mathbf{b}||_2^2$$

$||\mathbf{A}\mathbf{w} - \mathbf{b}||_2^2$ is a convex function in $\mathbf{w}$ and its solution can be found by setting the expression for the derivative with respect to the unknowns $\mathbf{w}$ to zero. The solution for $\mathbf{w}$ obtained in this way is:

$$\mathbf{w} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$$

Step 2: optimization

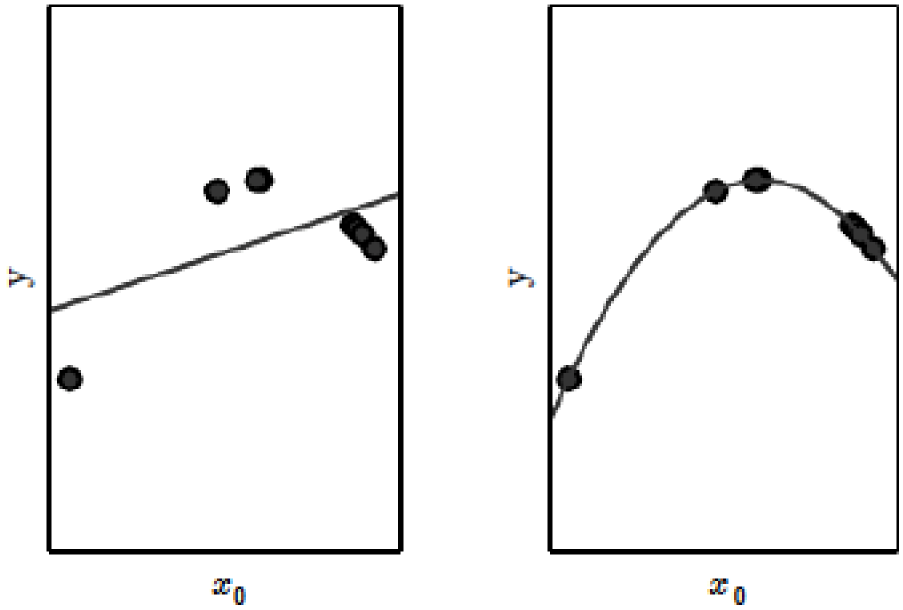
$$J(\boldsymbol{\theta}) = ||\mathbf{X}\boldsymbol{\theta} - \mathbf{y}||_2^2$$

$$\nabla_{\boldsymbol{\theta}} J = 0$$

$$\boldsymbol{\theta} = (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{y}$$

In some cases linear models are not sufficient and higher order polynomials are more suitable:

$$f(x) = wx + b \qquad\qquad f(x, x^2) = w_1 x + w_2 x^2 + b$$



Note that the quadratic model is linear in the parameters (does not contain quadratic terms of the parameters such as ${w_1}^2$).

What changes is the data. The quadratic model is equivalent to the linear model with two features where the second feature is x^2.

$$f(x) = wx + b \qquad f(x, x^2) = w_1 x + w_2 x^2 + b$$

$$f(\boldsymbol{\theta}, \mathbf{x}) = \boldsymbol{\theta}^\intercal \mathbf{x} \quad \mathbf{x} = \begin{bmatrix} x \\ 1 \end{bmatrix} \quad \mathbf{x}^{\text{new}} = \begin{bmatrix} x \\ x^2 \\ 1 \end{bmatrix}$$

More generally speaking, non-linear relationship can be modelled by creating a new feature vector with any non-linear mapping:

$$\mathbf{x}^{\text{new}} = \phi(\mathbf{x})$$

Generalization:

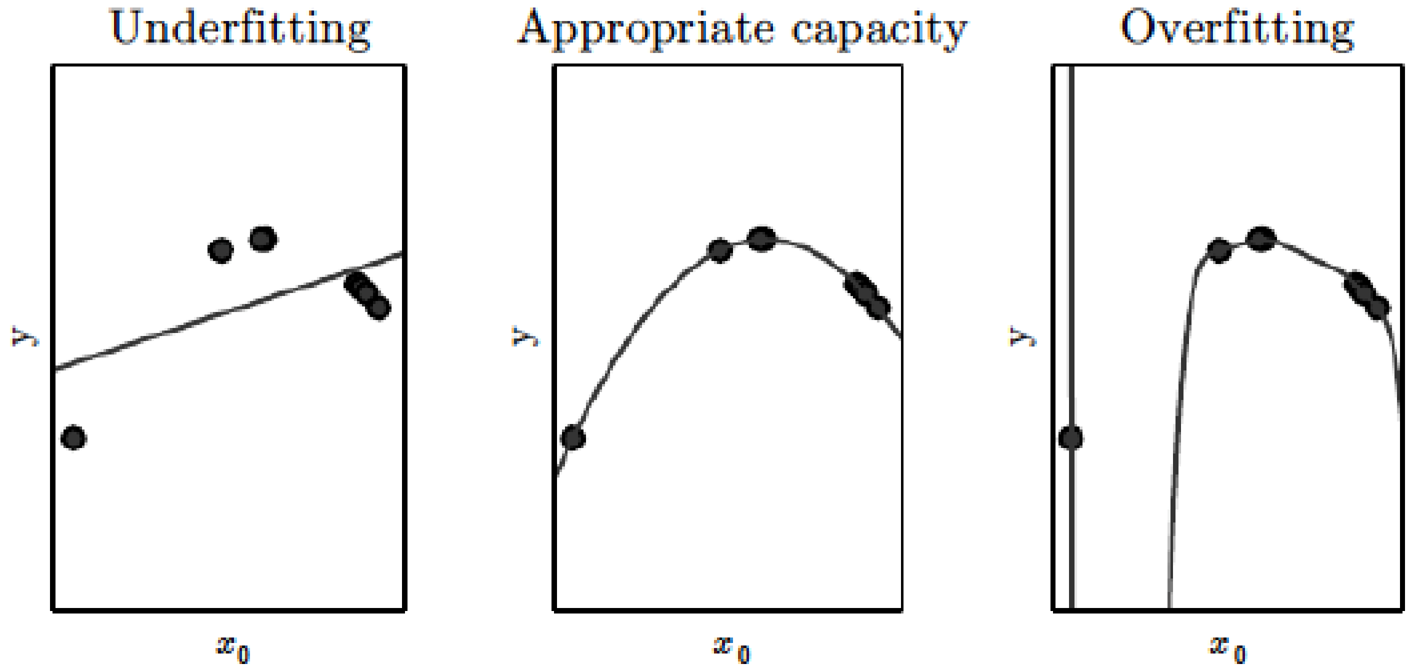
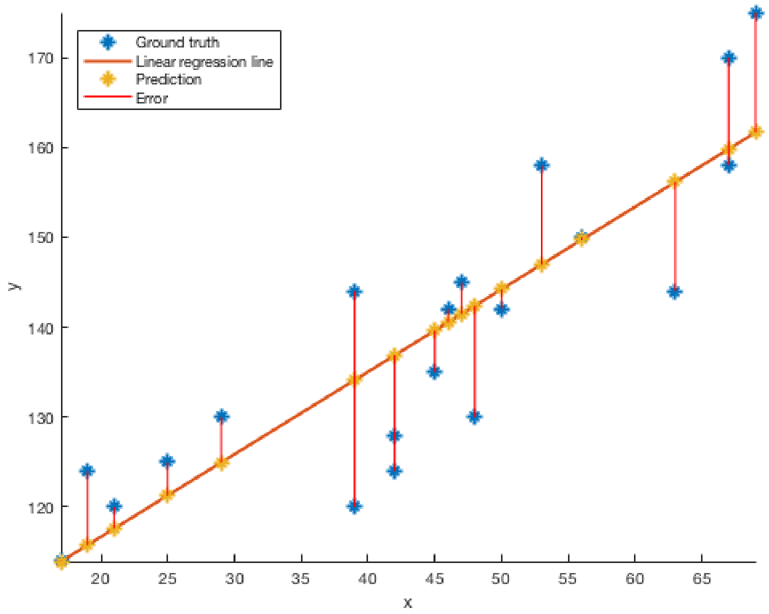
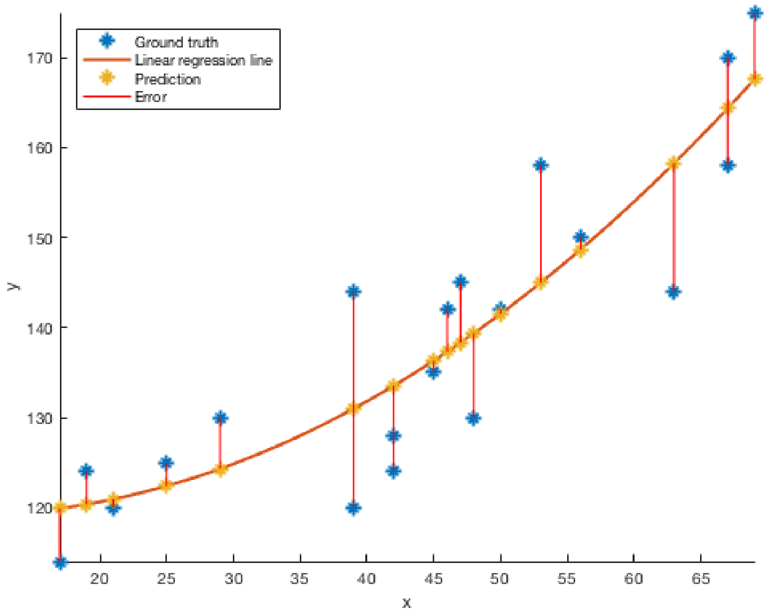


Figure from Deep Learning, Goodfellow et al.

A look ahead – exercises:





Linear regression for medical image analysis- measuring nuclei size:

$\mathbf{x}$

24 × 24 pixel images of nuclei from breast cancer histopathology slides

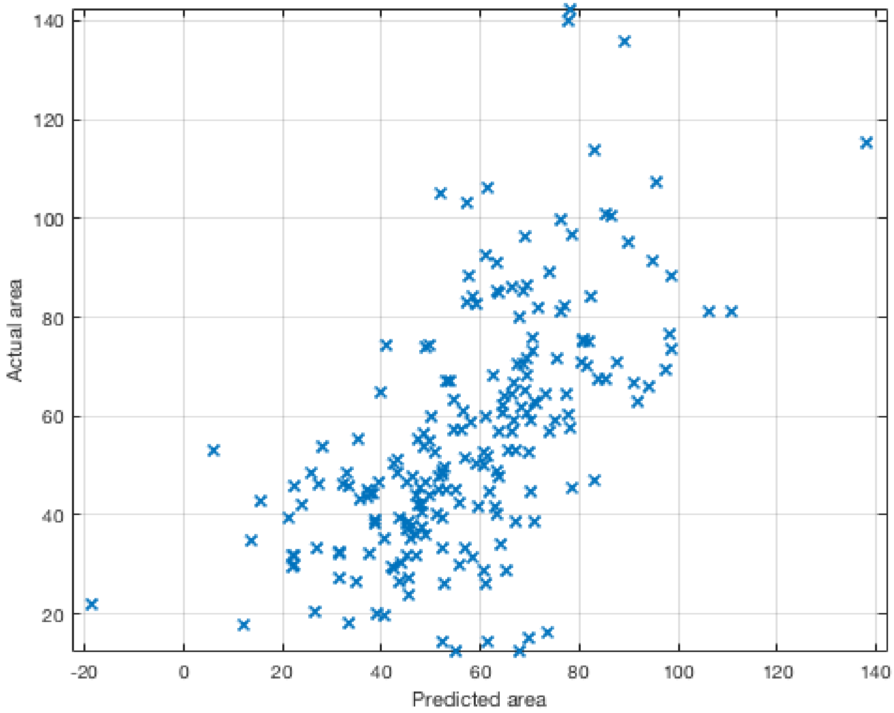
98.25 102.06 68.81 20.13 27.44 81.00 48.06 43.75 33.69 98.25

y

area of the nucleus in the center of the image in μm^2

Linear regression for medical image analysis- measuring nuclei size:

y_i



$\theta^{\mathsf{T}} \mathbf{x}_i$

NOTE: This is not a plot of Feature **x** vs. target y but predicted vs. actual y.

When we continue:

How to properly perform machine learning experiments and report results?
Extend linear regression to a classification model.
Combine these classification models into a neural network.

Discussion point 1

You are given a dataset of cardiac MRI images. You want to train a machine learning model that will “roughly” segment the left ventricle with a contour that is a perfect circle (e.g. the red circle in the figure below).

Can you do this with a regression model? If not, explain why? If yes, what will be the inputs and the outputs of the moodel?

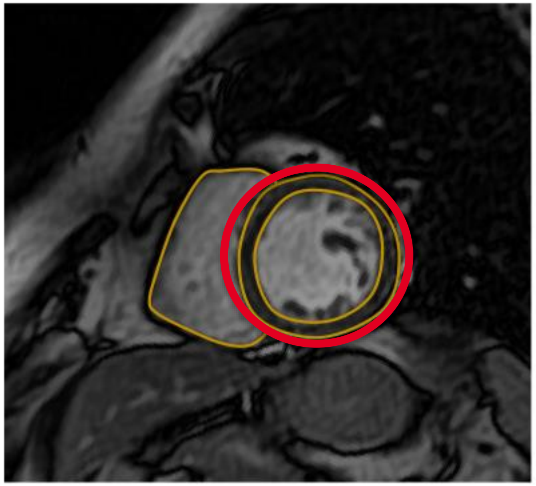


Figure from: https://www.ub.edu/mnms/

Discussion point 2

You have trained a linear regression model for a problem with two input features. It turned out that $w_2=0$. What information does that give you about the second input feature x_2.

$$f(w_1, w_2, b, x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$

Figure from: https://www.ub.edu/mnms/