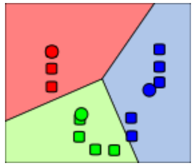
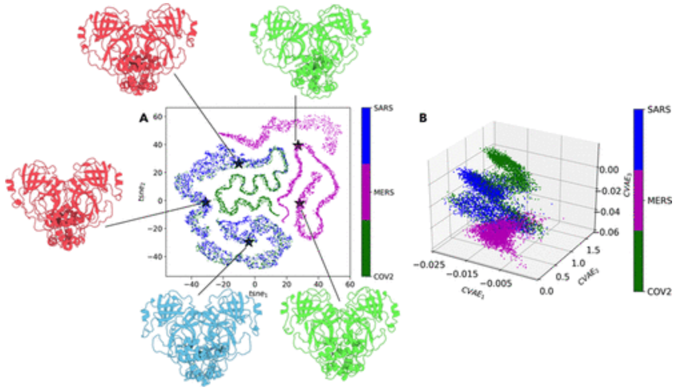
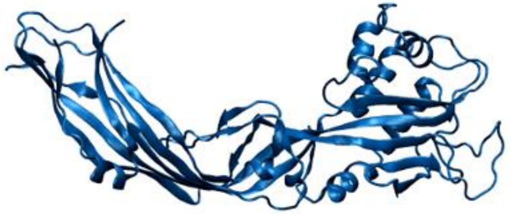
From biomolecular data to information

Antonia Mey

Matteo Degiacomi







Durham University

matteo.t.degiacomi@dur.ac.uk

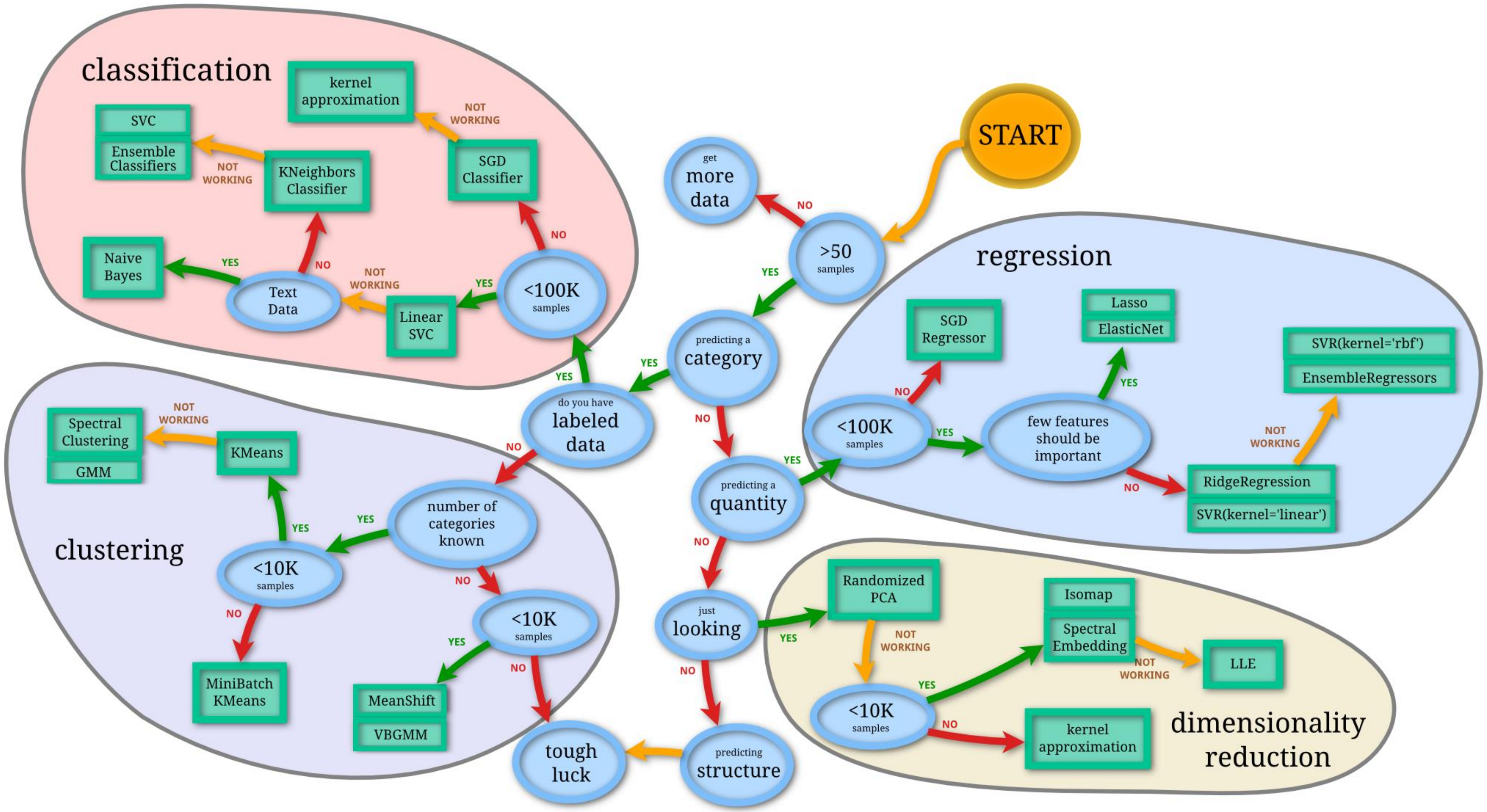
@MatteoDegiacomi

antonia.mey@ed.ac.uk

@ppxasjsm

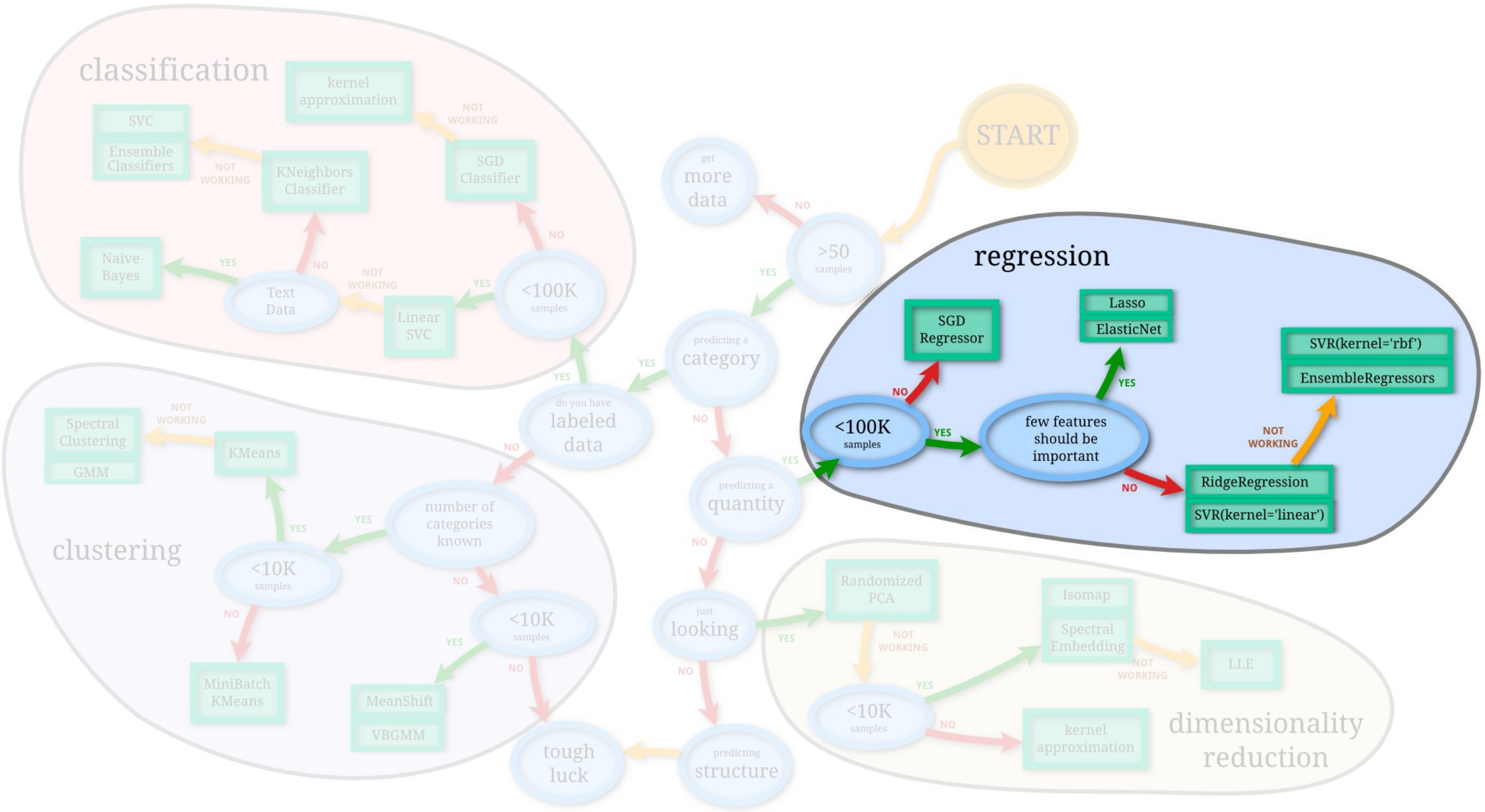
THE UNIVERSITY of EDINBURGH

The Data Mining world



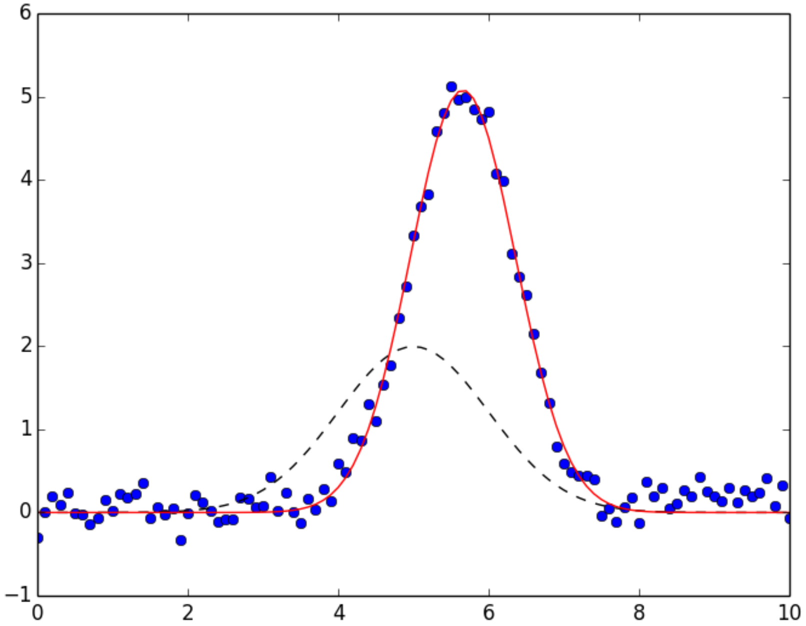
From **scikit-learn.org**

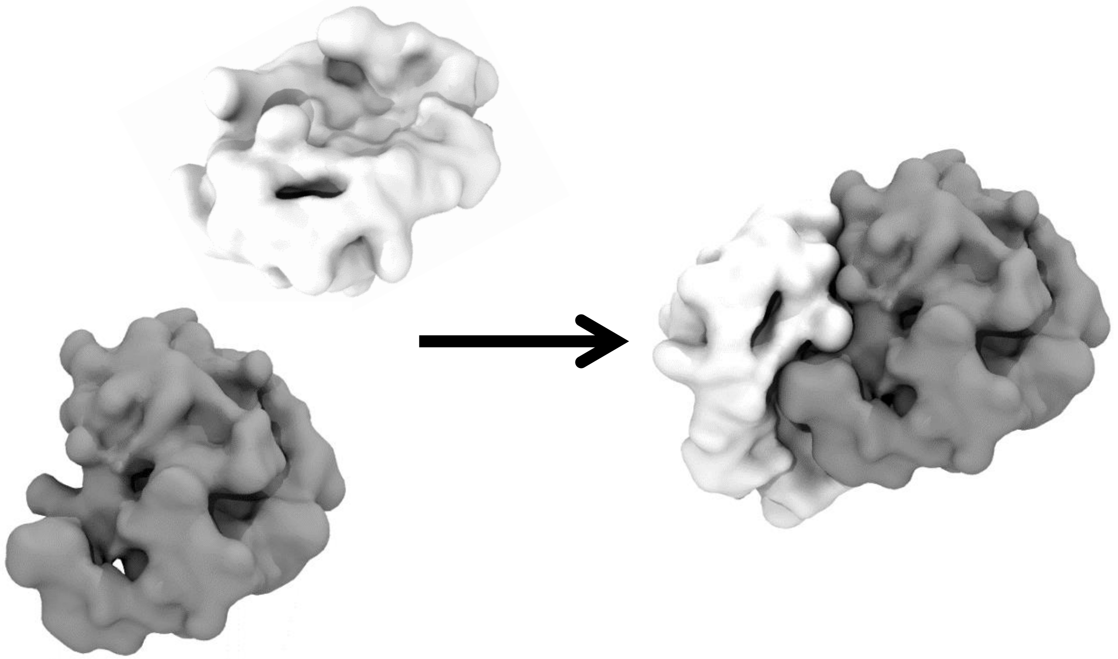
The Data Mining world

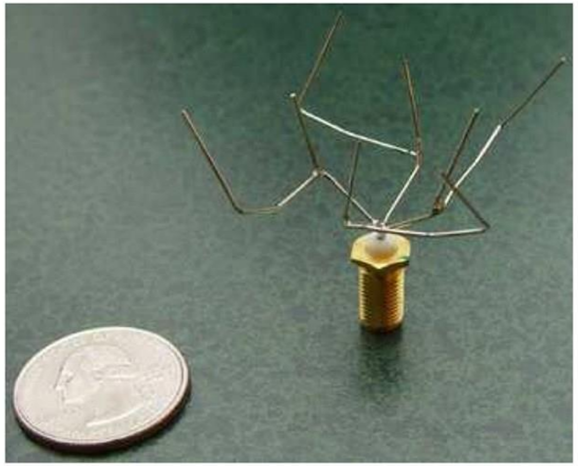


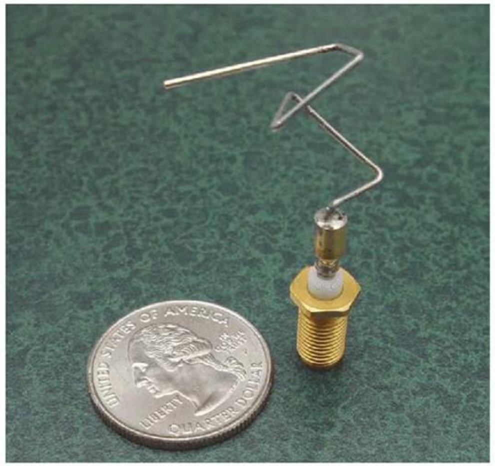
From **scikit-learn.org**

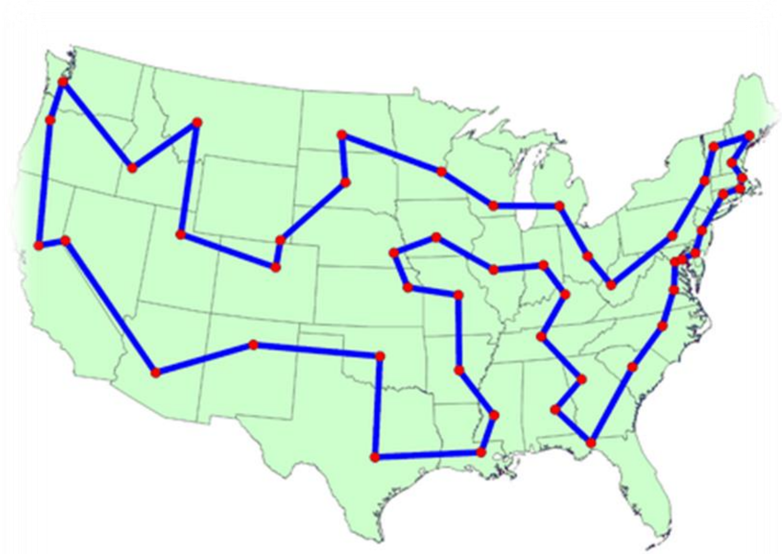
Data fitting is an optimization problem



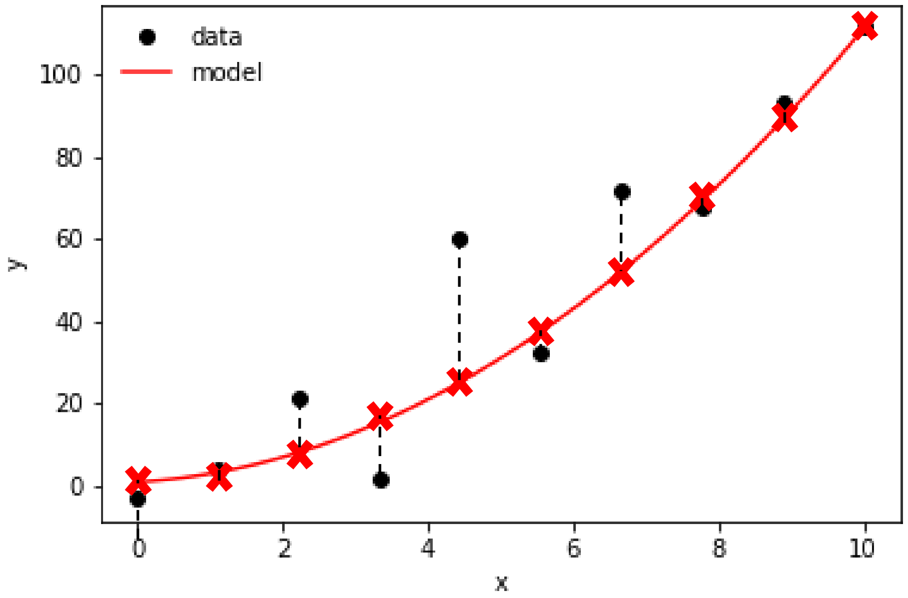








Linear Least squares



- Model predicts values for each datapoint
  $\hat{y}_i = f(x_i; a, b, \ldots)$

- **Residuals** quantify prediction error
  $r_i = y_i - f(x_i; a, b, \ldots)$

The best model minimizes the sum of squared residuals («loss function»)

$$\mathrm{E}(\mathrm{a, b, \ldots}) = \sum_{i=1}^{N} r_i^2$$ **, solve ∇E = 0**

Use least squares (i.e. analytical solution) if:

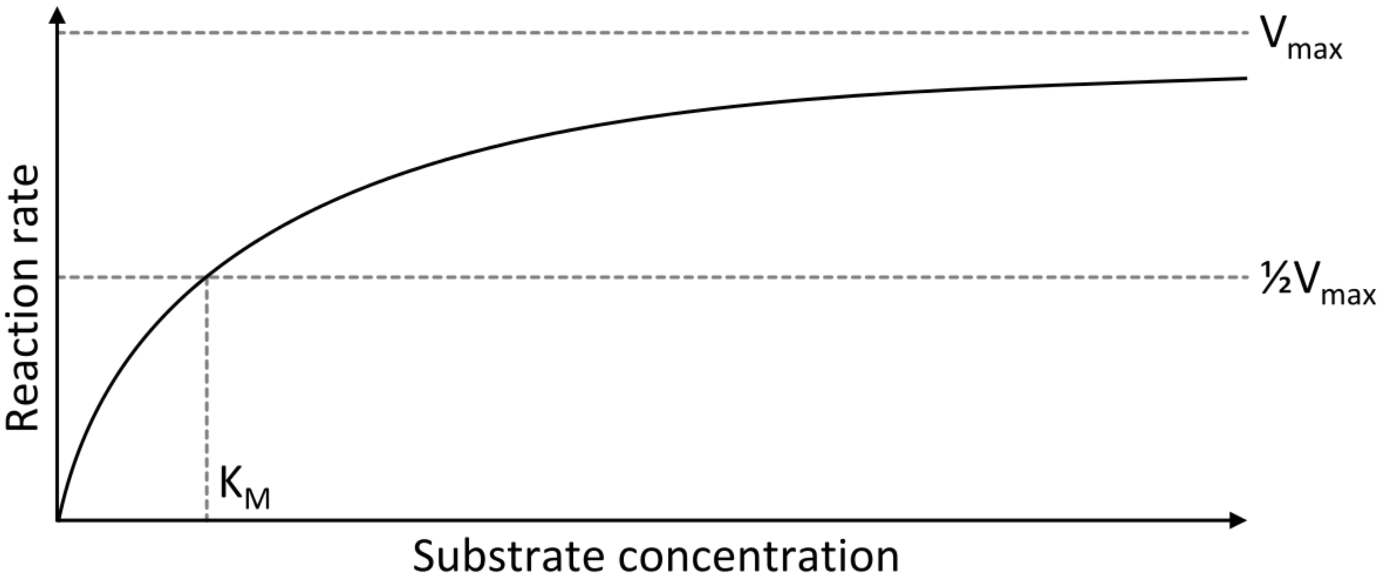
- there are more data points than parameters (*overdetermined*)
- model's parameters combine linearly *(linear least squares)*

Non-linear least squares

Non-linear combinations of model parameters. E.g. Michelis-Menten model:

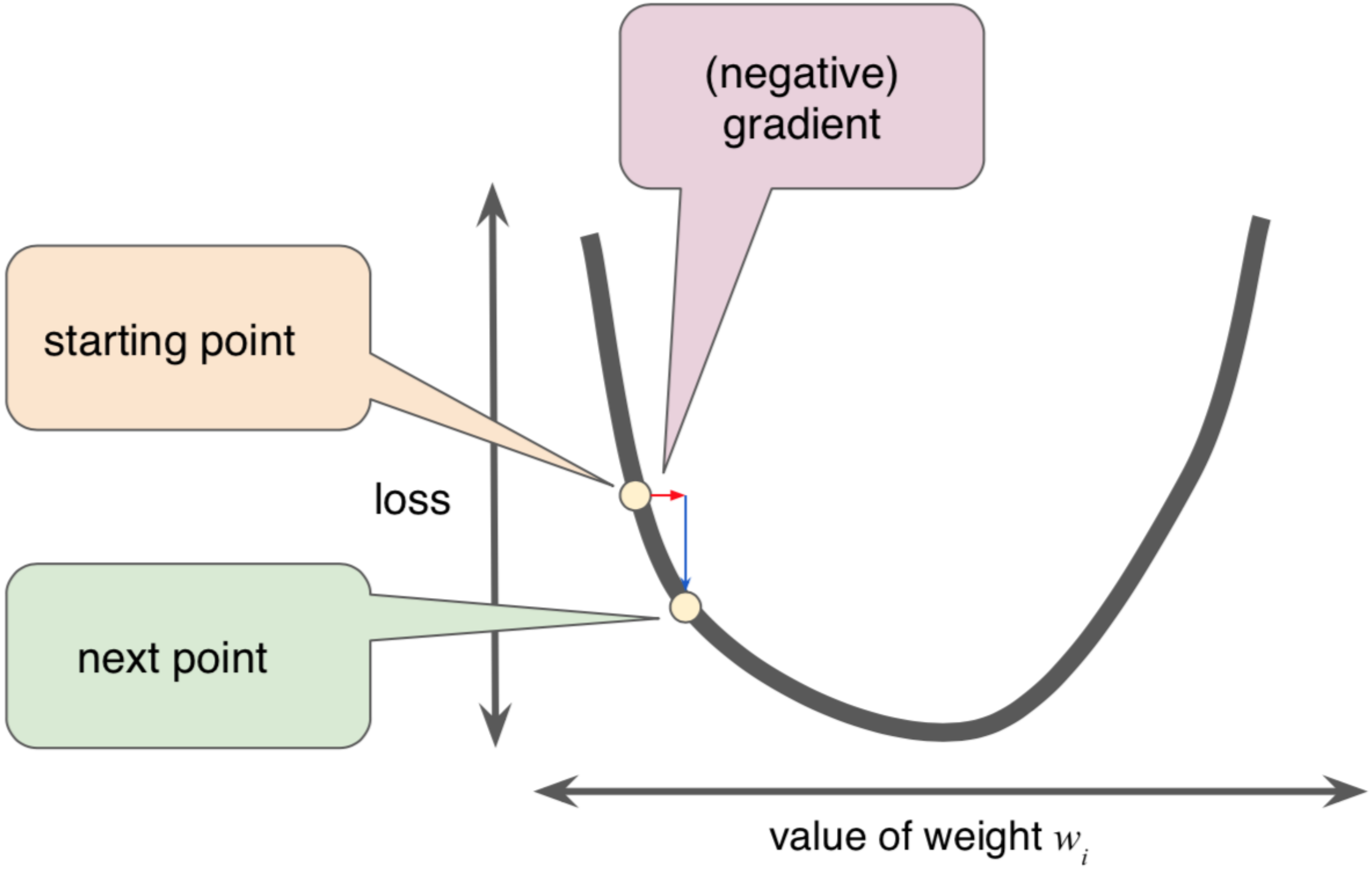
$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]}$$

$$f(x, a, b) = \frac{a\,x}{b + x}$$



- Cannot be solved using least squares
- Needs to be solved via *iterative methods*

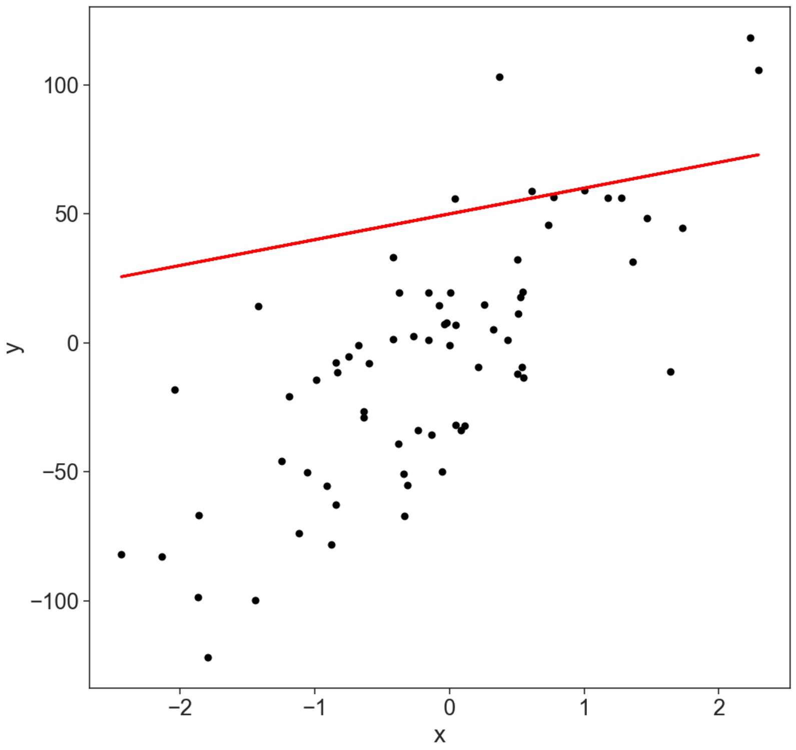
Gradient descent



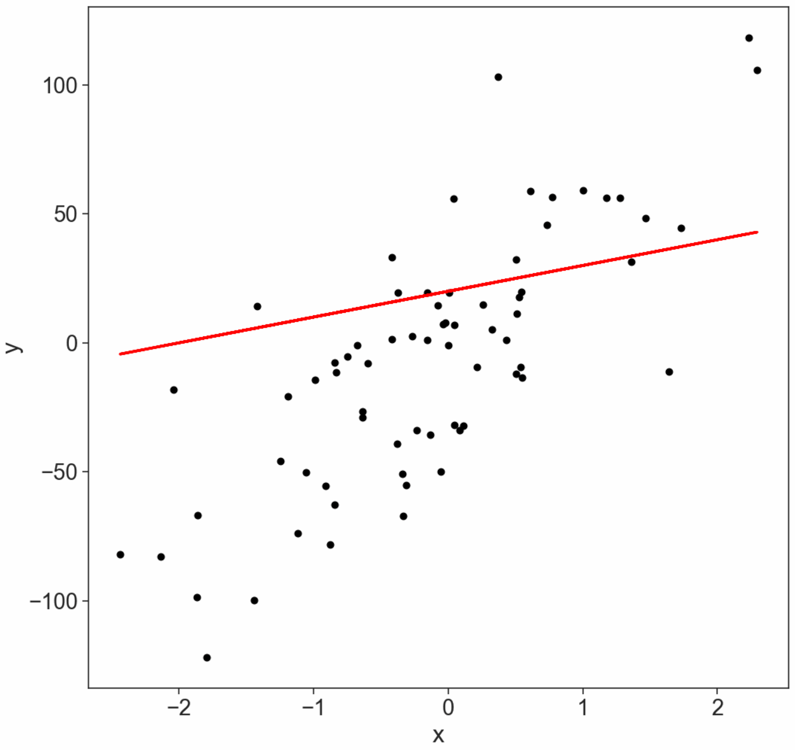
In machine learning, the step size of the gradient descent is called the **learning rate**

There is an optimal learning rate for every regression problem

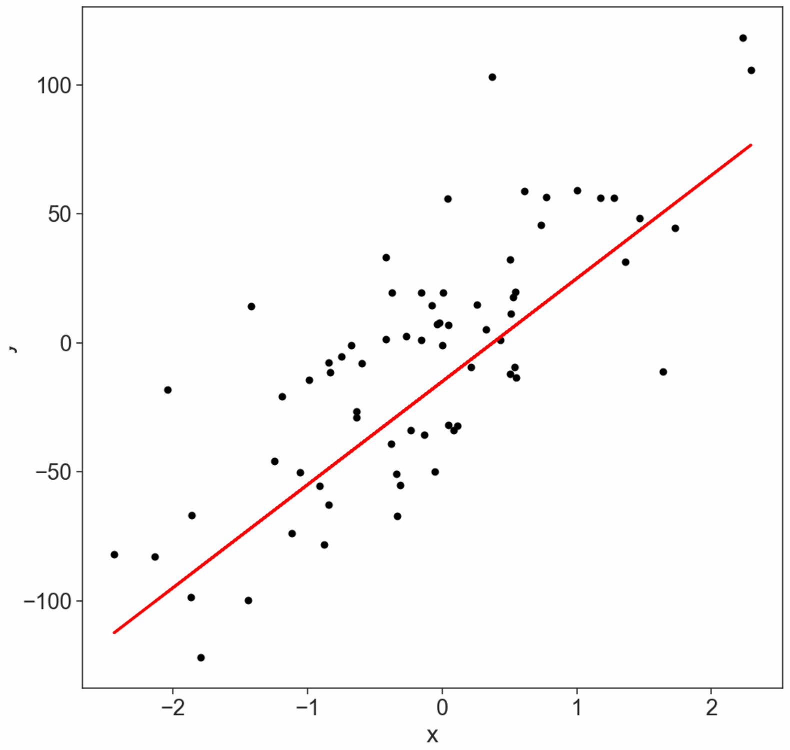
Gradient descent: learning rate and initial conditions



Low learning rate
bad starting point
424 steps

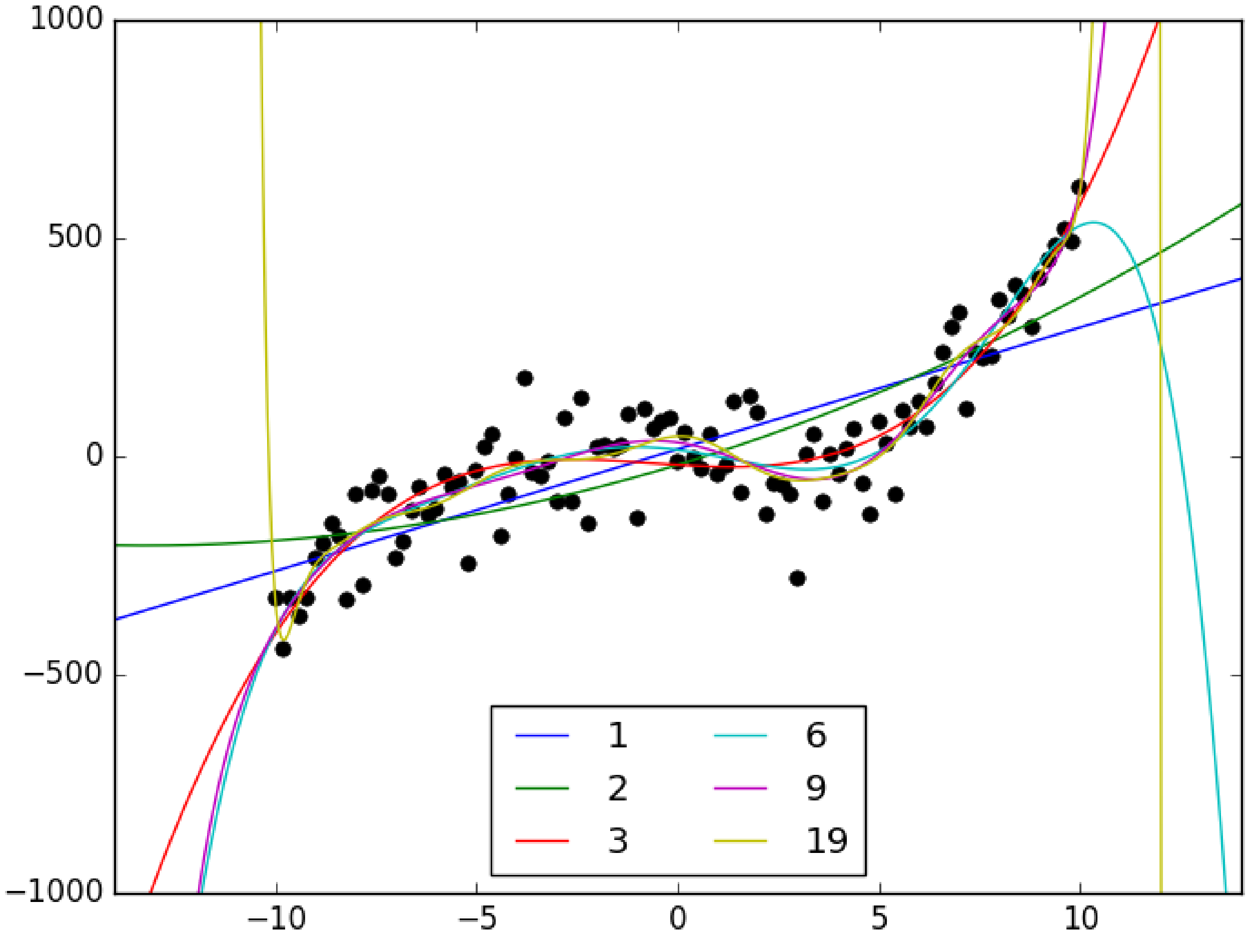


High learning rate
bad starting point
52 steps



Low learning rate
Good starting point
379 steps

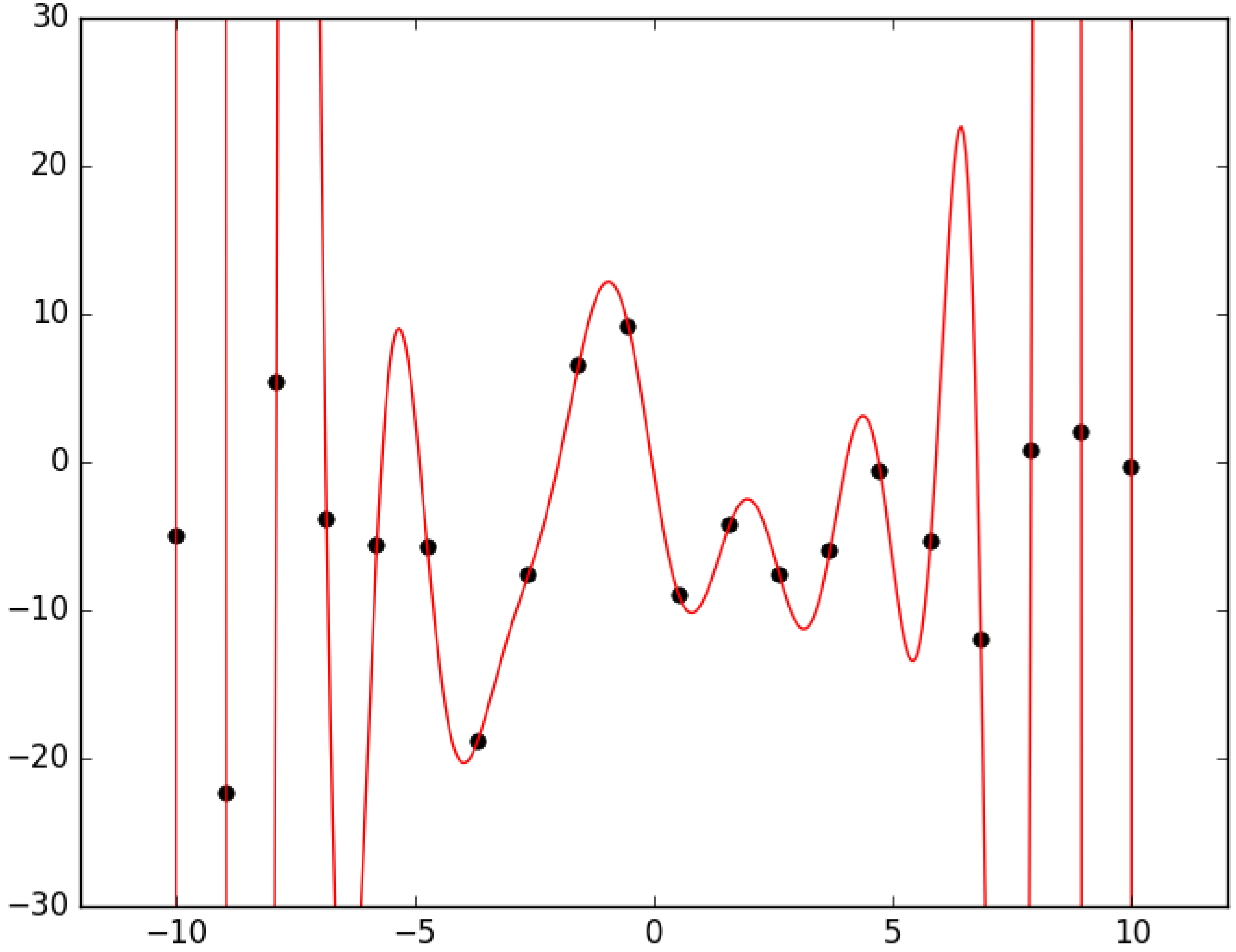
The more parameters, the better the fit?



"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk"

John von Neumann

The more parameters, the better the fit?

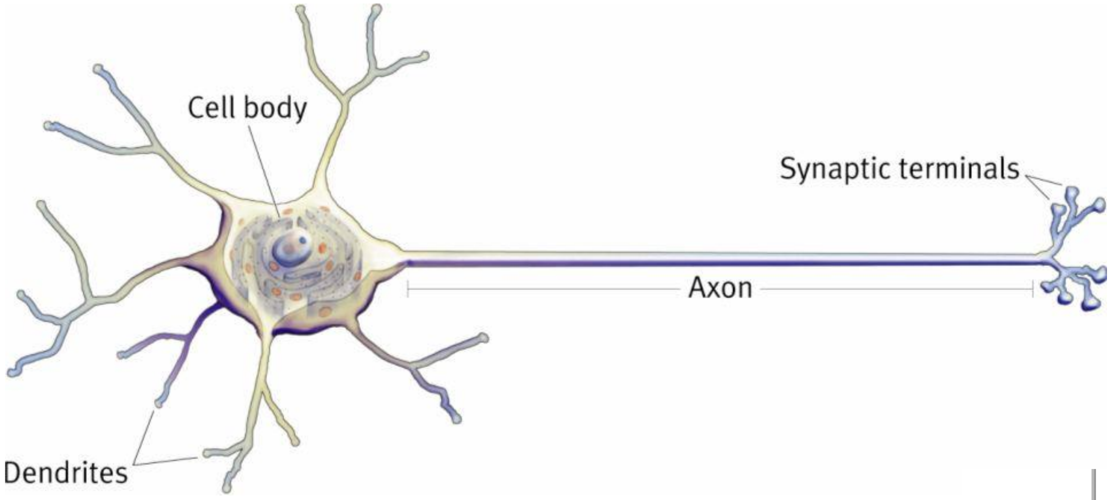


N points can be perfectly fit with a *N-1* order polynomial

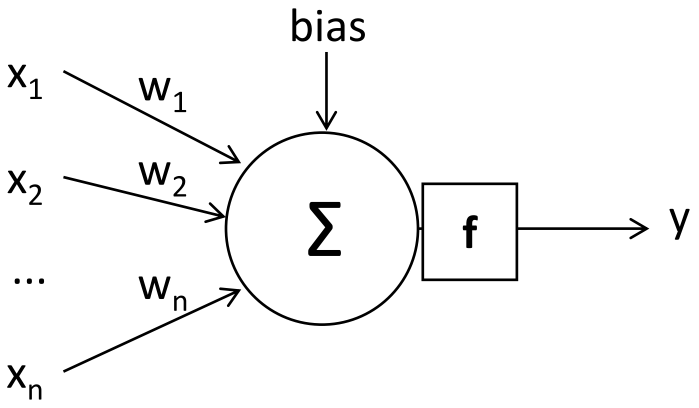
"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk"

John von Neumann

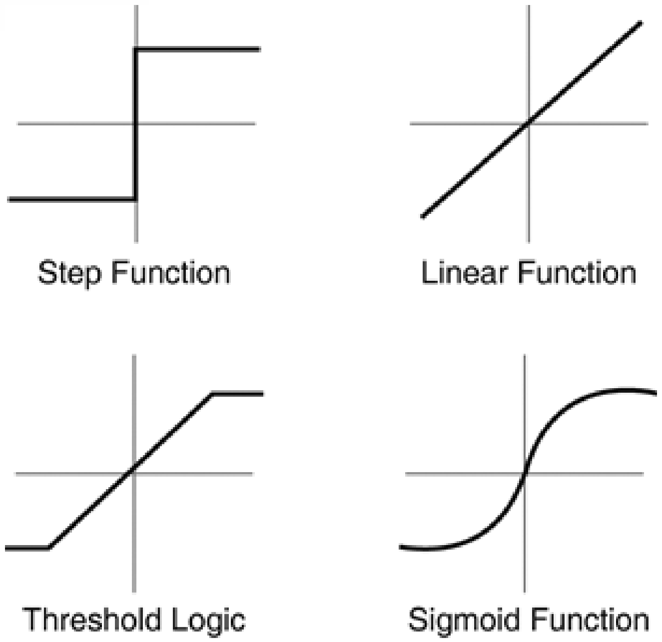
Artificial Neural Network (ANN)



A **neuron** fires if input signal is above a threshold



$f(x_1w_1 + x_2w_2 + ... + x_nw_n+b) = y$

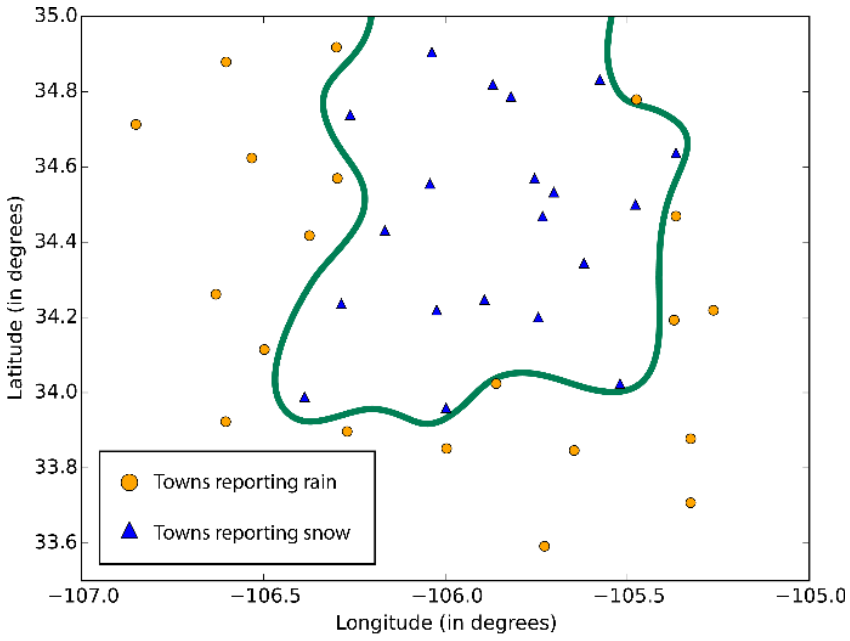


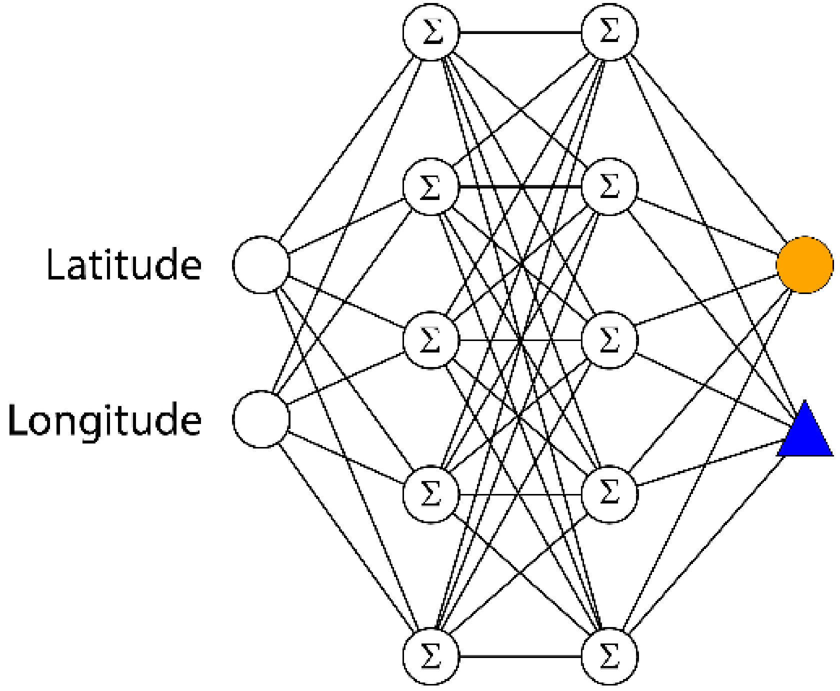
The **activation function *f*** can take several shapes

Artificial Neural Network (ANN)

Neurons can be arranged in **networks**

Hidden layers enable producing any complex boundary (in classification) or data fitting model (in regression)



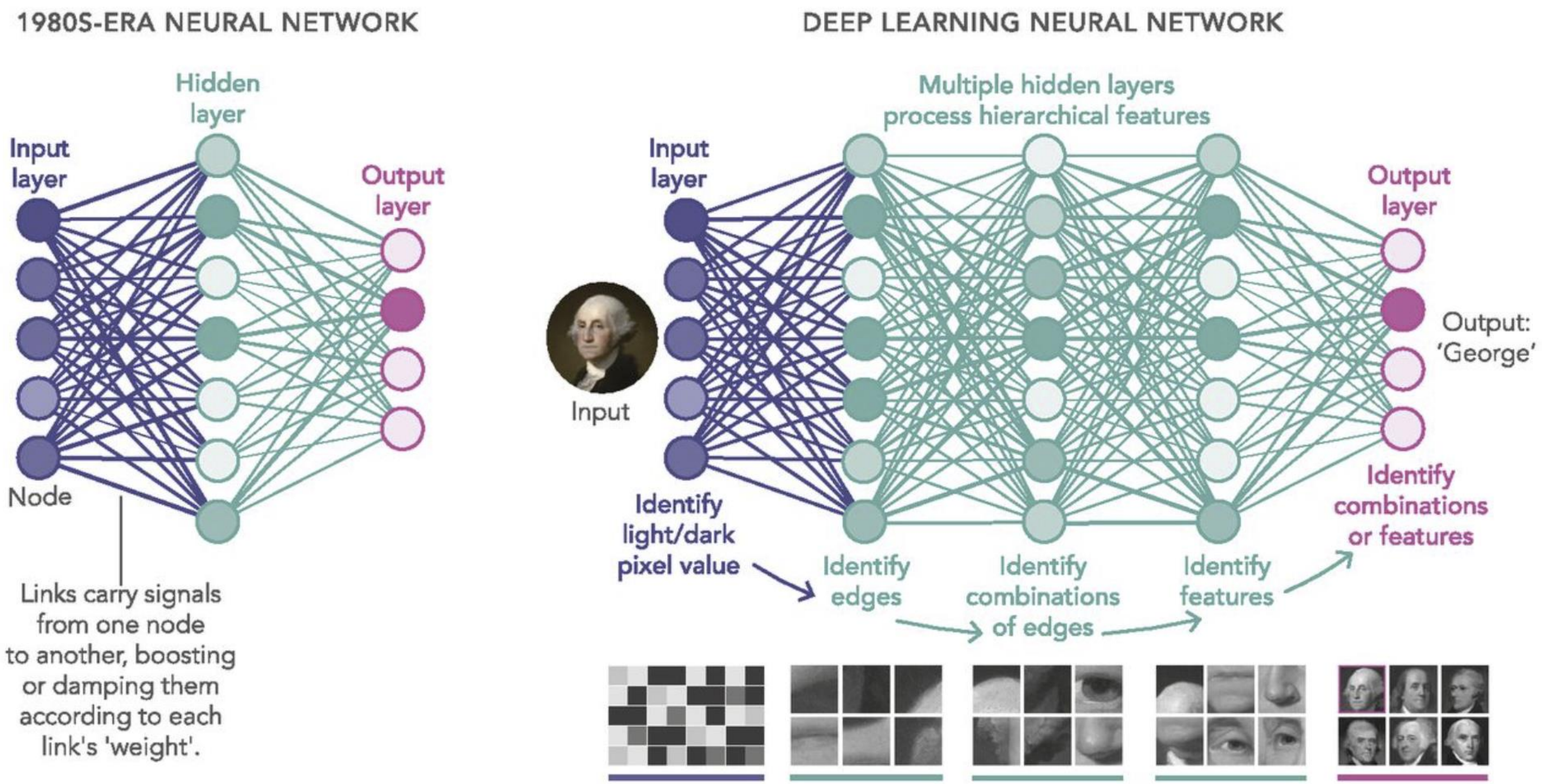


Training the ANN = finding synapses weights w_i minimizing error

ANN can fit any data, but are not easily interpretable

azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice

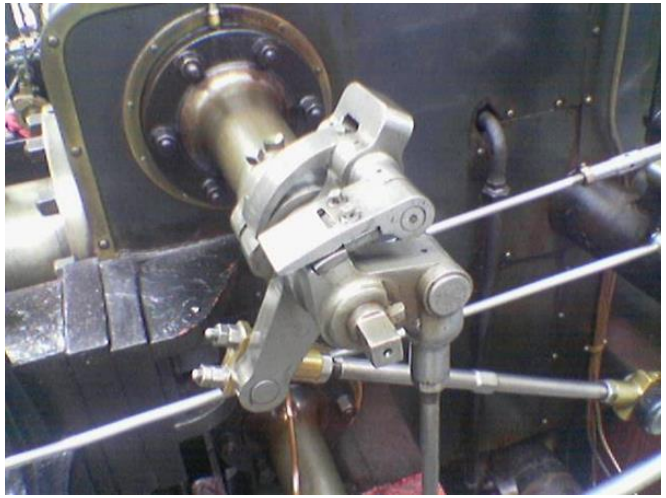
Shallow vs Deep Learning



M.M. Waldrop, *News Feature: What are the limits of deep learning?*, **PNAS**, 2019

[Extra] Convolutional neural networks

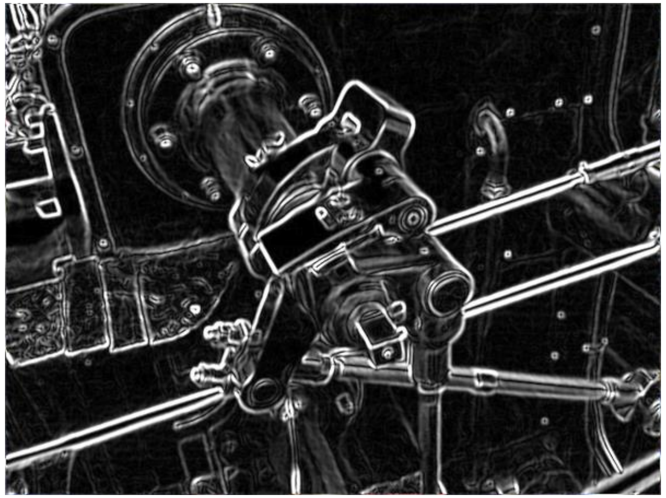
Convolution: a mathematical operation, “sliding a filter” (kernel) over the signal. *Example, edge detection:*



$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathbf{A}$$

and

$$\mathbf{G}_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \mathbf{A}$$



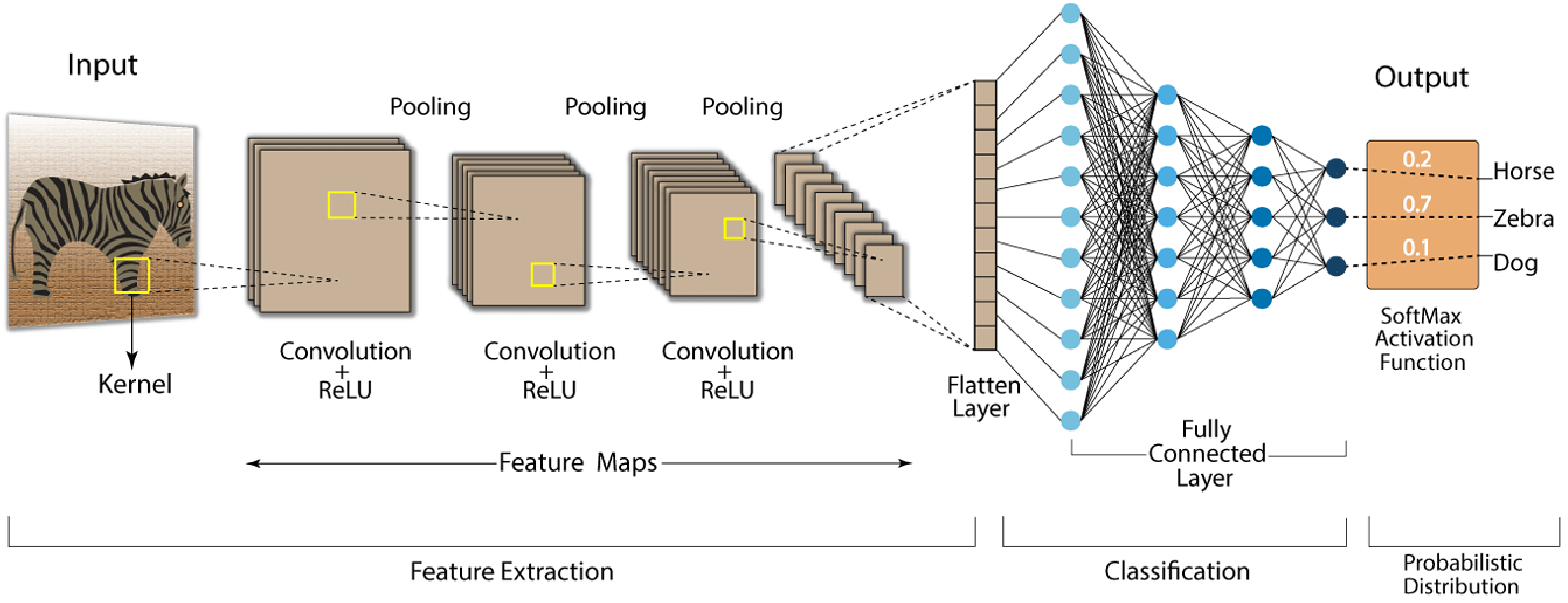
Convolutional neural network: instead of neurons, has many kernels.

Learning = optimising kernel weights

[Extra] Convolutional neural networks

Exploit local correlation in data (e.g. images, spectra, ...).

Can deal with inputs of arbitrary sizes with less parameters to learn



Problem with neural networks: interpretability

Conclusion

- Know what algorithms do, what their limitations are, and how their parameters may affect results

- Pick your algorithm depending on the nature of your data

- Better data often beats better algorithms