AI Whiteboarding

Machine Learning
Deep Learning
Data Science

Christian.schramm@dell.com

artificial intelligence / data science

machine learning

Samuel Arthur (1959):
"Field of study that gives computers the ability to learn without beingexplicitly programmed"

data science

“extract knowledge and insights from data”

not mutually exclusive

findings of one can be fed into the other
e.g. explore customer base and act upon findings

artificial intelligence

three major distinctions:

supervised learning

known inputs („labelled data“), desired outputs („known target“)
apply derived knowledge to new data

unsupervised learning

unlabelled data, unknown target, „make sense of data“

anomaly detection: monitoring, credit card fraud

reinforcement learning

touches both supervised and unsupervised learning
basic components: environment, agent, action, reward

supervised learning

the simplest of the three and the most common

regression

output: numerical value

classification

output: probability of being part of a category

cat or dog?

training algorithm

data

feed

model (e.g. linear regression)

repeat compare

objective function: evaluate result (e.g. SL: loss, RL: reward)

vary

optimization algorithm (e.g. gradient descent)

number of repeated steps = „epochs". fixed value versus adaptive stopping

three kinds of data

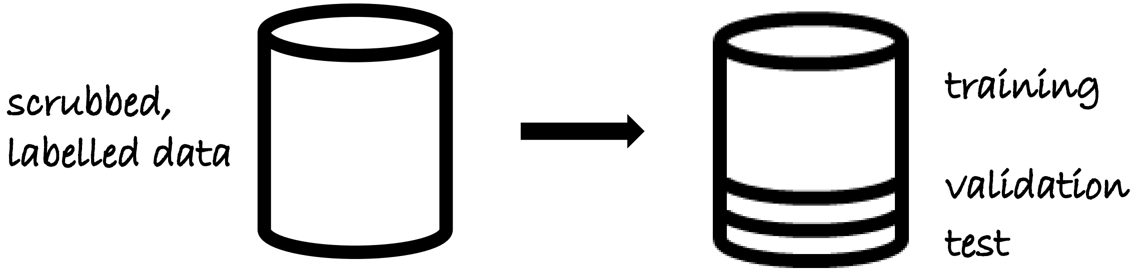
numerical: weight, height, dollar-euro-exchange rate, ...

ordinal: first, second, third, ...

categorical: france/italy/spain, cats/dog, large/medium/small, ...

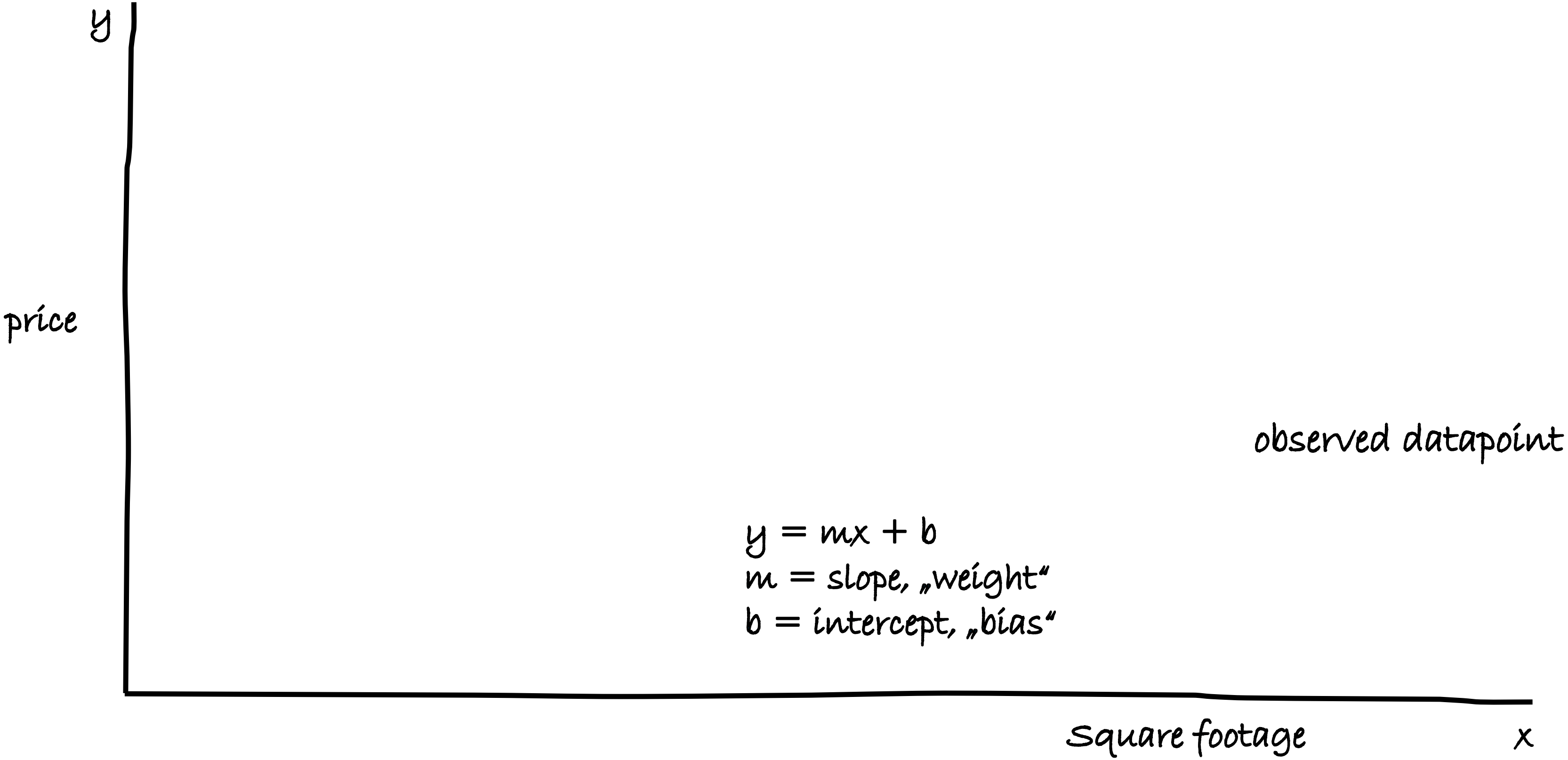
machine learning likes numerical data best

train test split



only one training-validation set

supervised learning: (linear) regression

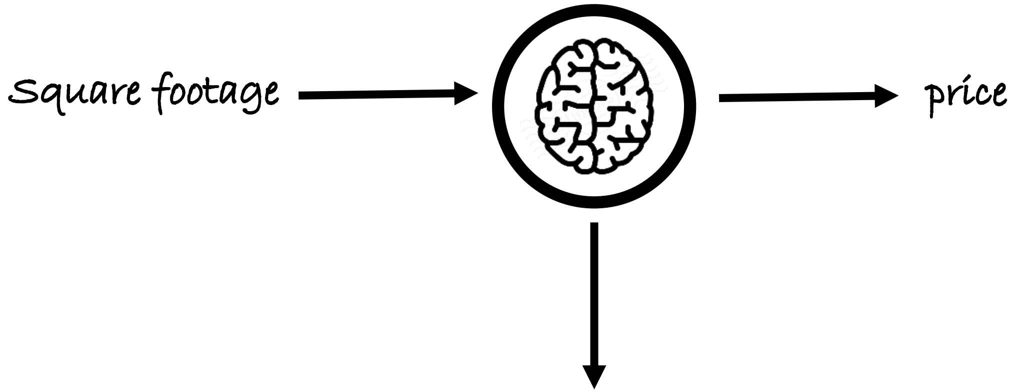


$y = mx + b$

m = slope, „weight“

b = intercept, „bias“

linear regression

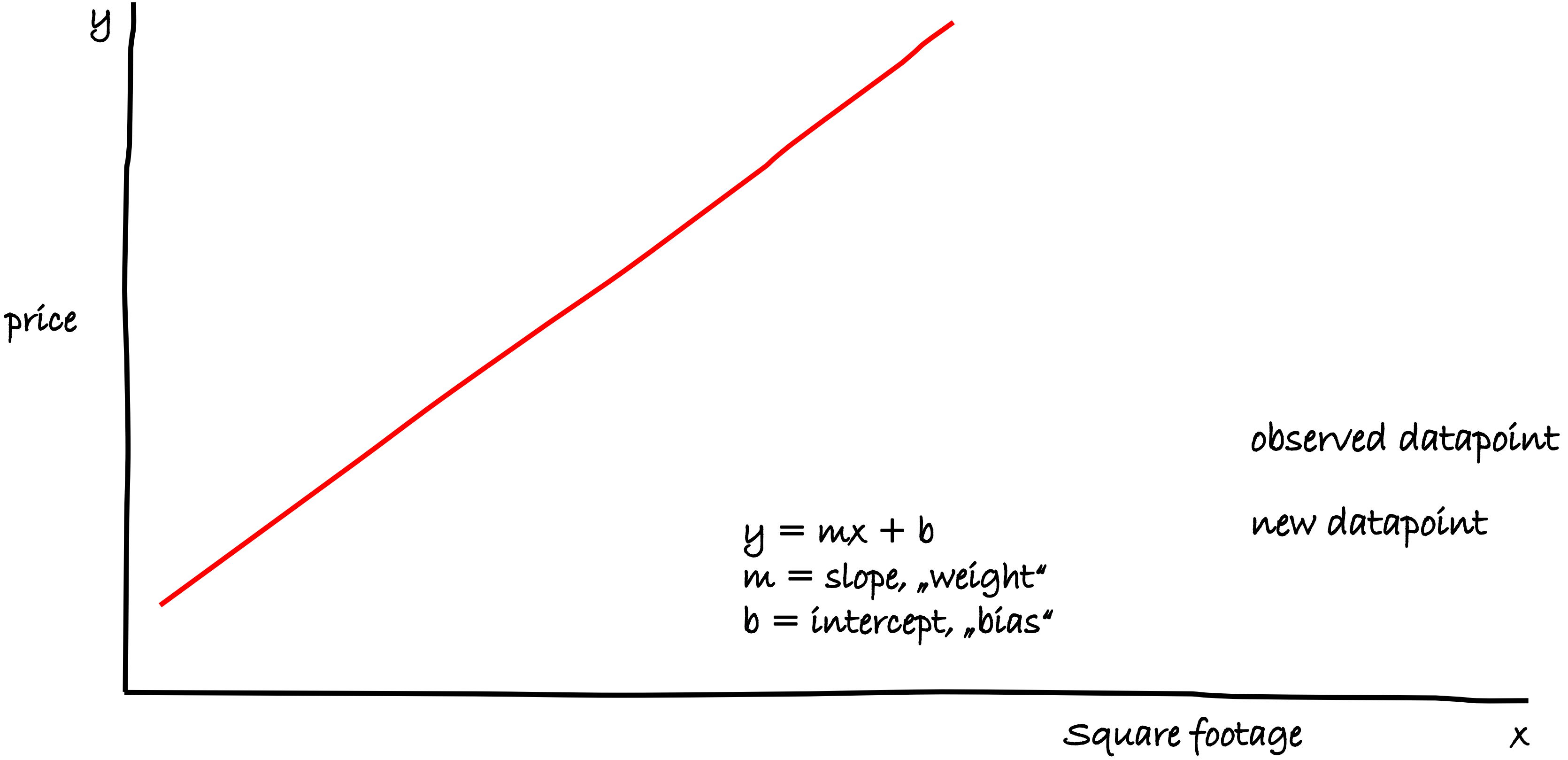


$y = mx + b$

m = slope, „weight“

b = intercept, „bias“

supervised learning: (linear) regression

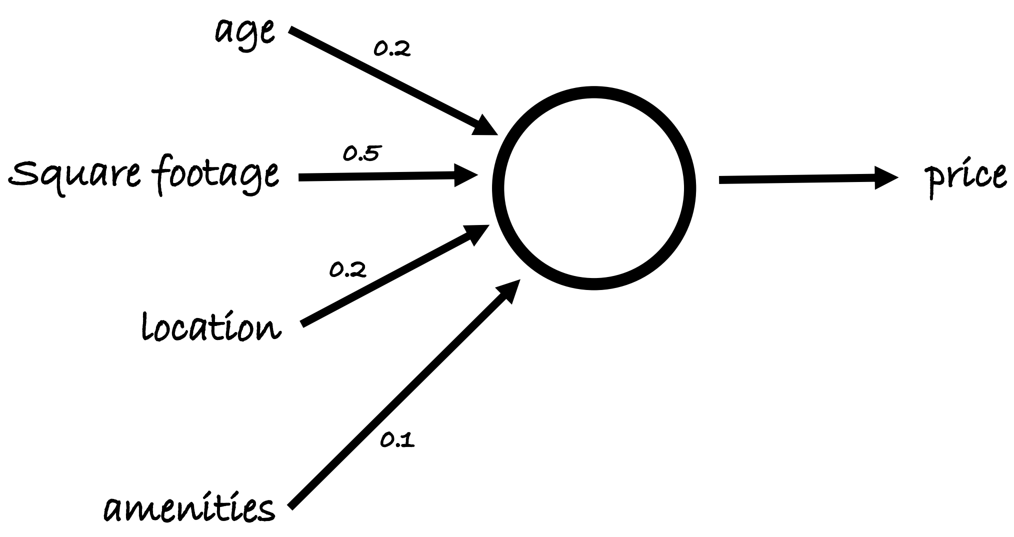


$y = mx + b$

m = slope, „weight"

b = intercept, „bias"

multiple linear regression (with weights)



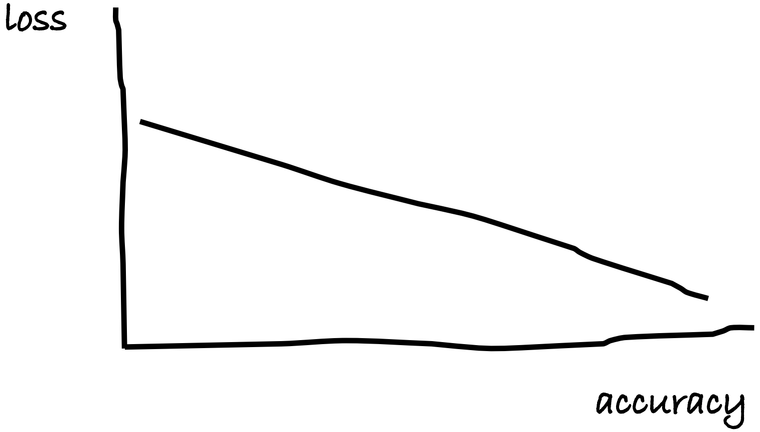
0.2 x age + 0.5 x square f. + 0.2 x location + 0.1 x amenities = price

we aim to predict precisely on prices in the future

objective function

evaluate model: does the model's output match the desired correct values?

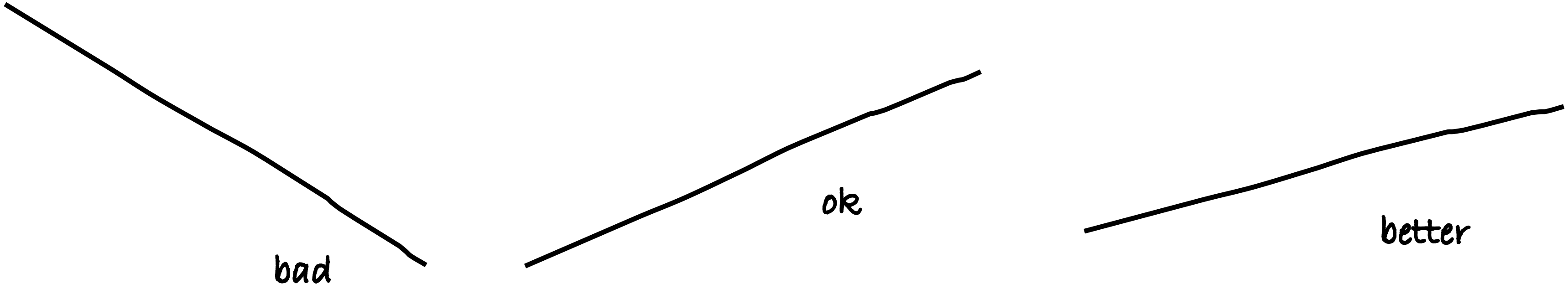
„loss" / „cost" in supervised learning
„reward" reinforcement learning, the exact opposite



in supervised learning:
regression: l2-norm
classification: cross-entropy

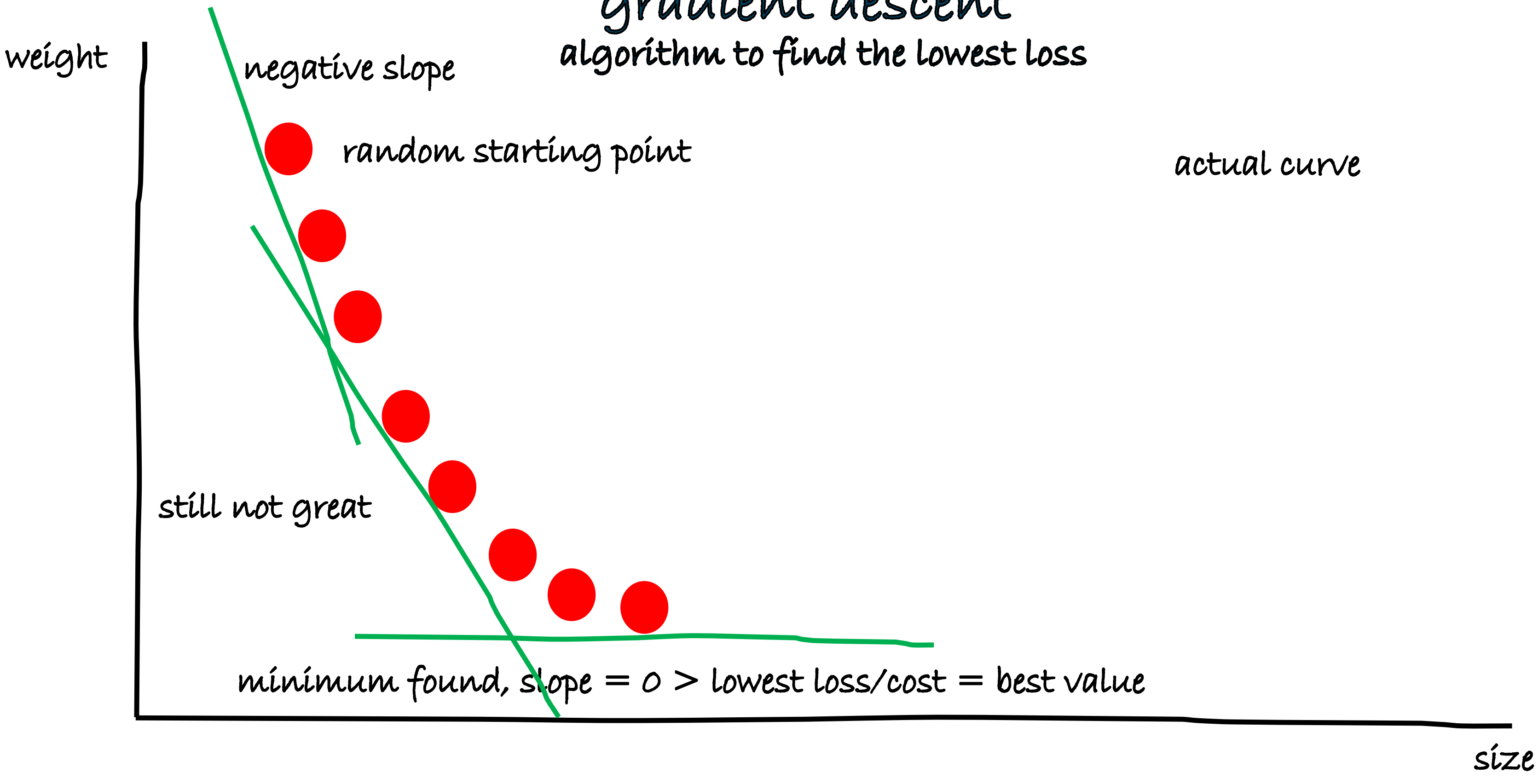
loss / cost function
how good is the model?

Sum(y-yhat)^2 > min
y = observed value
yhat = model value

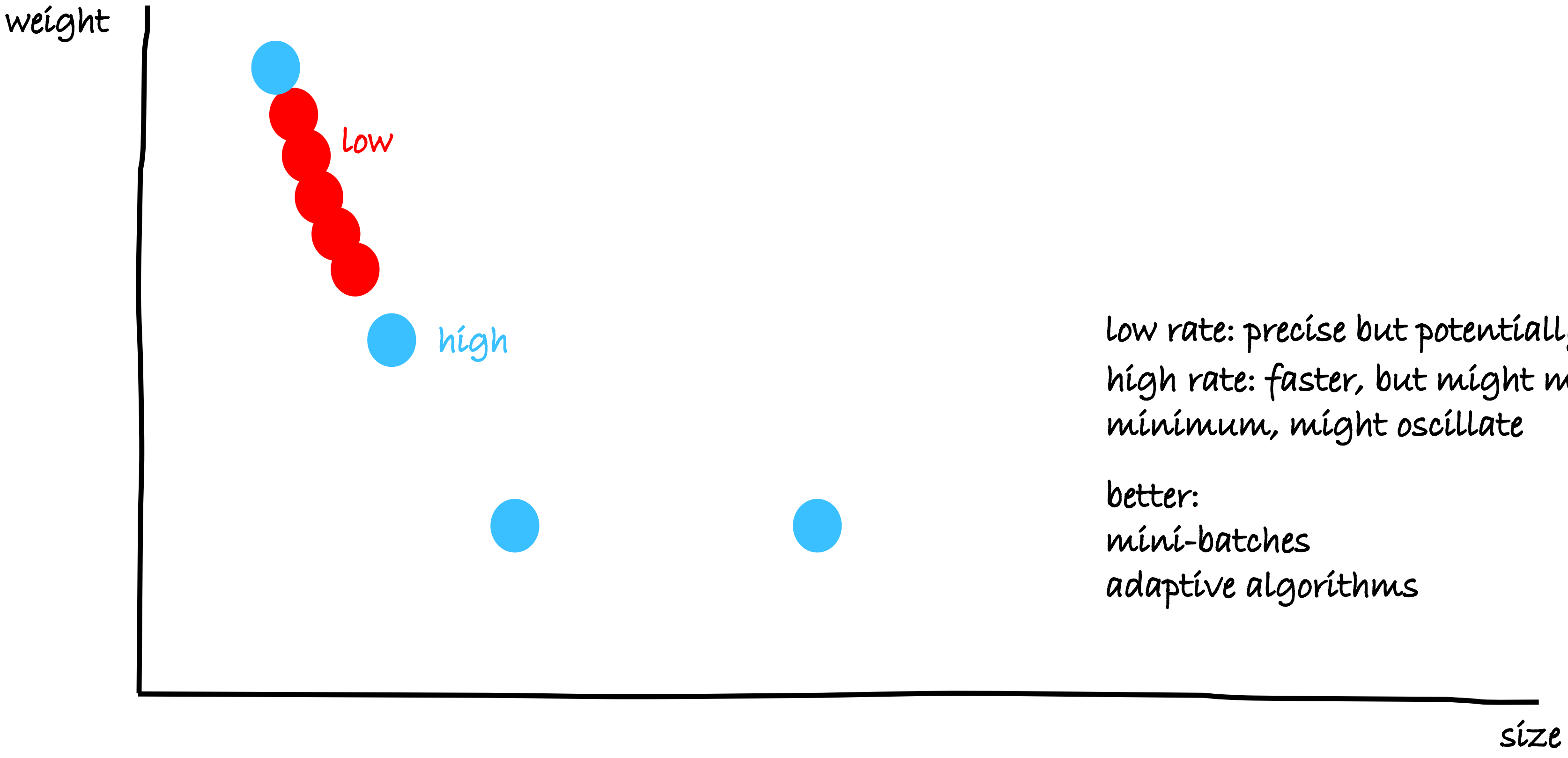


gradient descent

algorithm to find the lowest loss



learning rate



low rate: precise but potentially slow

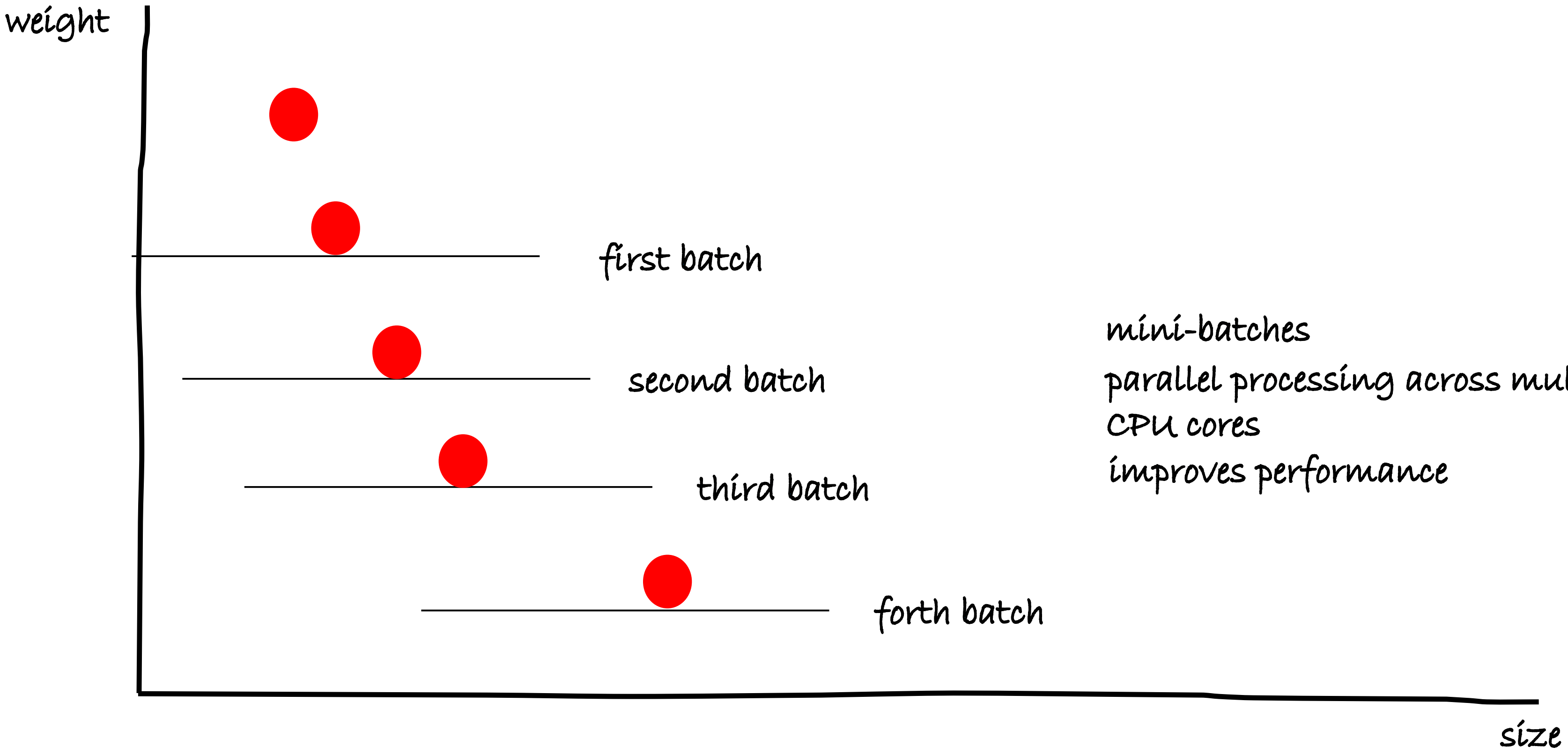
high rate: faster, but might miss the minimum, might oscillate

better:

mini-batches

adaptive algorithms

stochastic gradient descent



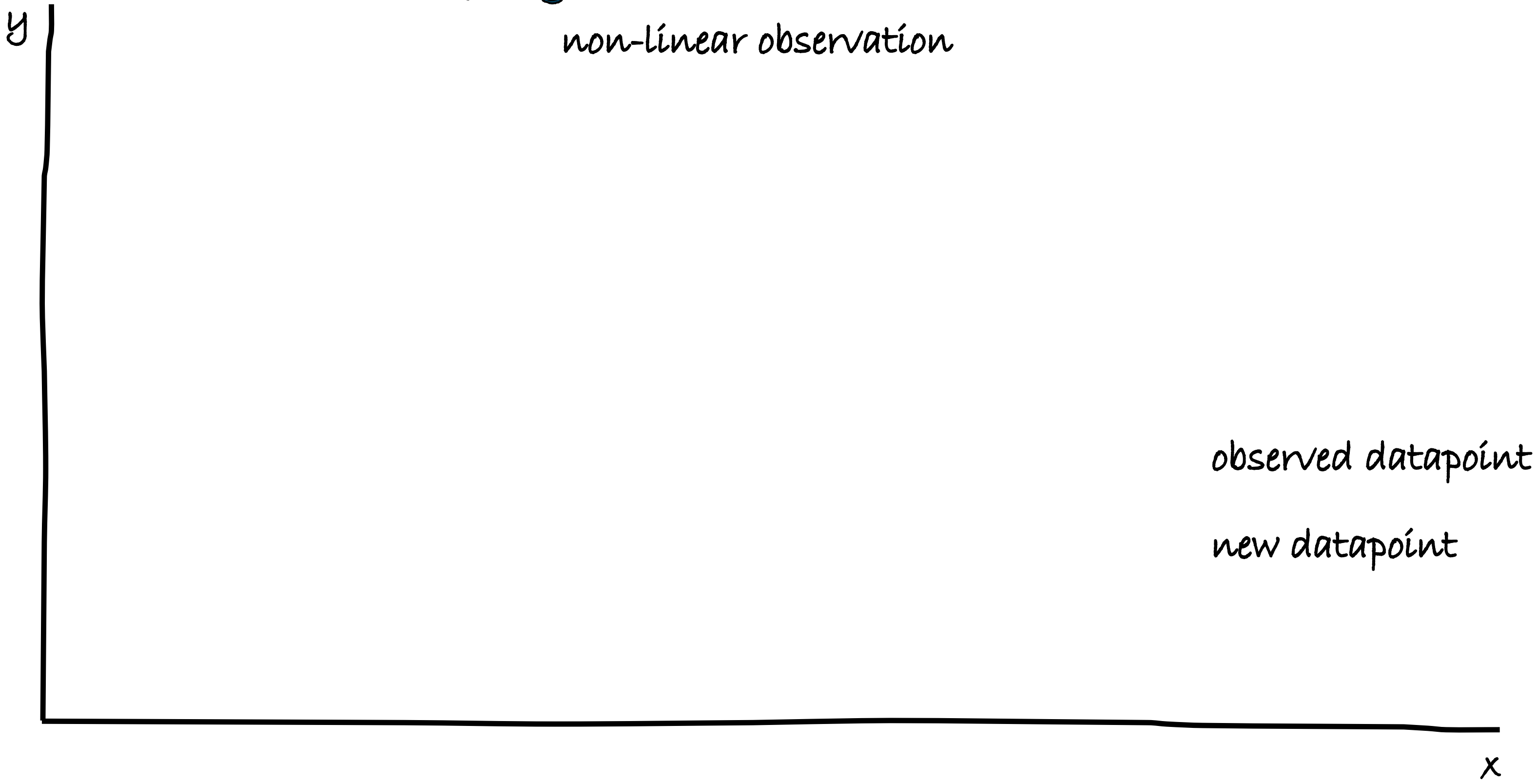
mini-batches

parallel processing across multiple CPU cores

improves performance

polynomial regression

non-linear observation



underfitting

too generic, not focused on training data

overfitting

too specialized, too focused on training data

y

observed datapoint

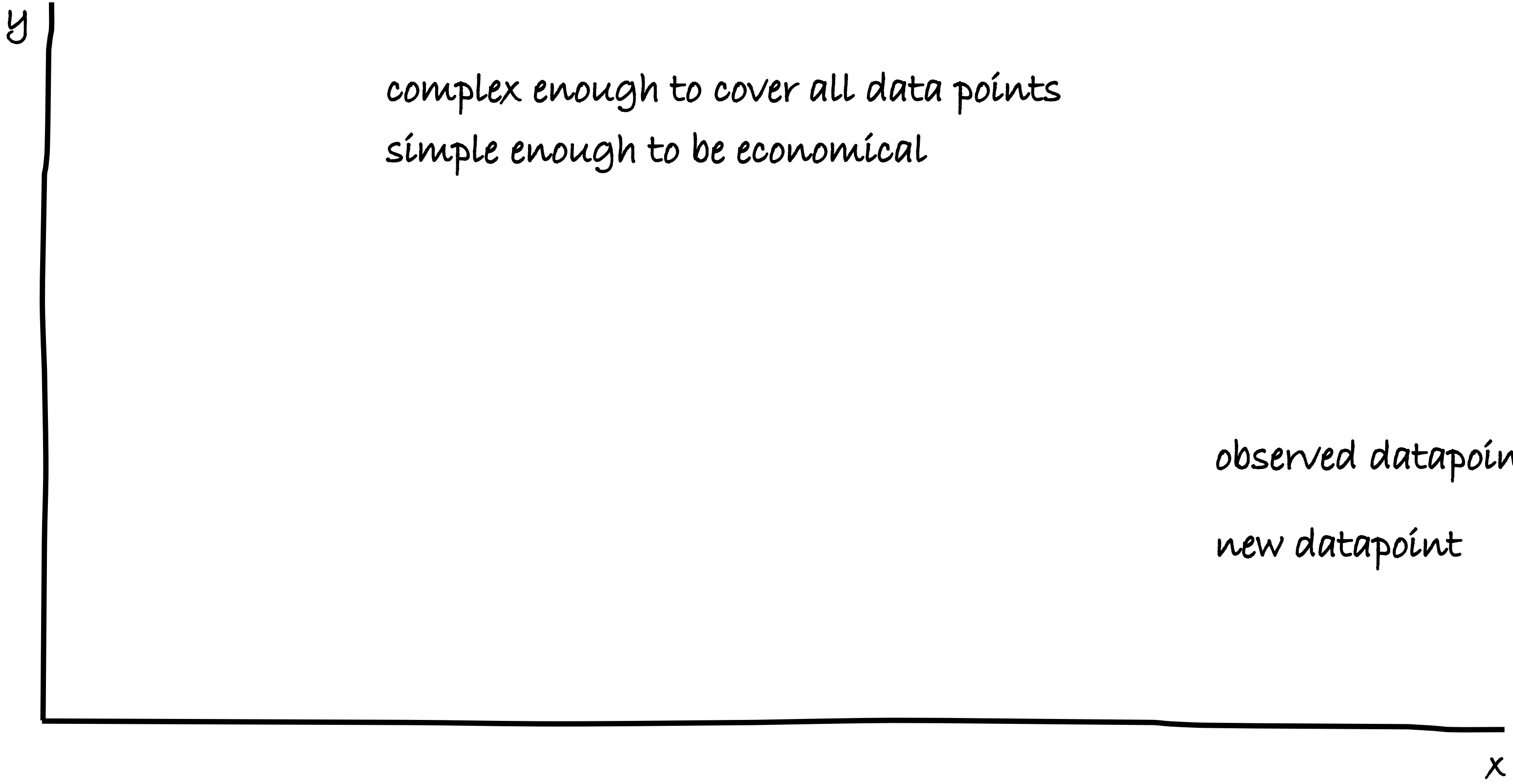
new datapoint

x

fitting

complex enough to cover all data points

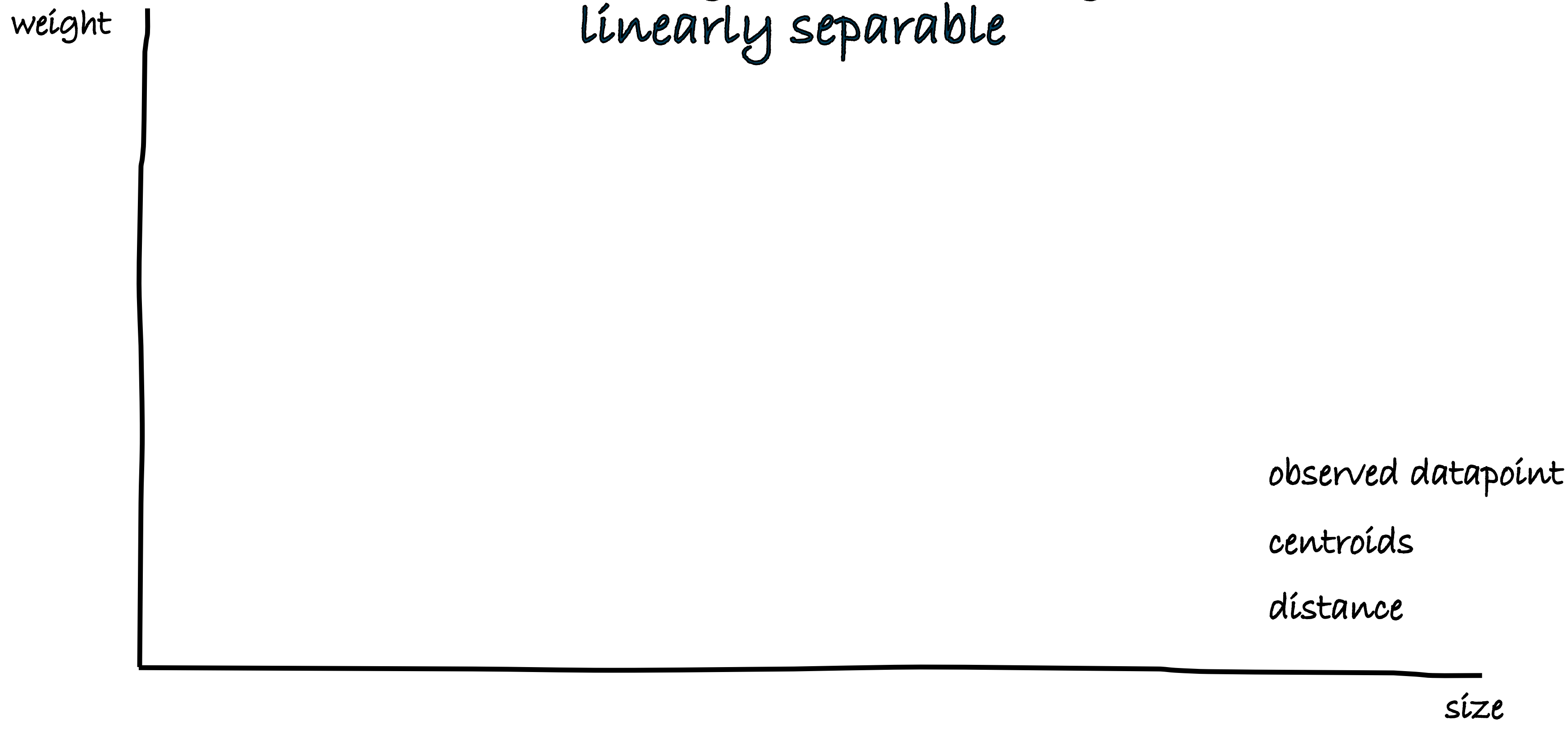
simple enough to be economical



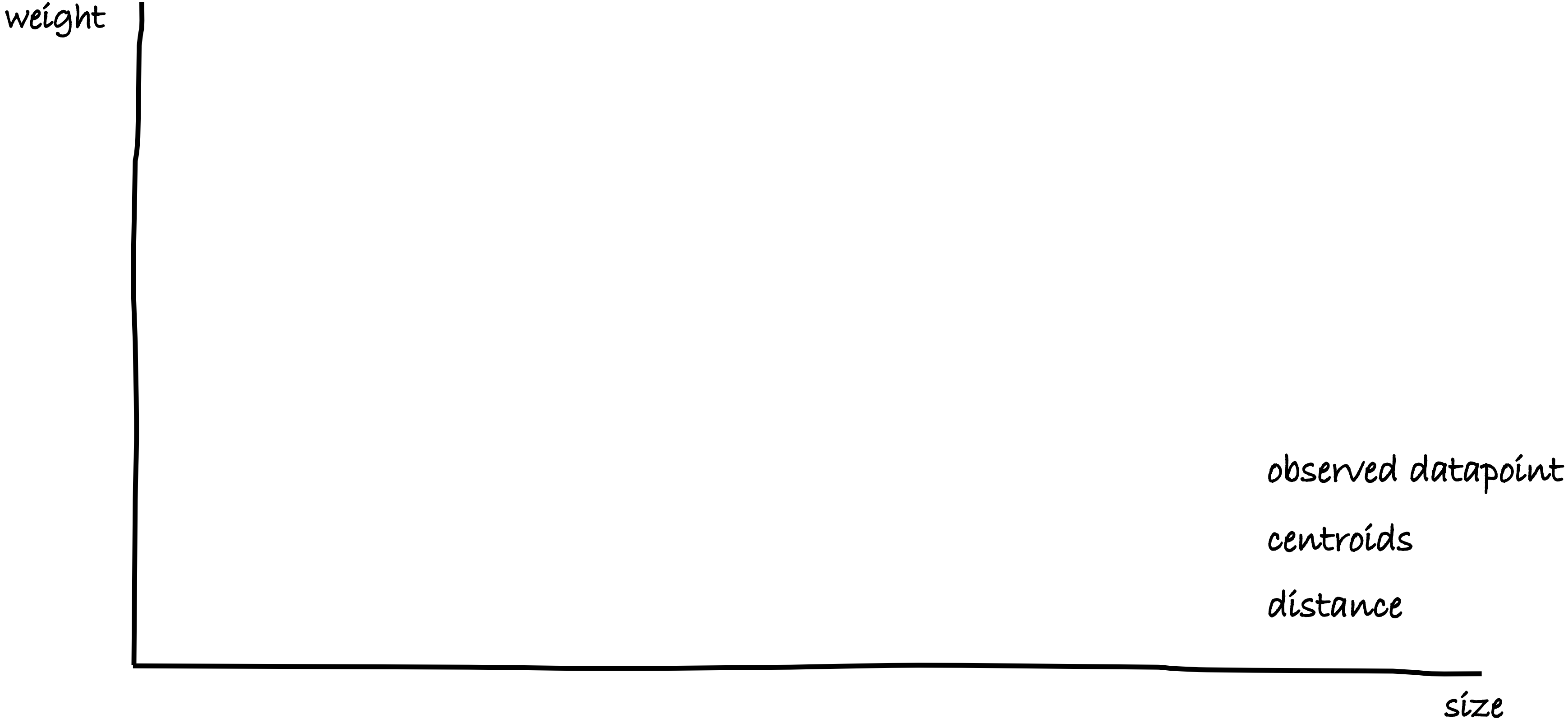
observed datapoint

new datapoint

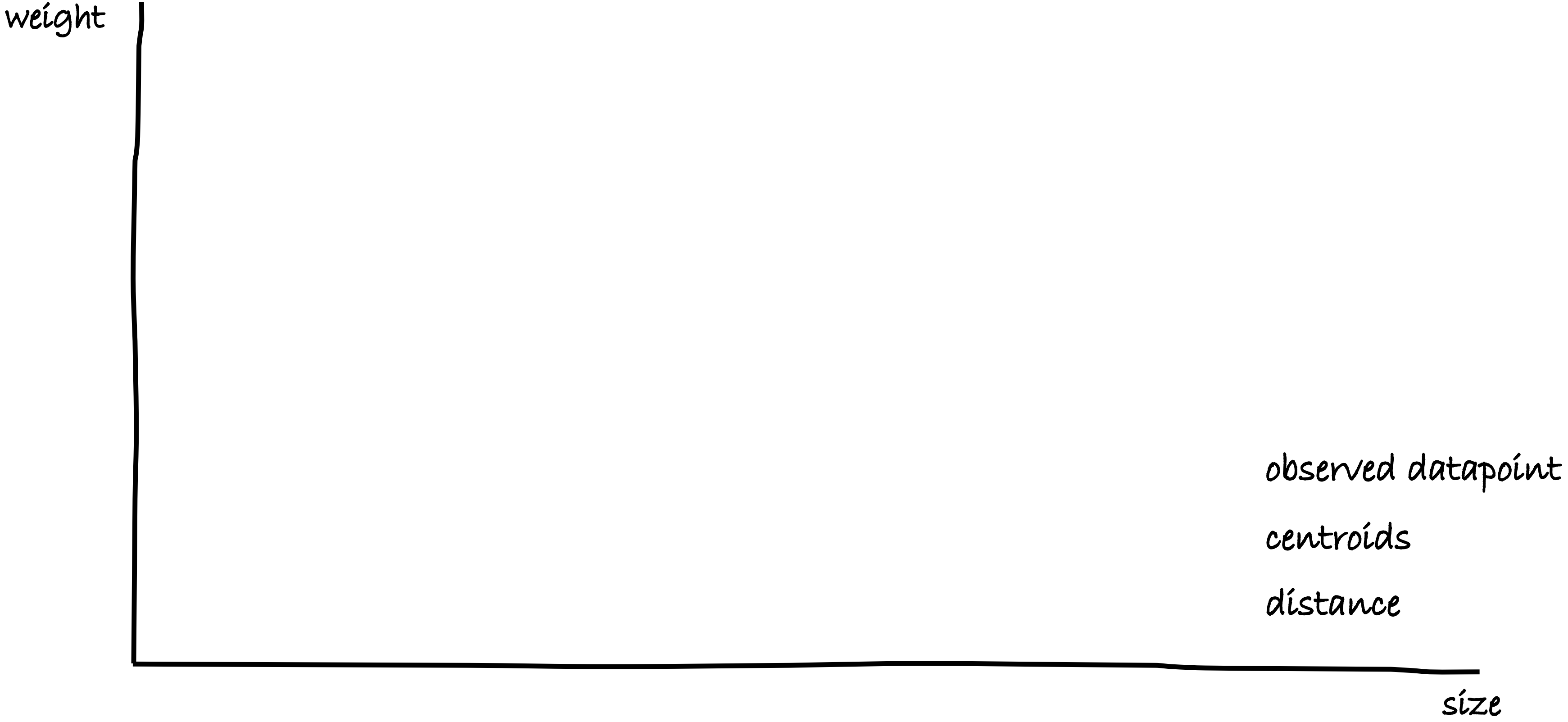
UL: clustering: k-nearest neighbor
linearly separable



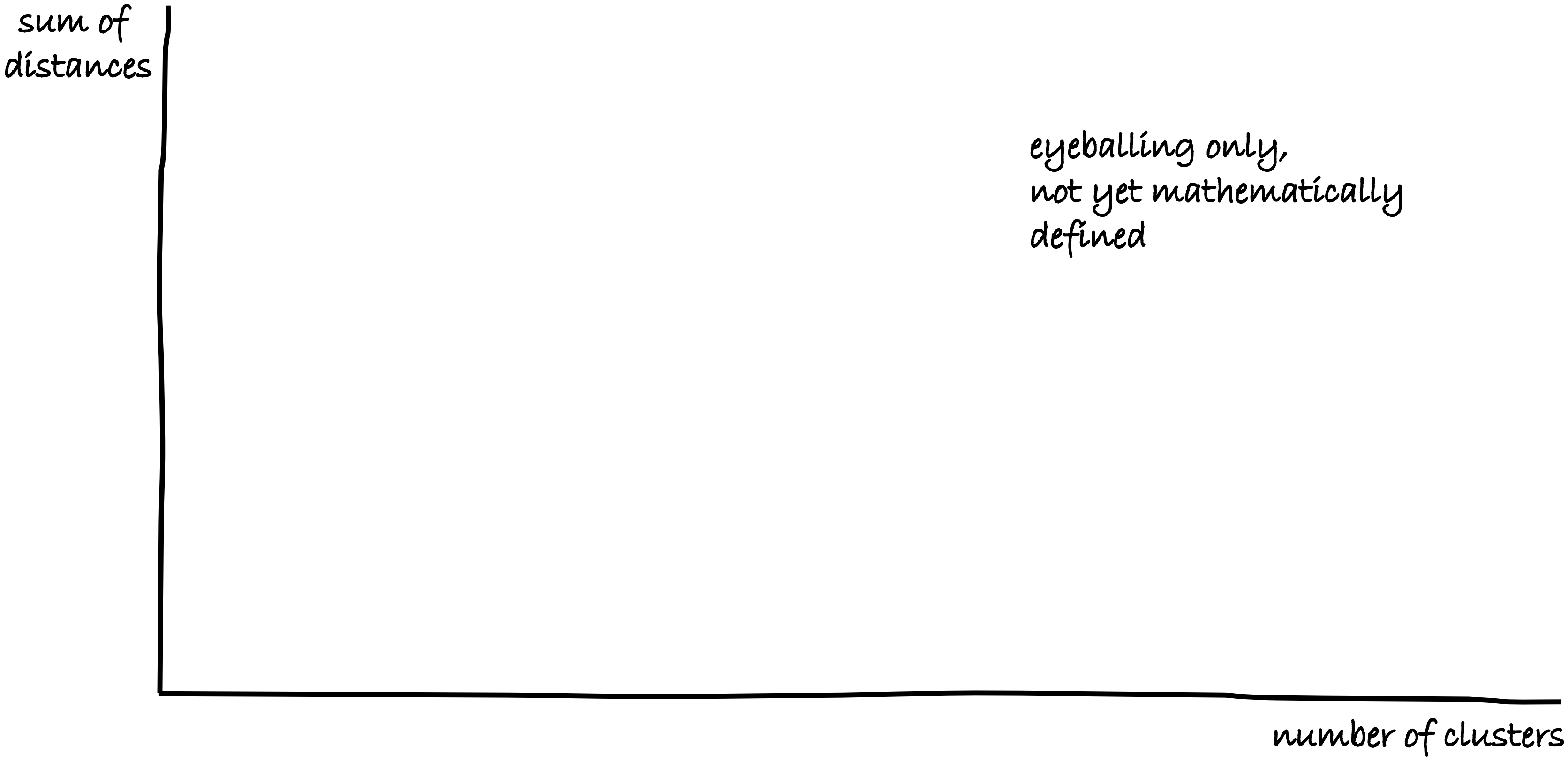
UL: k-nearest neighbor



UL: k-nearest neighbor

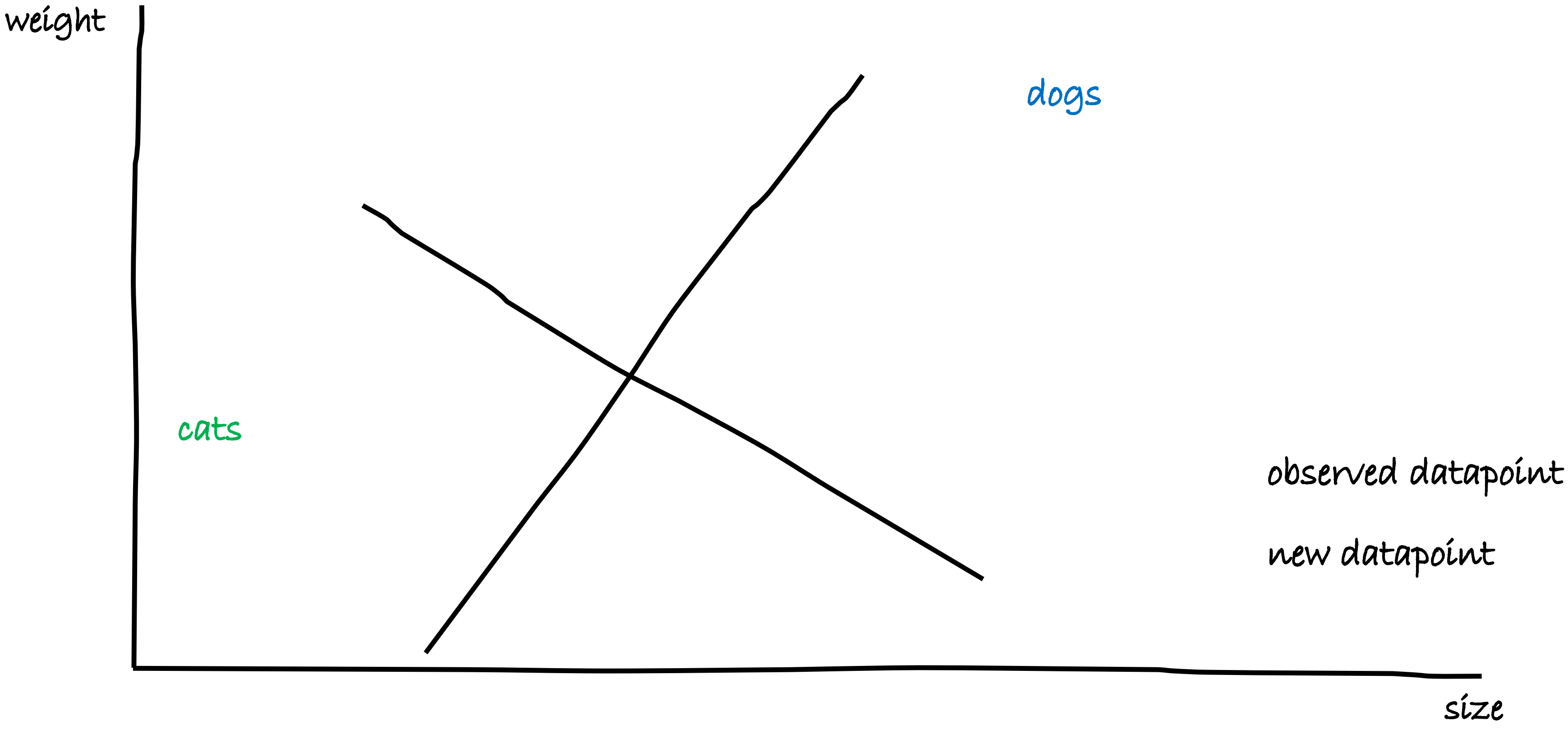


UL: elbow method

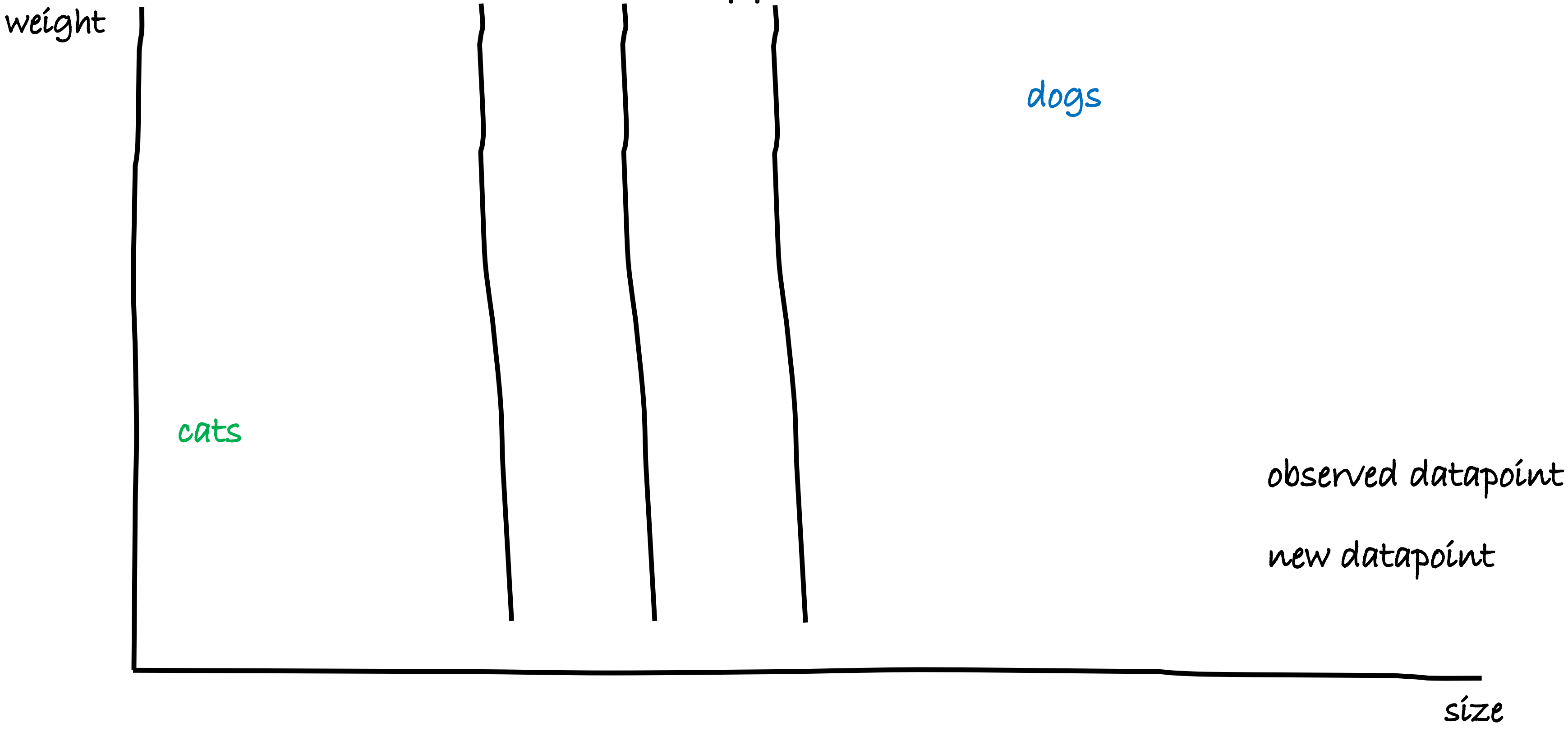


eyeballing only,
not yet mathematically
defined

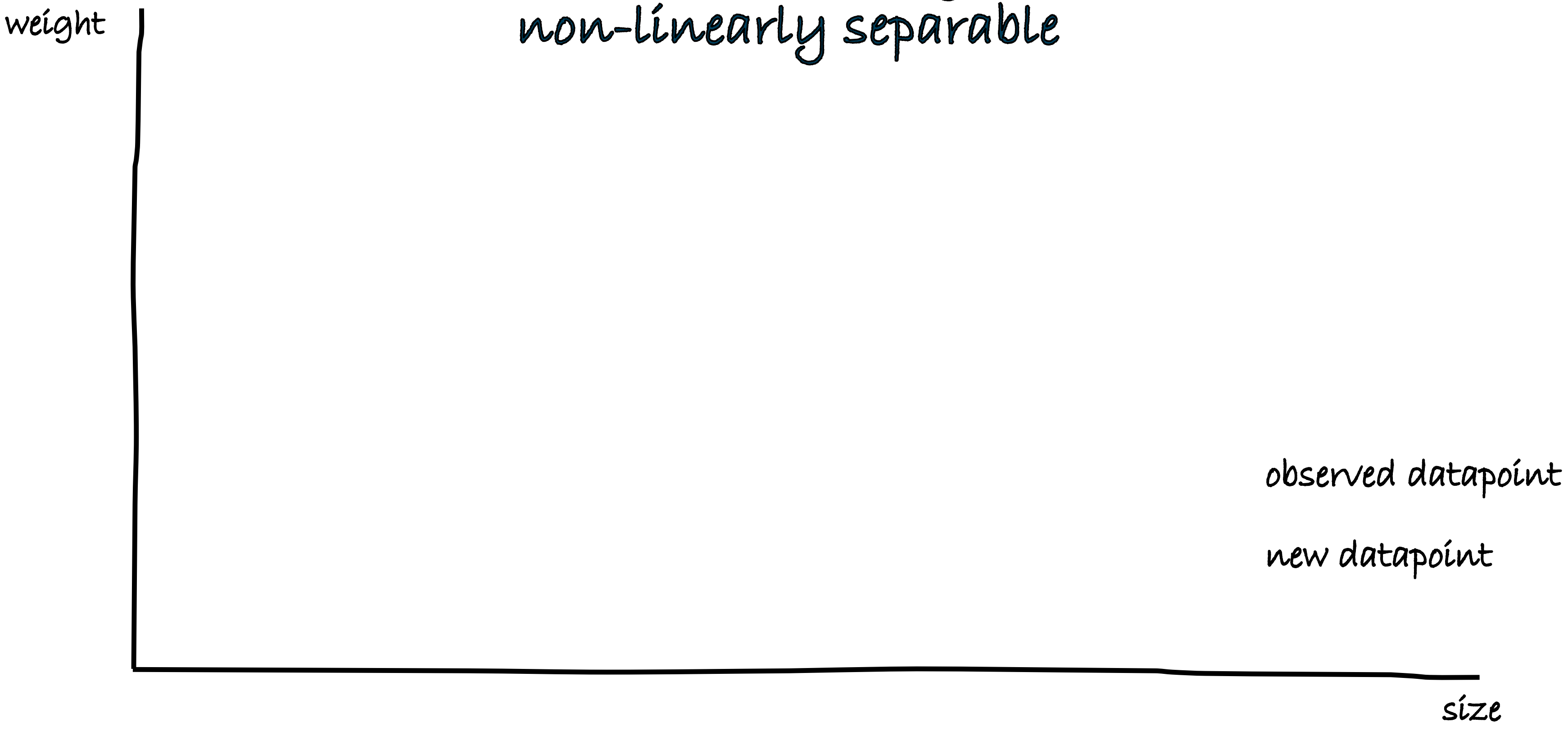
UL: clustering



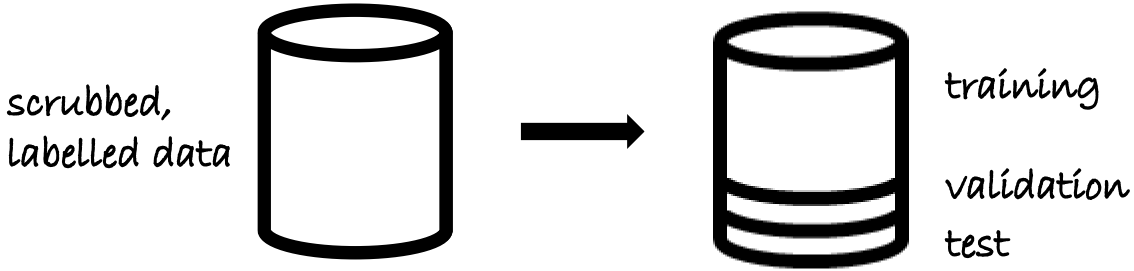
UL: support vectors



UL: clustering
non-linearly separable



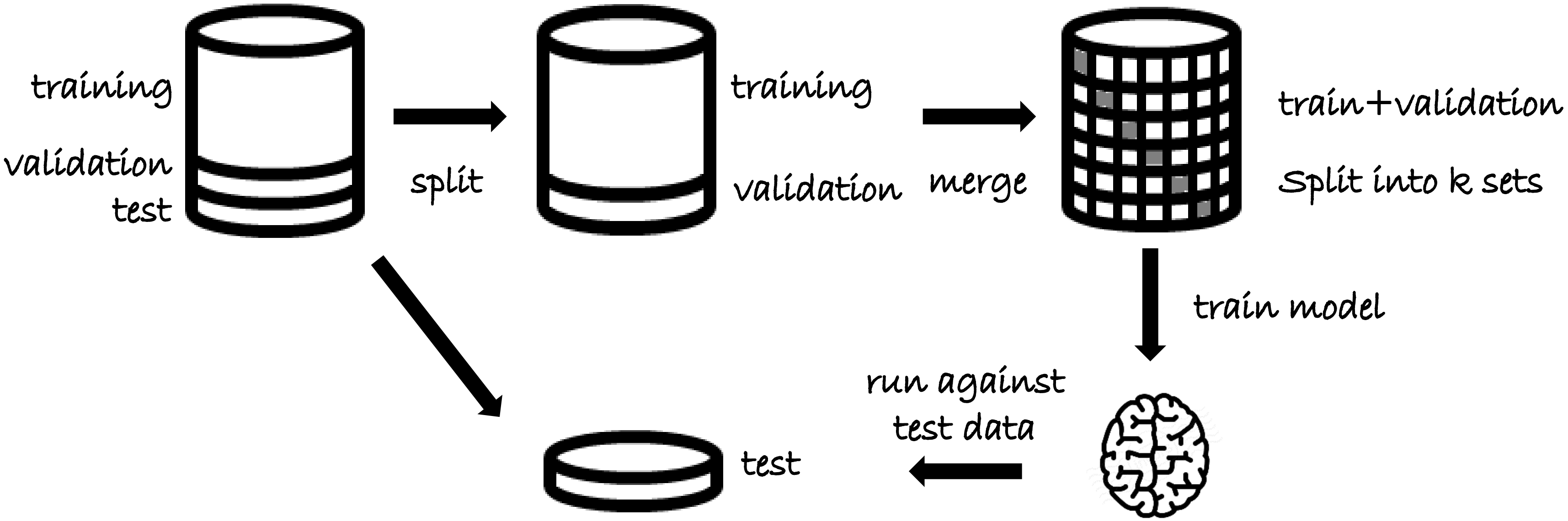
recap: train test split



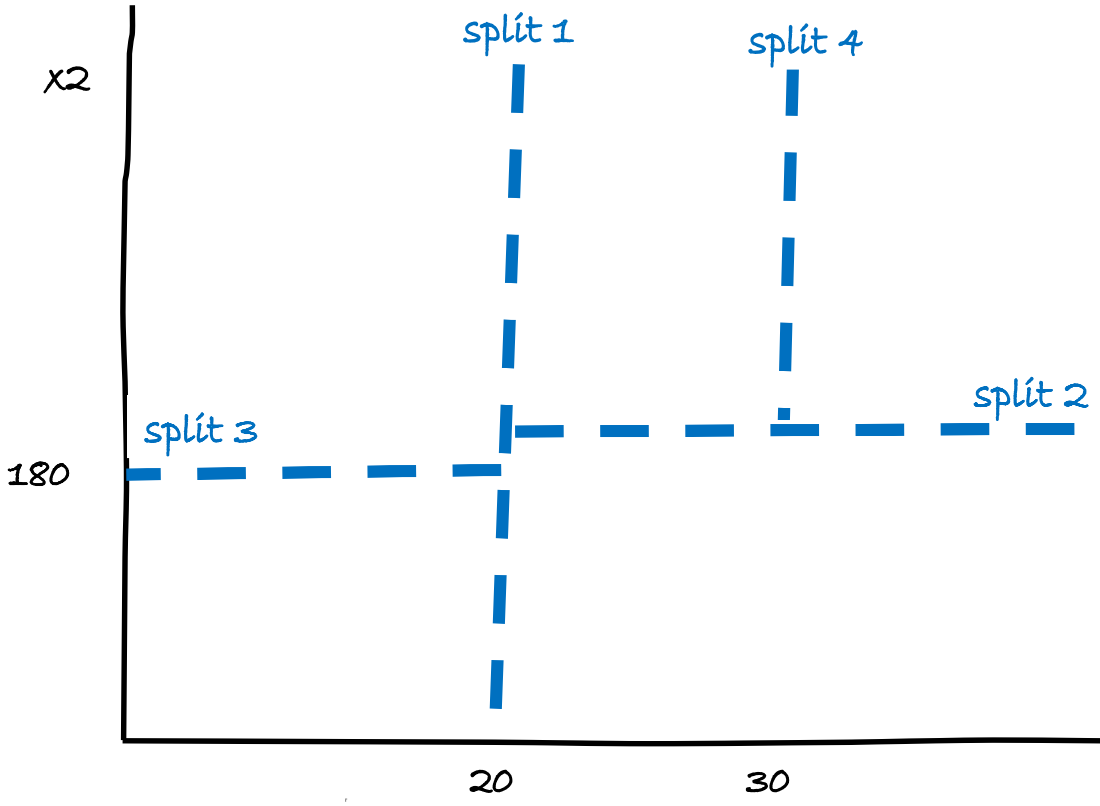
only one training-validation set

k-fold cross validation

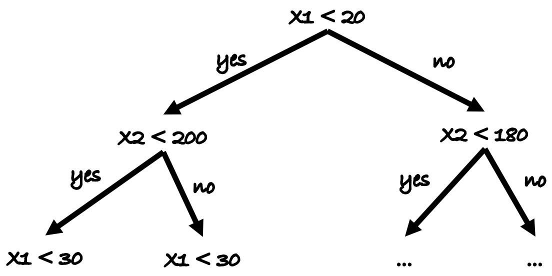
subsetting train-test data



decision tree

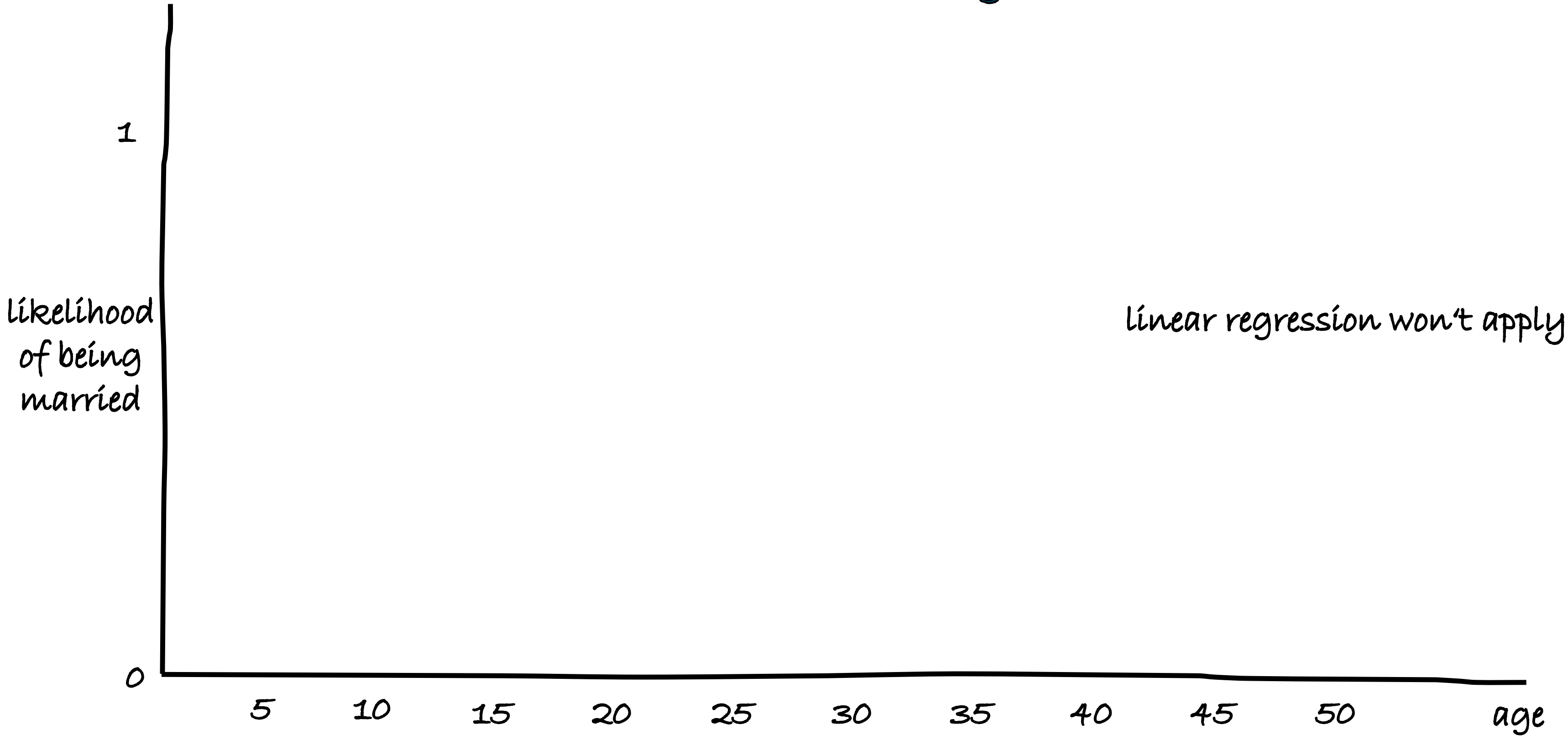


dependent variable y
is a 3D variable

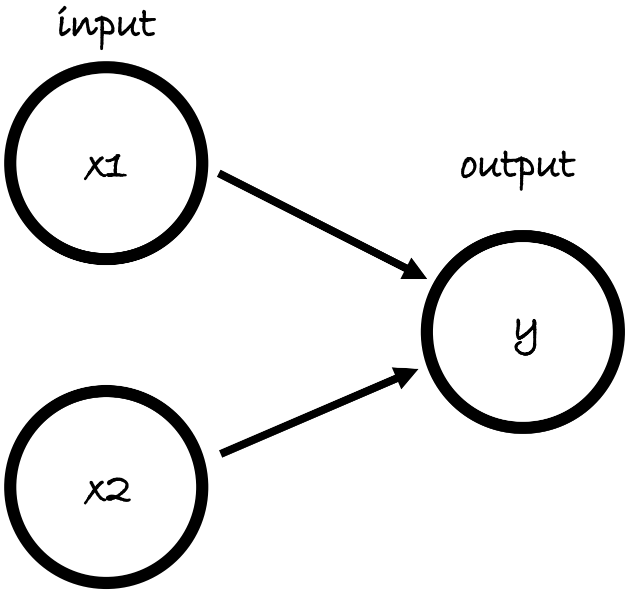


$$y = x1 * x2$$

logistic regression for (binary) classification



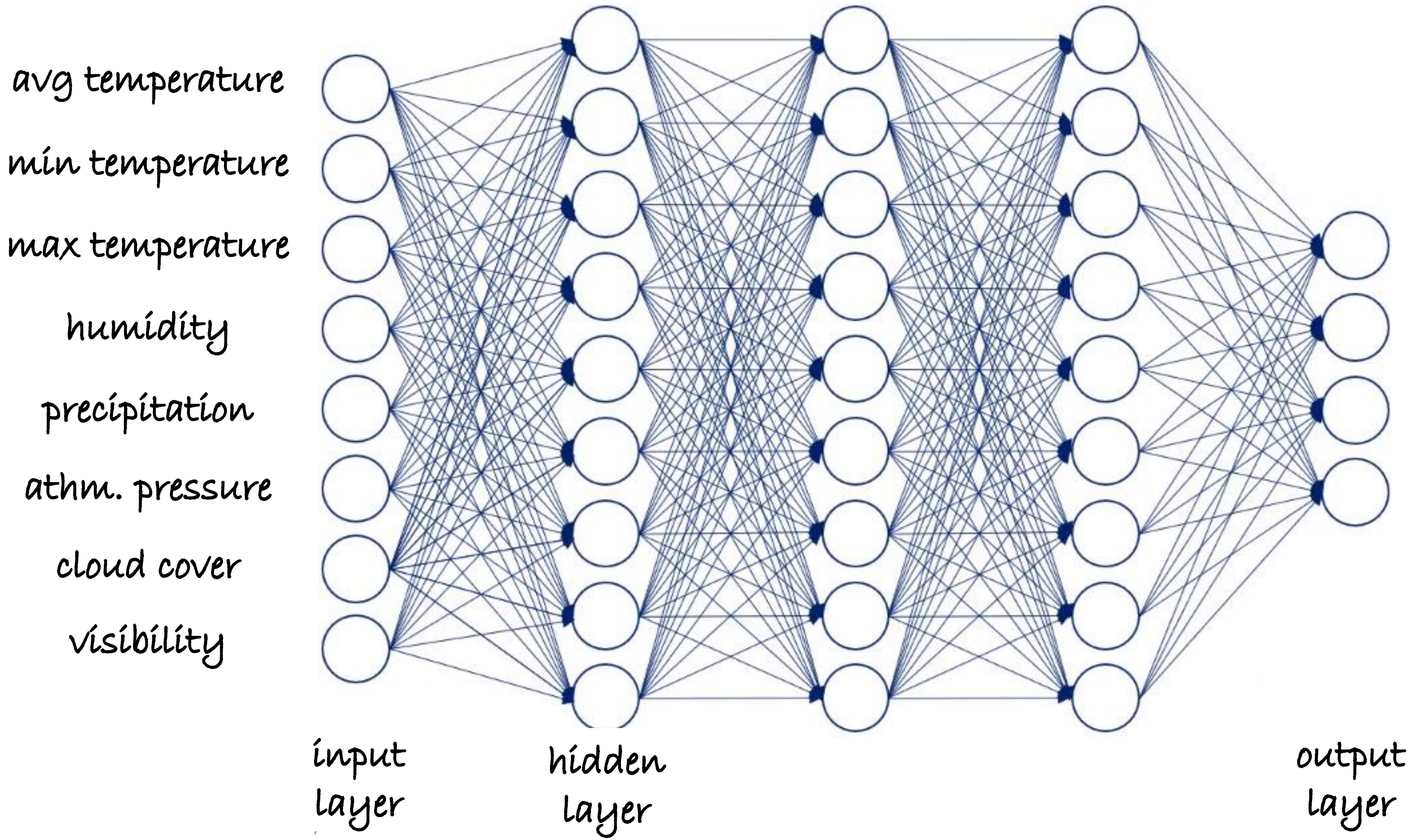
deep learning: linearity



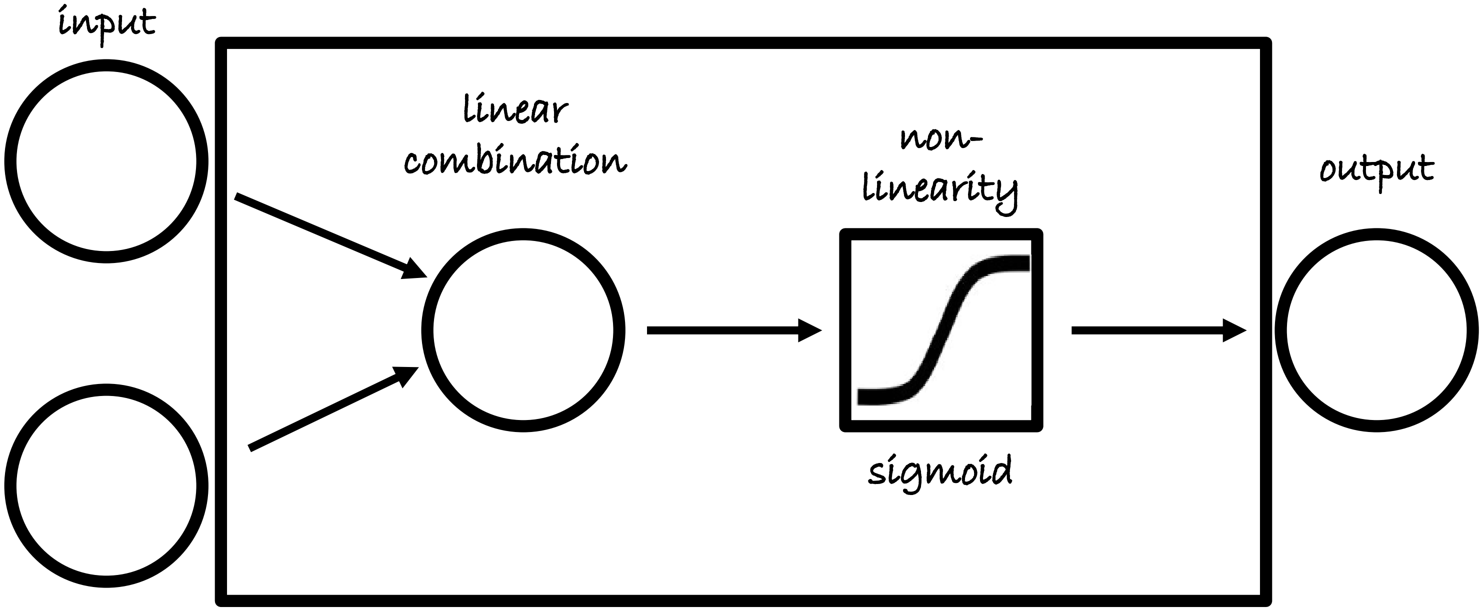
linear model
Sum(y-yhat) ^2 > min
y = observed value
yhat = model value

most real-life scenarios are
not linear

DL: deep net


avg temperature
min temperature
max temperature
humidity
precipitation
athm. pressure
cloud cover
visibility
input layer
hidden layer
output layer

DL: combining linearity and non-linearity

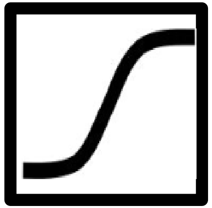


Layer = building block of neural networks

hidden layers cannot be stacked without non-linearity

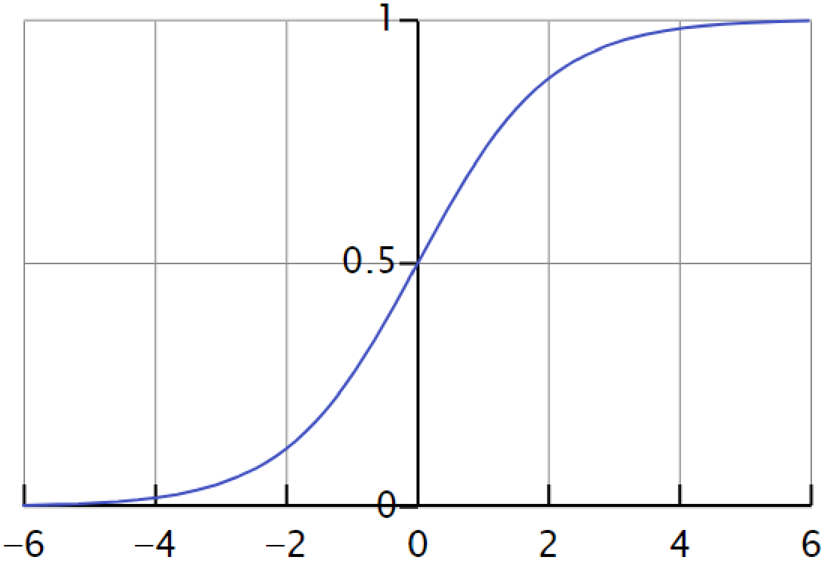
DL: sigmoid

non-linearity as basis for neural networks



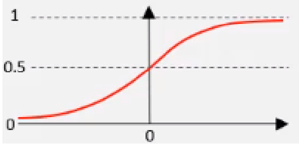
õ(x)=1/1+e^-x

sigmoid



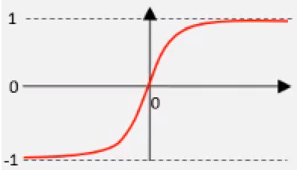
DL: activation functions

sigmoid
(logistic function)



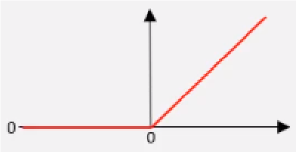
(0,1)

tanh
(hyperbolic tangent)



(-1,1)

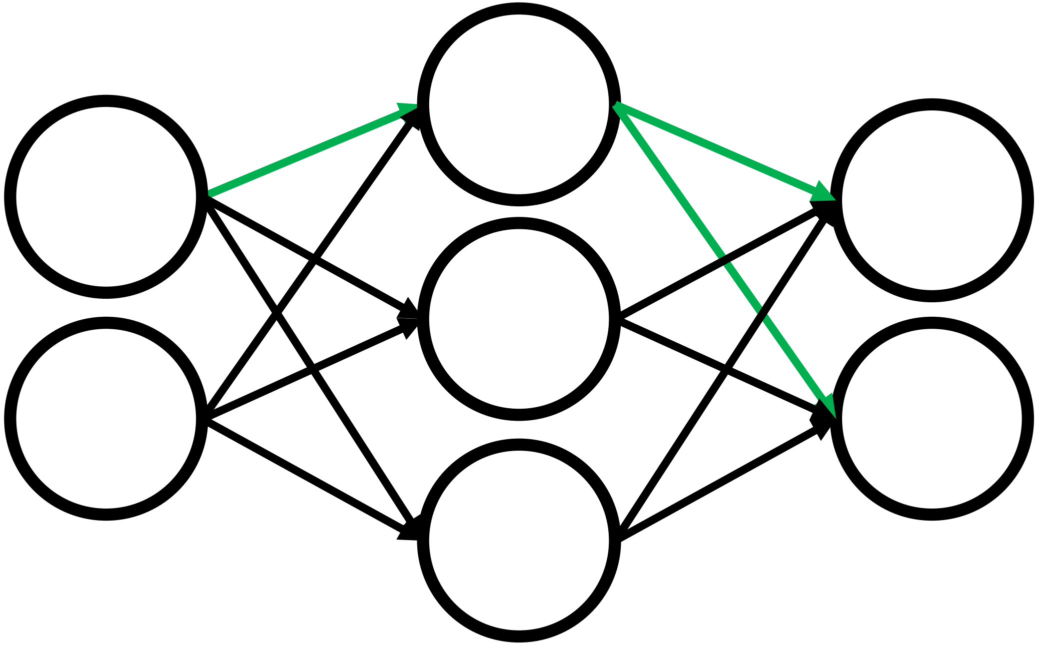
ReLU
(rectified linear unit)



(0,∞)

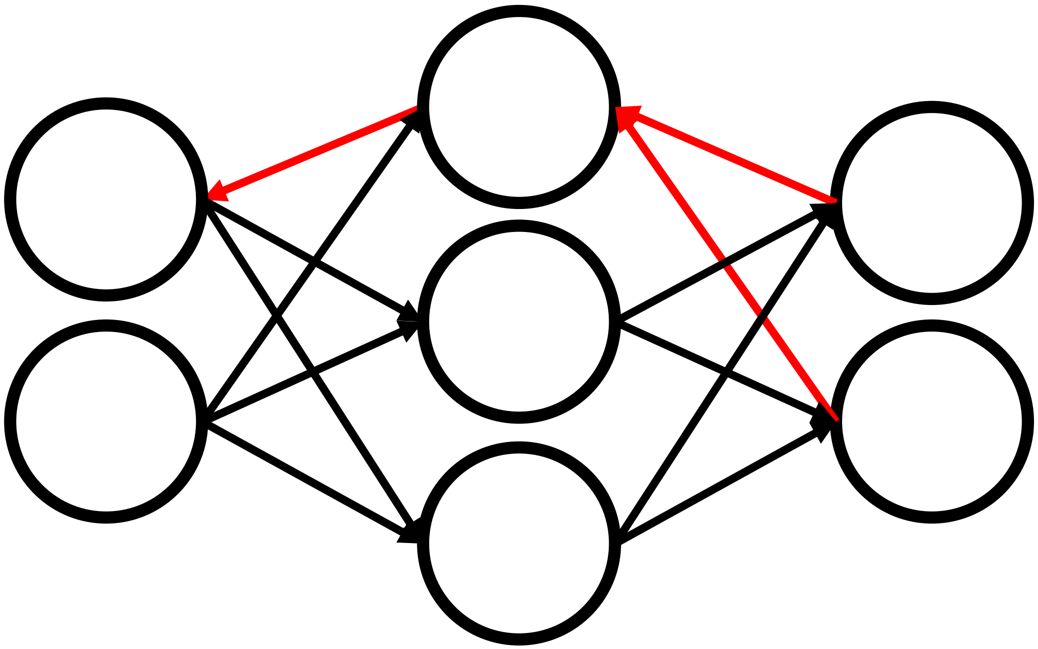
softmax different every time (0,1) sum is always 1

backpropagation: feed forward

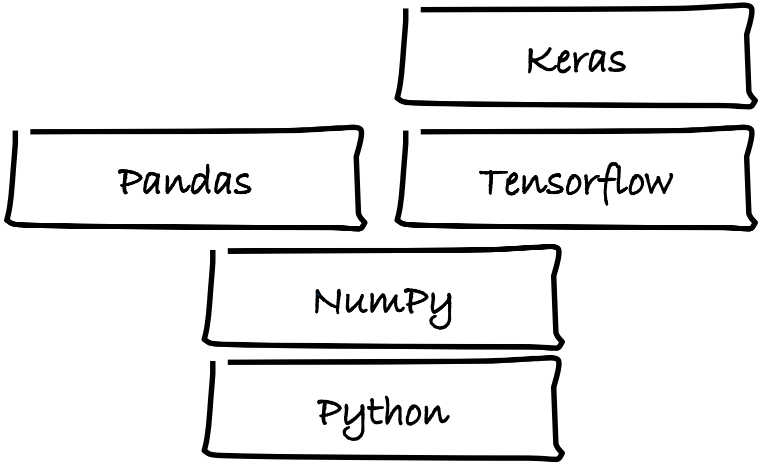


backpropagation

back prop algorithm identifies which weight contributed to which errors and feeds it back for adjustment

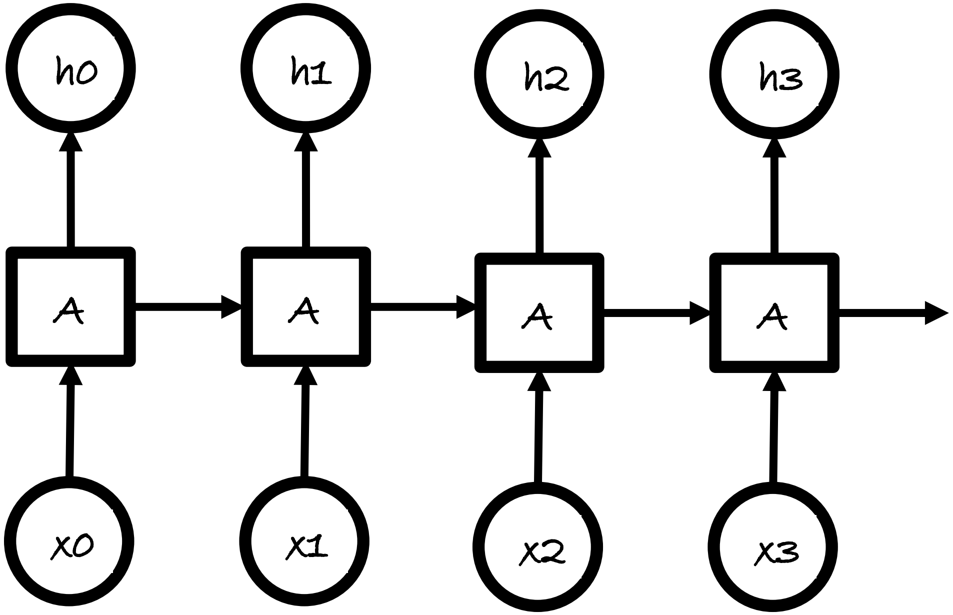


development frameworks examples

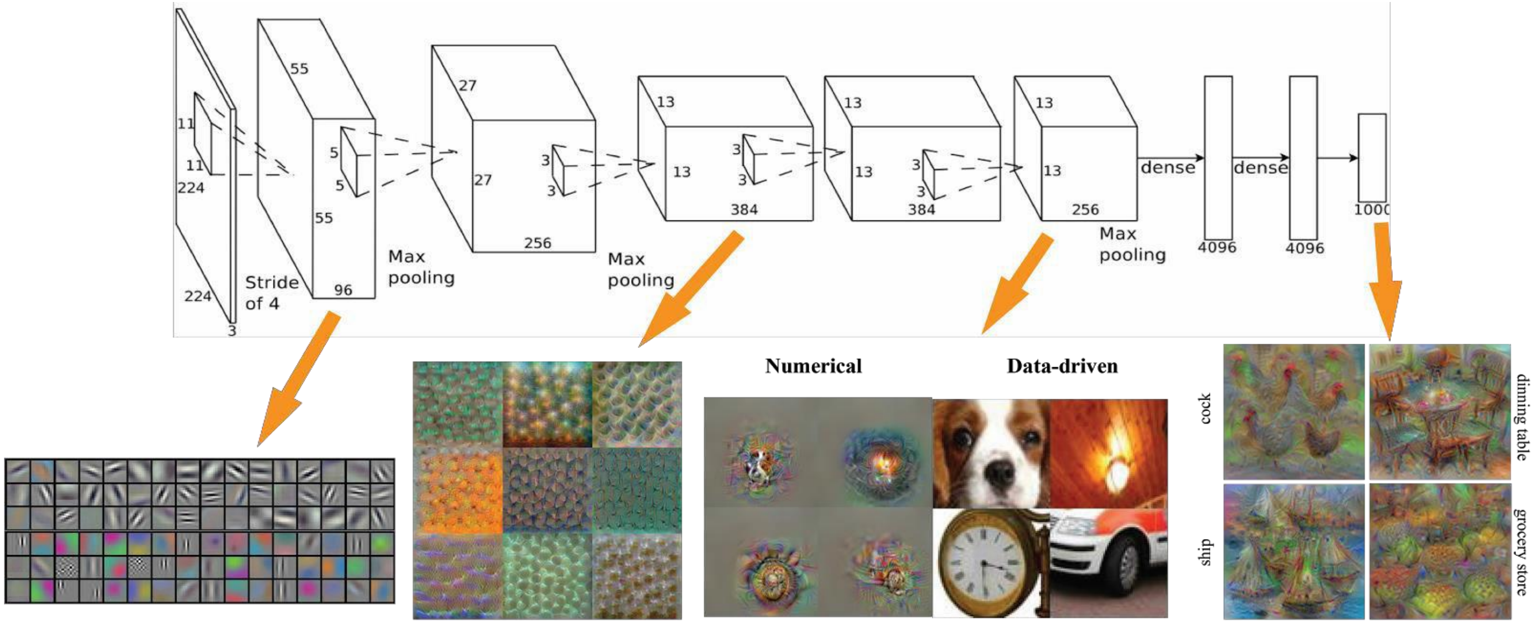


recurrent neural network

time-series processing (stock-market, sensor data etc.)



convolutional neural network



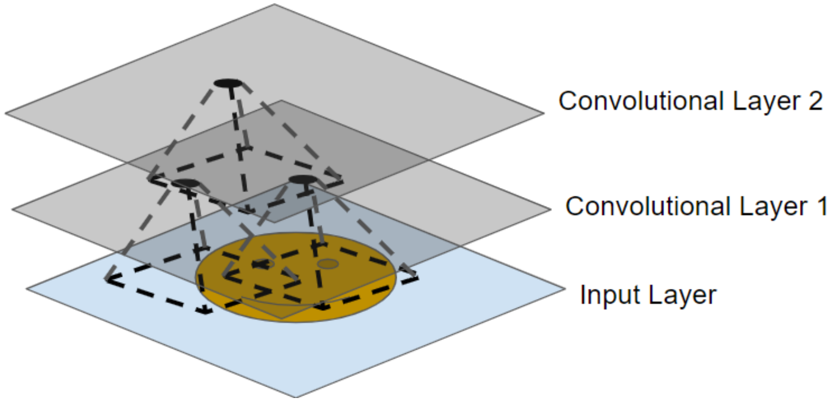
Conv 1: Edge+Blob

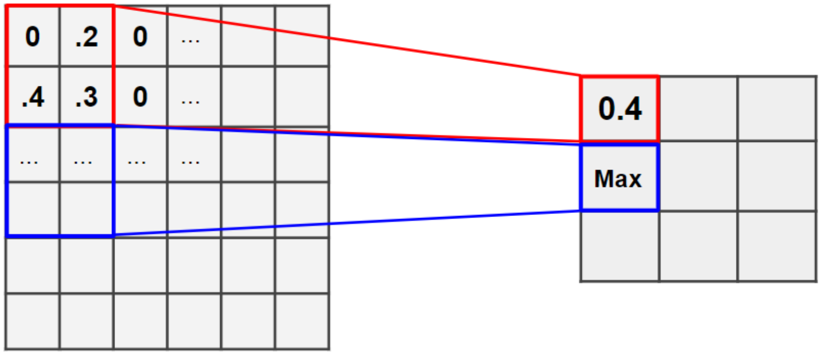
Conv 3: Texture

Conv 5: Object Parts

Fc8: Object Classes

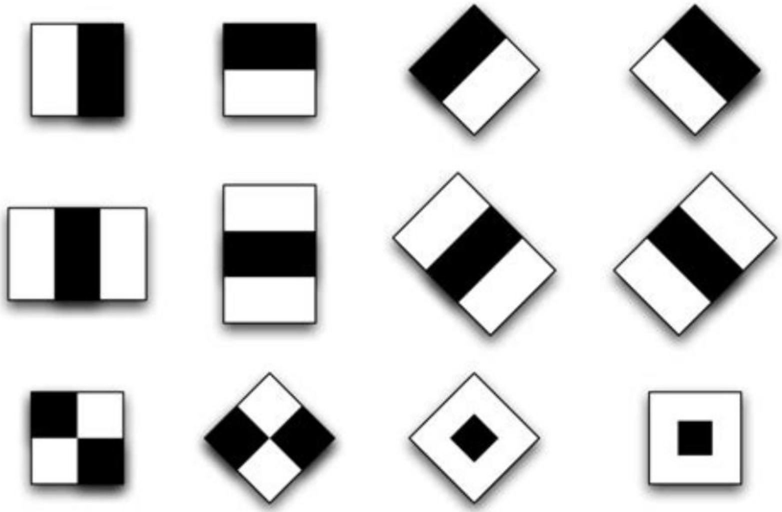
convolutional neural network





object/facial recognition

Haar-like features (after Alfred Haar)



edge features, line features and four-rectangle features

CNN: object detection

practical application in agricultur:

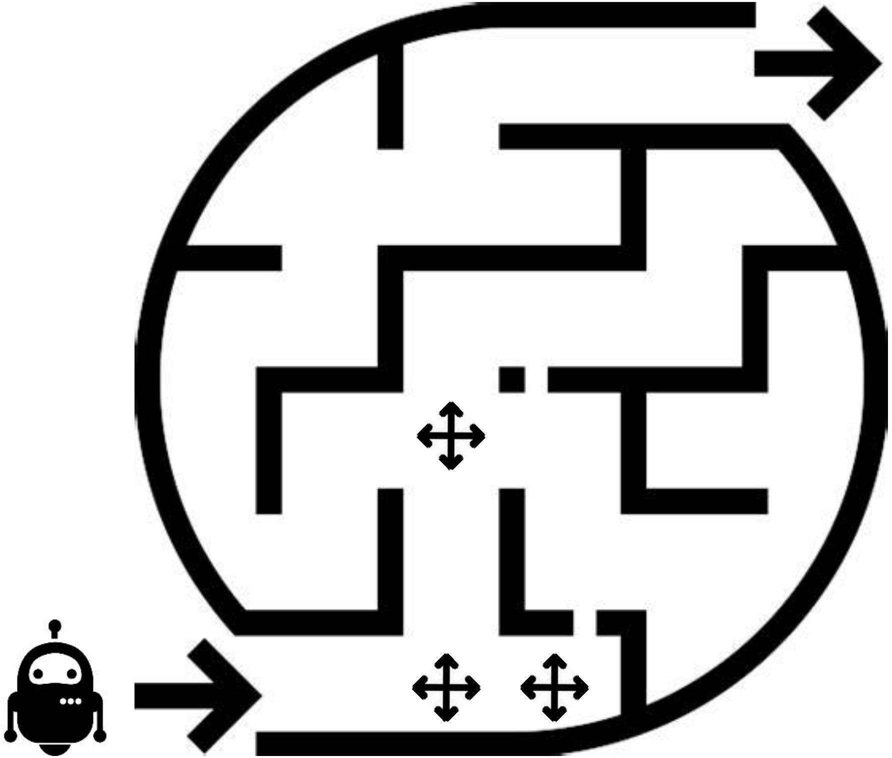
use drones regularly to scan crop fields for pests, lack of nutrients etc

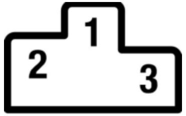
apply required amounts weed killer in affected areas, fertilizer and water where needed

increase efficiency of above mentioned agents and be more environmentally conscious

reinforcement learning

getting better by trial and error





components:

- environment
- agent
- actions
- reward

practical application:

- robotics
- manufacturing
- logistics
- optimization

reinforcement learning

roomba

visually scan area

plan route

vacuum clean area

return to dock before battery runs out

repeat according to schedule