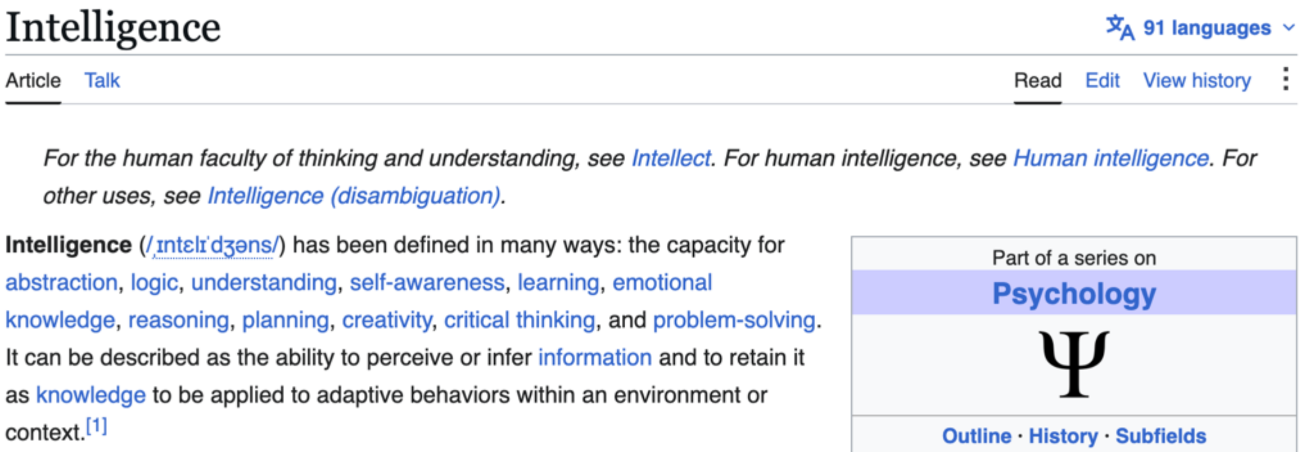
Machine Learning

Guillem Guigó i Corominas

guillem.guigo@uvic.cat


Intelligence?


Intelligence

91 languages

Article Talk Read Edit View history

For the human faculty of thinking and understanding, see Intellect. For human intelligence, see Human intelligence. For other uses, see Intelligence (disambiguation).

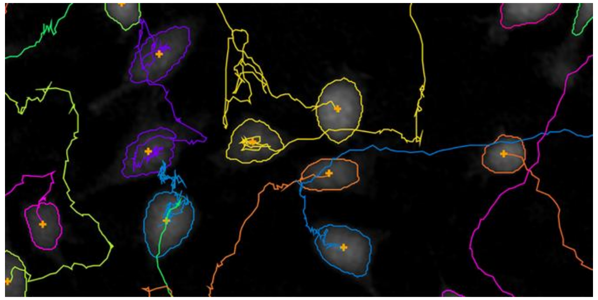
Intelligence (/ɪnˈtɛlɪdʒəns/) has been defined in many ways: the capacity for abstraction, logic, understanding, self-awareness, learning, emotional knowledge, reasoning, planning, creativity, critical thinking, and problem-solving. It can be described as the ability to perceive or infer information and to retain it as knowledge to be applied to adaptive behaviors within an environment or context.[1]

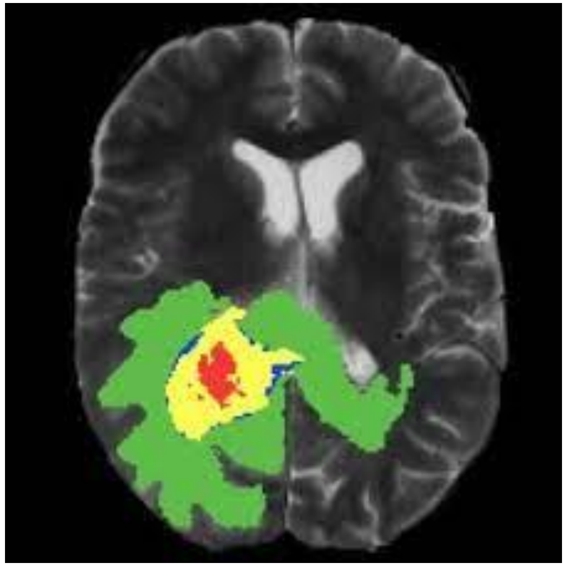
Part of a series on
Psychology
Ψ
Outline · History · Subfields

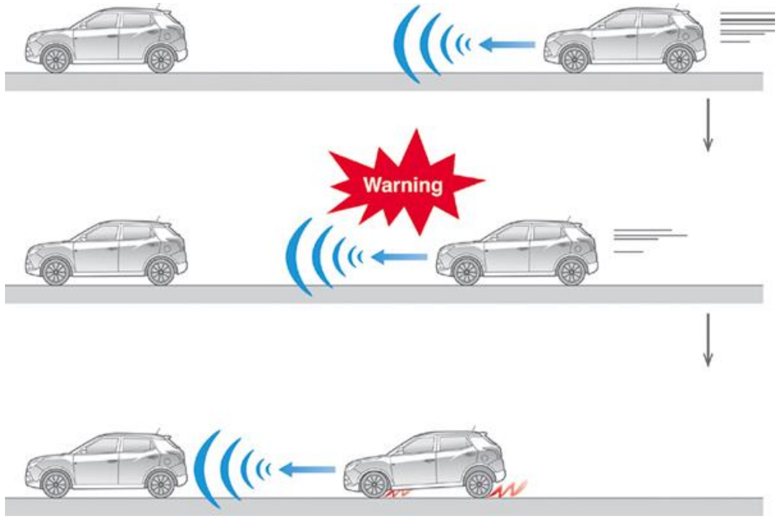
Intelligence is the cognitive capacity to learn, understand, apply knowledge, and adapt to new situations. It encompasses abilities like logic, abstract thinking, problem-solving, and self-awareness

Natural intelligence: displayed by animals and some humans

Artificial intelligence: displayed by machines





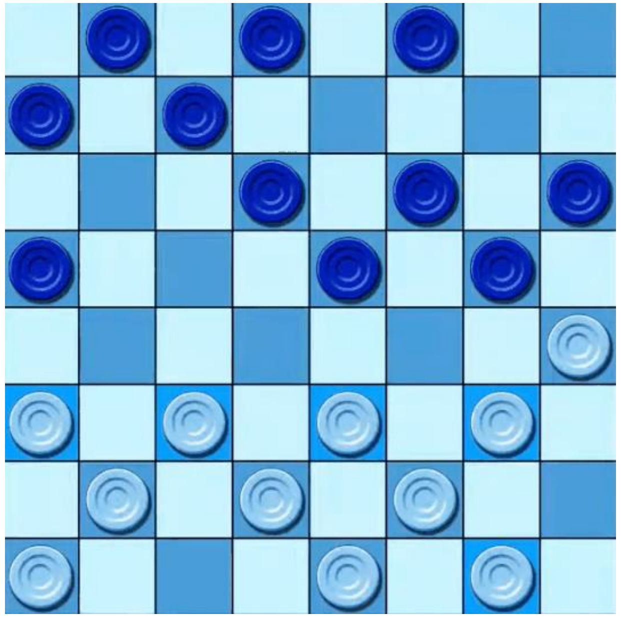


Learning

- The acquisition of knowledge or skills through study, experience, or being taught – *Oxford English Dictionary*
- Memory
- Repetition
- Improvement
- Understanding

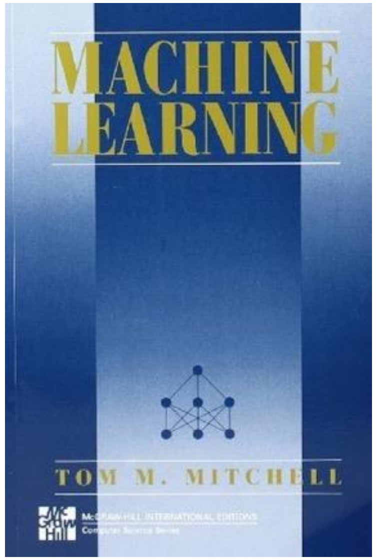
Machine learning

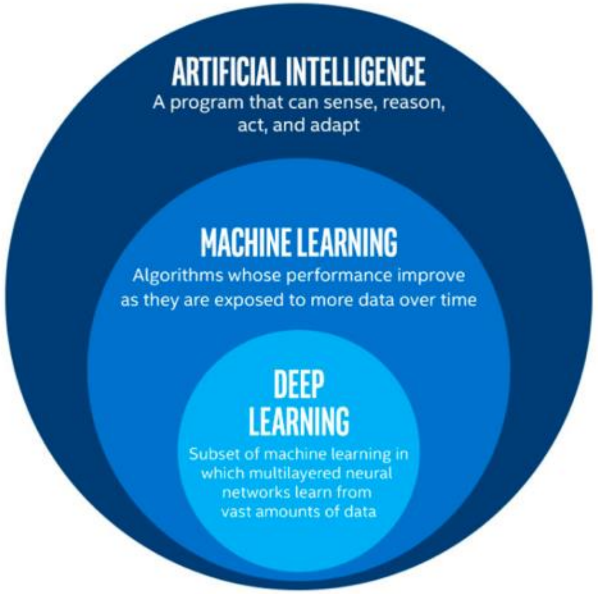
- Arthur Samuel in 1959: Computer that plays checkers that improves with experience
- Mini-max algorithm with alpha-beta pruning
- Rote learning: news inputs were compared with previously stored input-output pairs, amplifying the search depth of the tree
- Machine learning: “Field of study that gives computers the ability to learn without being explicitly programmed”

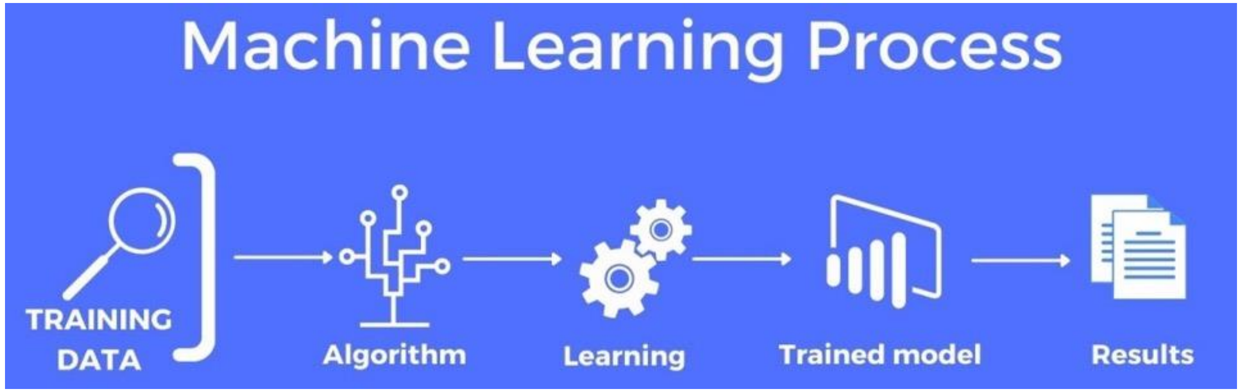


Machine learning

- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E ” – Tom Mitchell 1997







Computation

- Systematic approach to manipulate data
- Observations → from qualitative to quantitative descriptions
- Computer: machine that performs logical and arithmetic operations following instructions – programmable
- Algorithm: sequence of well-defined instructions

Data

- Data is represented by numbers → *D* dimensional vectors

Features

Observations

Height	Weight
178	77
186	82
168	66

$D = 2$

$N = 3$

$x^1 = [178, 77]$

$x^2 = [186, 82]$

$x^3 = [168, 66]$

$D = 3$

$N = 4$

m^2	*n* rooms	year
80	2	2014
75	3	1987
92	3	1999
130	4	2005

$$\begin{bmatrix} 80 & 2 & 2014 \\ 75 & 3 & 1987 \\ 92 & 3 & 1999 \\ 130 & 4 & 2005 \end{bmatrix}$$

4x3 matrix

$D = 5$

$N = 5$

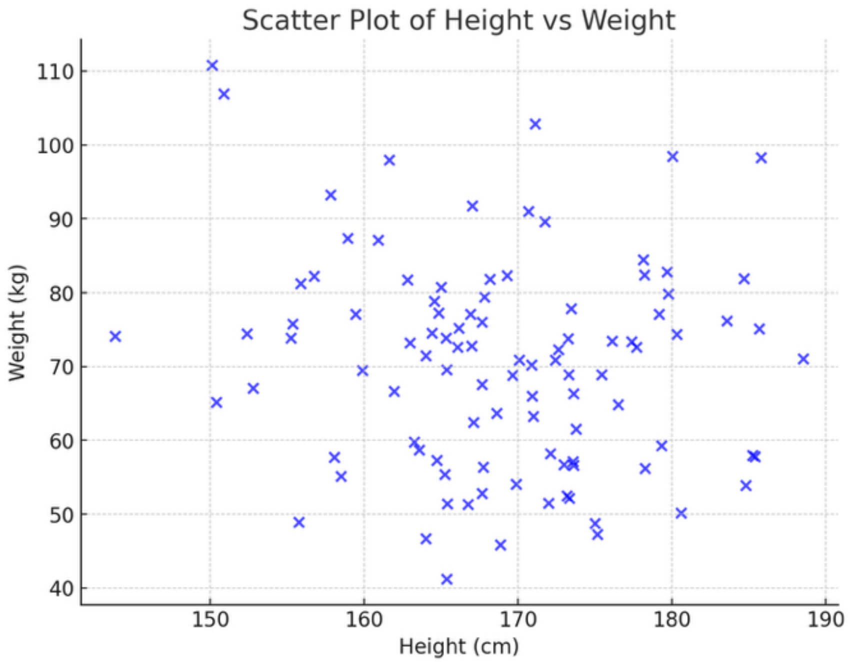
x_1	x_2	x_3	x_4	x_5
0.01	0.41	0.78	0.93	0.53
0.86	0.09	0.61	0.07	0.39
0.22	0.02	0.54	0.47	0.80
0.28	0.30	0.56	0.73	0.64
0.42	0.66	0.76	0.04	0.24

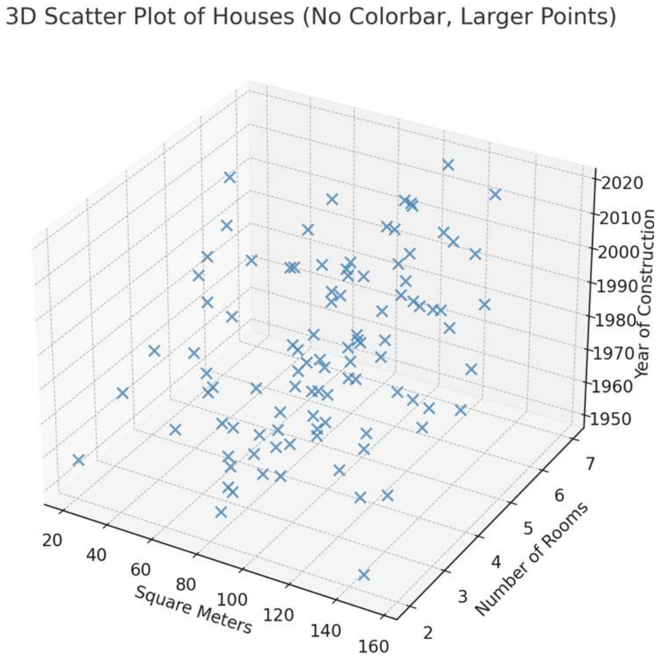
x_i^j → ith feature of the jth observation

Dataset → Set of all observations

Data representation

- Each observation is represented by a point in the dimension of the feature space

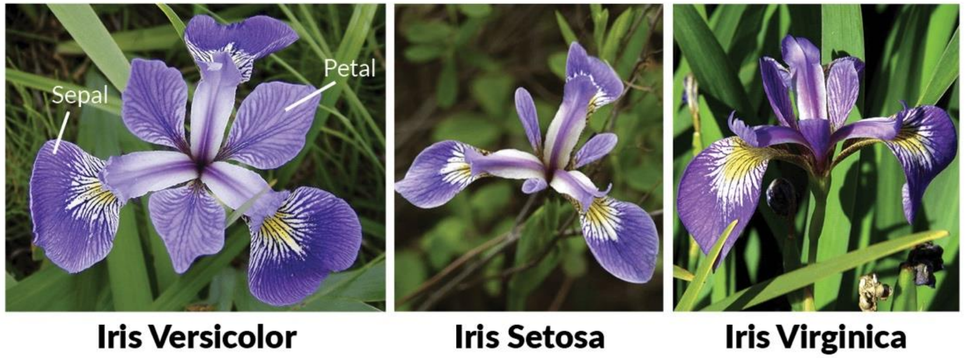


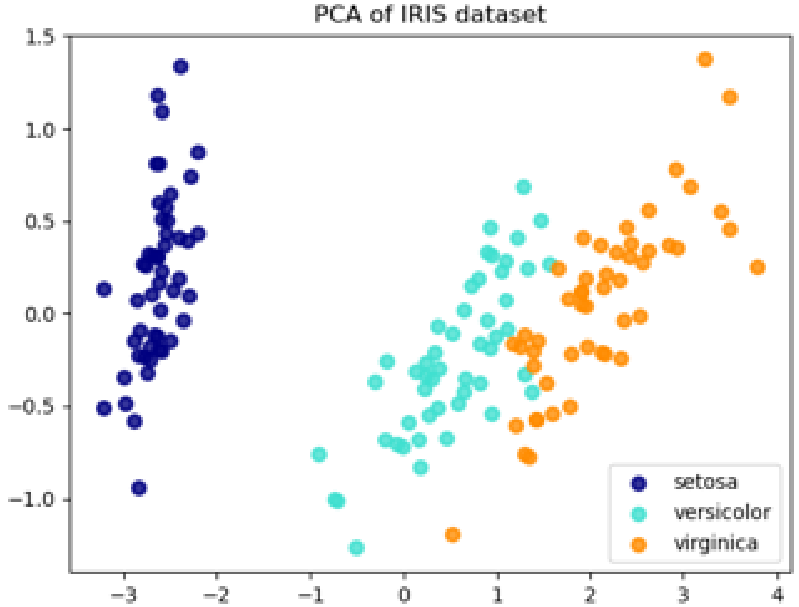


Euclidean distance→ $d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}.$

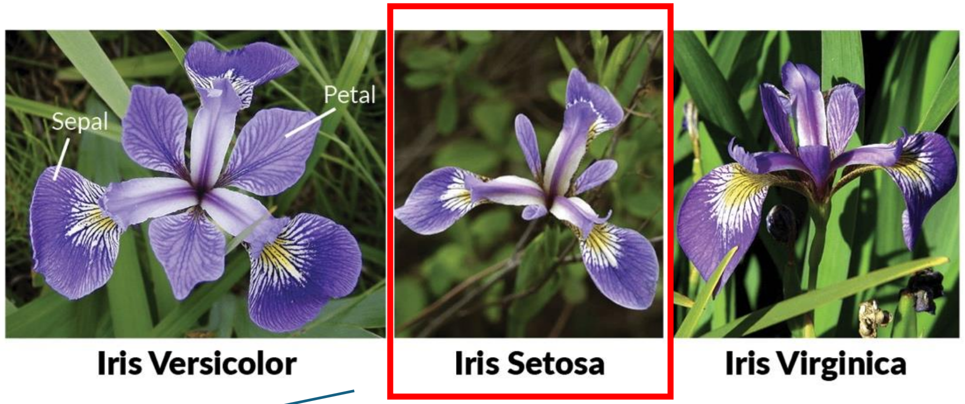
Example: Iris dataset

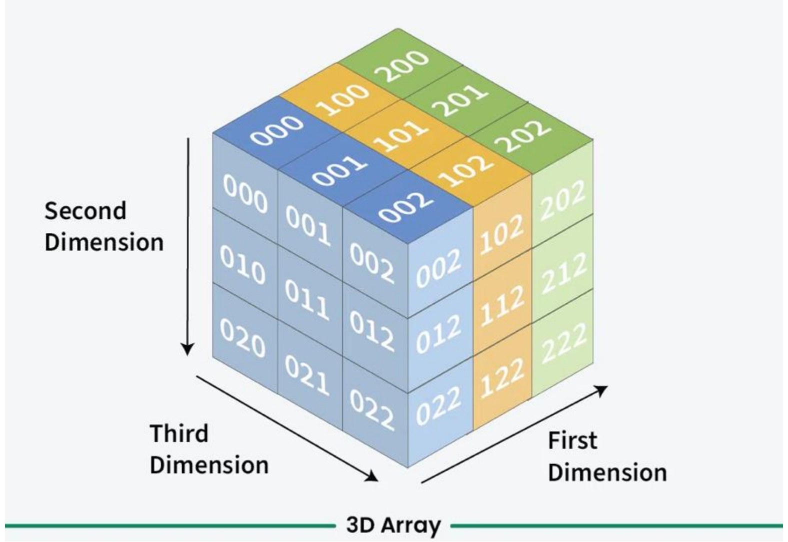
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa
15	5.8	4	1.2	0.2	Iris-setosa
16	5.7	4.4	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa
19	5.7	3.8	1.7	0.3	Iris-setosa



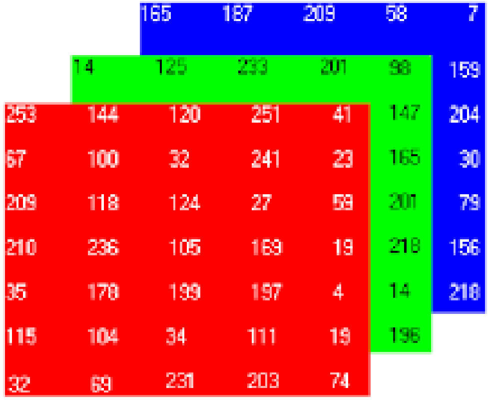


Example: Iris dataset





Tensor



RGB

$$X_1 = [0.99, 0.64, 0.83, 0.37, 0.56, 0.92, 0.90, 0.58, 0.27, 0.09, \ldots]$$

Data representation

Pointy ears	Eye separation	Snout length	Mouth width	Label
yes	3cm	4cm	8cm	Cat
no	4cm	12cm	20cm	Dog

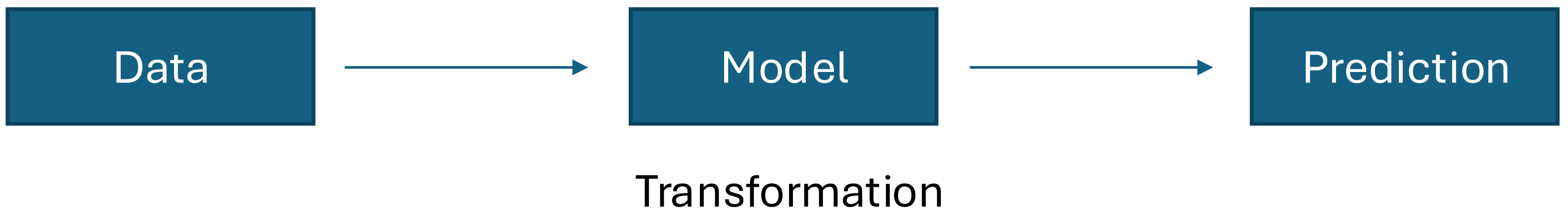
1	3	4	8
0	4	12	20

0
1

Labels

Predictions from data

- A model applies a transformation to the input data to generate a prediction



m^2	n rooms	year
80	2	2014
75	3	1987
92	3	1999
130	4	2005

Model of house price

$$price = 100 \cdot m^2 + 5000 \cdot n\ rooms + 5 \cdot year + 10000$$

Predictions from data

$f(m^2, n\ rooms, year) \rightarrow f: \mathbb{R}^3 \rightarrow \mathbb{R}^1$

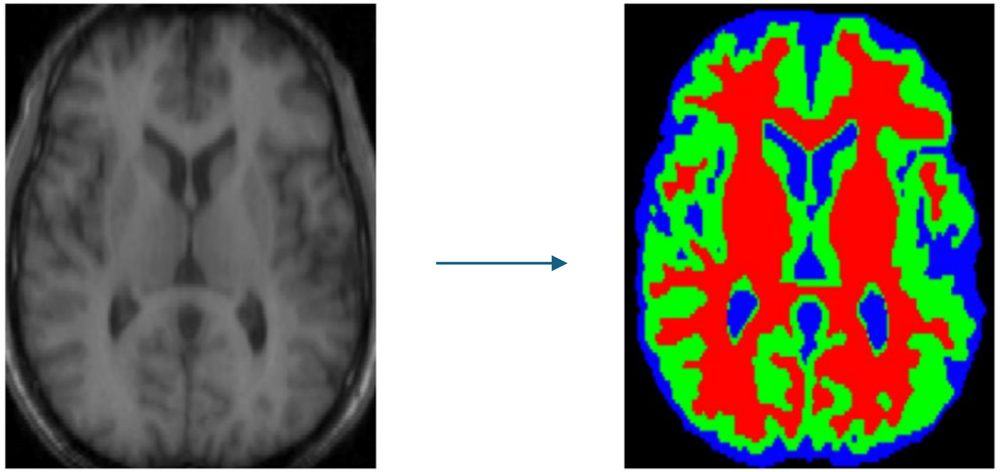
Regression: prediction of a continuous value

(cat and dog feature representation)
$f(x_1, x_2, x_3 x_4) \rightarrow f: \mathbb{R}^4 \rightarrow \mathbb{R}^1$

Classification: prediction of a discrete category

(cat and dog image representation)
$f(P) \rightarrow f: \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^1$

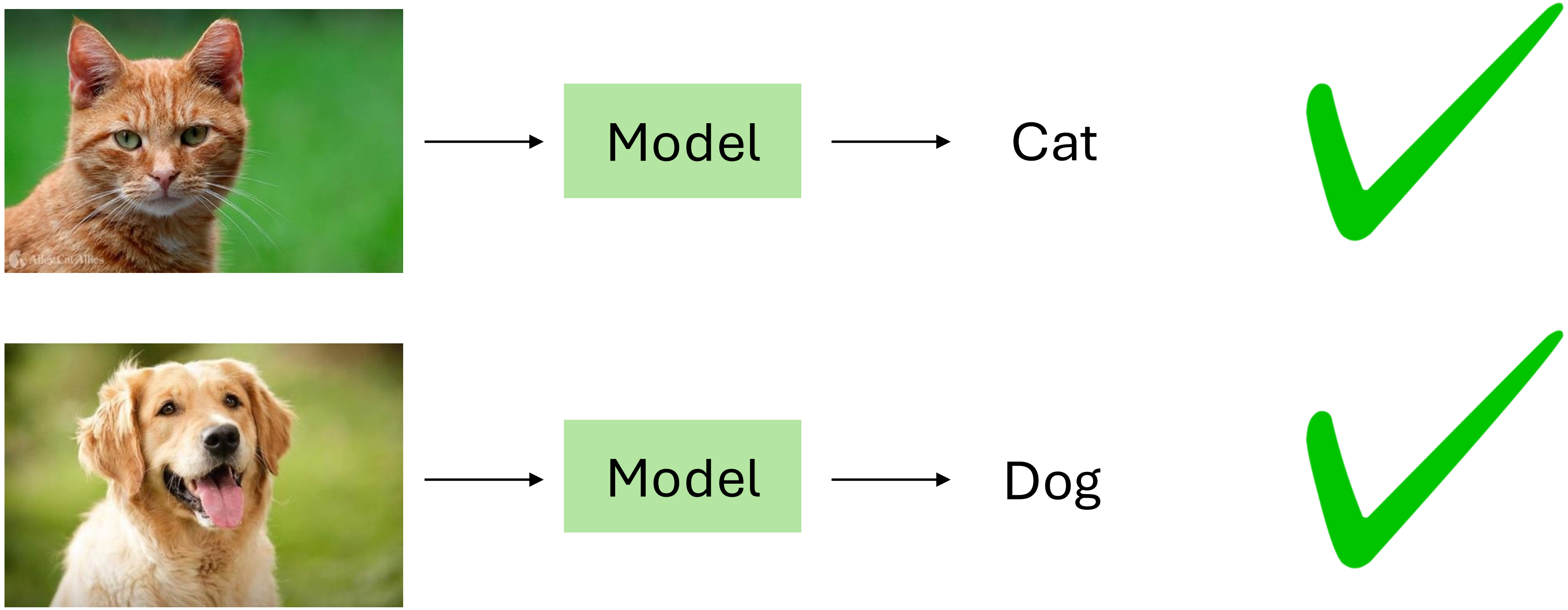
Brain tissue segmentation
$f(P) \rightarrow f: \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H \times W \times C}$



(a) Axial slice (b) Tissue segmentation

How do we come up with the correct model?

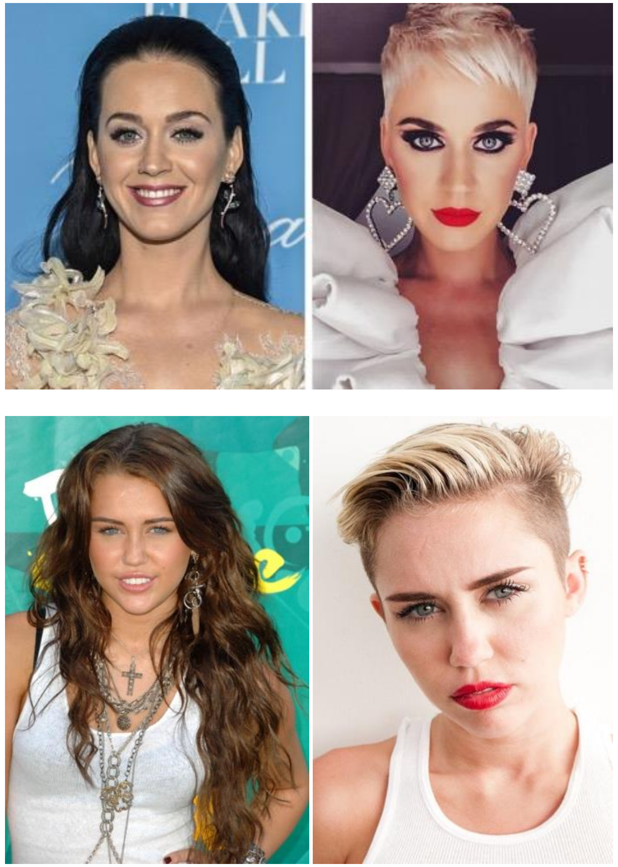
- How can we find the transformation that for a given input returns the correct output?



- We assume there exists a true underlying function $\mathrm{F}: \mathrm{X} \rightarrow \mathrm{Y}$
- We use available data + algorithms to approximate it $F' \approx F$

$$\begin{bmatrix} RBC \\ HB \\ HCT \\ WBC \\ MCV \\ \ldots \end{bmatrix}$$

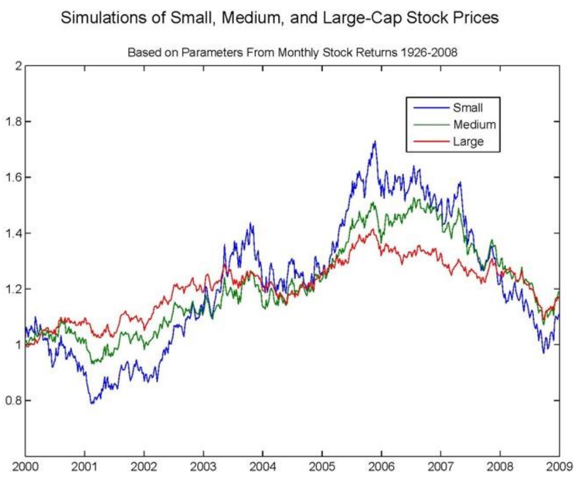
Leukemia or healthy?

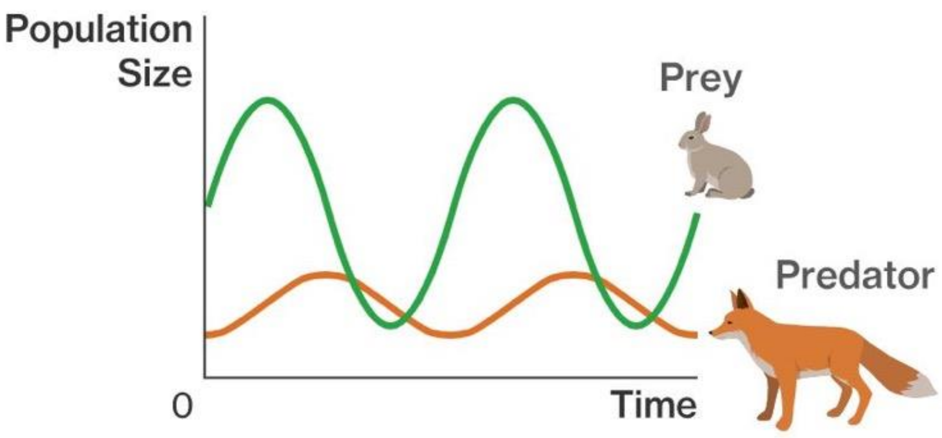


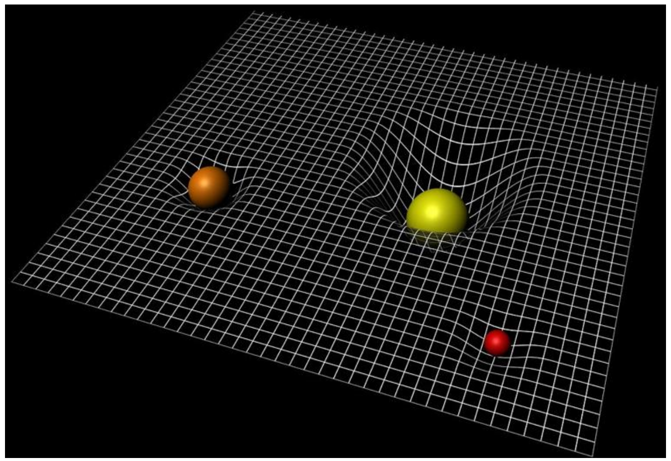
Katy Perry or Miley Cirus?

Modeling

- A mathematical model is a formal description of how a system behaves
- We usually make assumptions about this behavior
- Mathematical models are described by functions

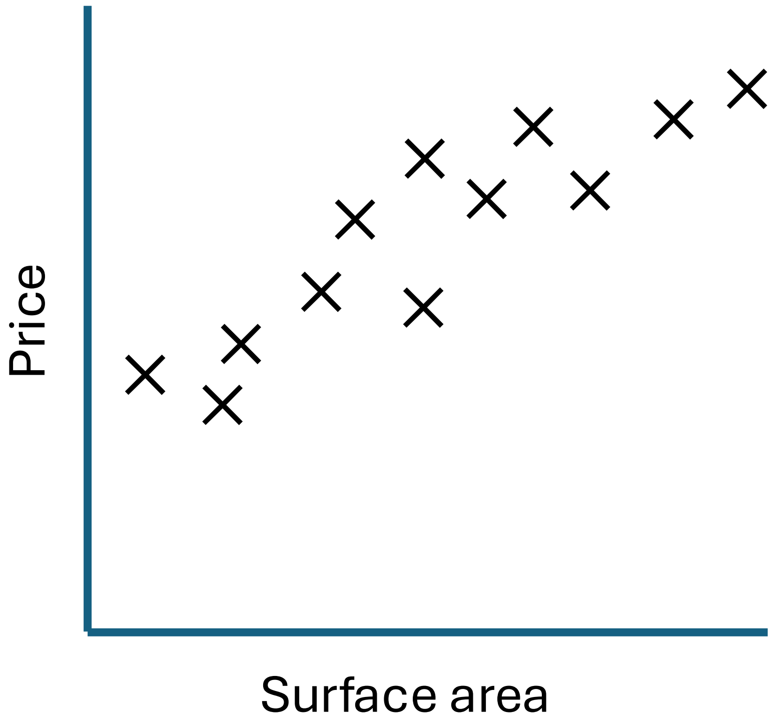






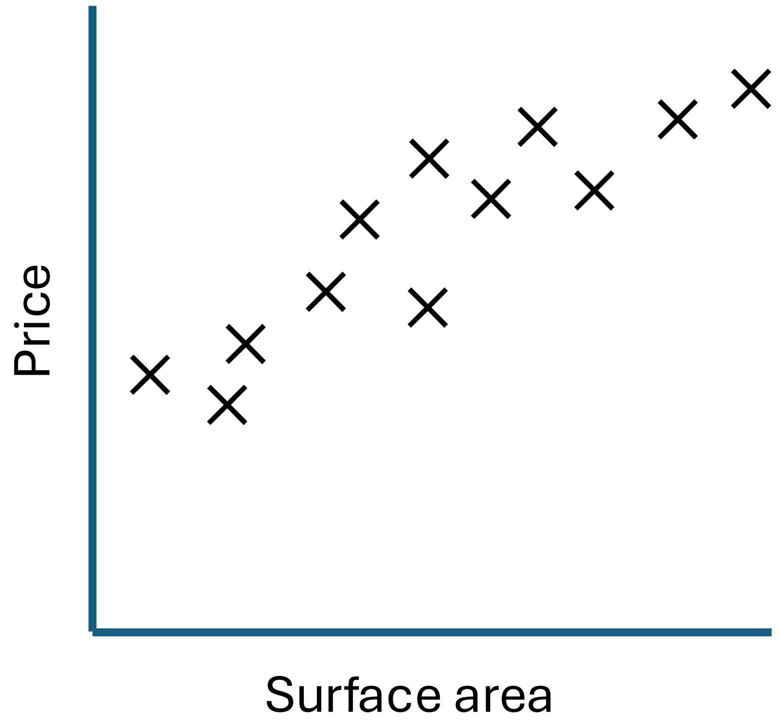
House price model

- A model to predict the price of a house based on its surface area in square meters
- Linear model: $y = ax + b$
- How do we find a and b ?
- We will use the observations
- Iterative process

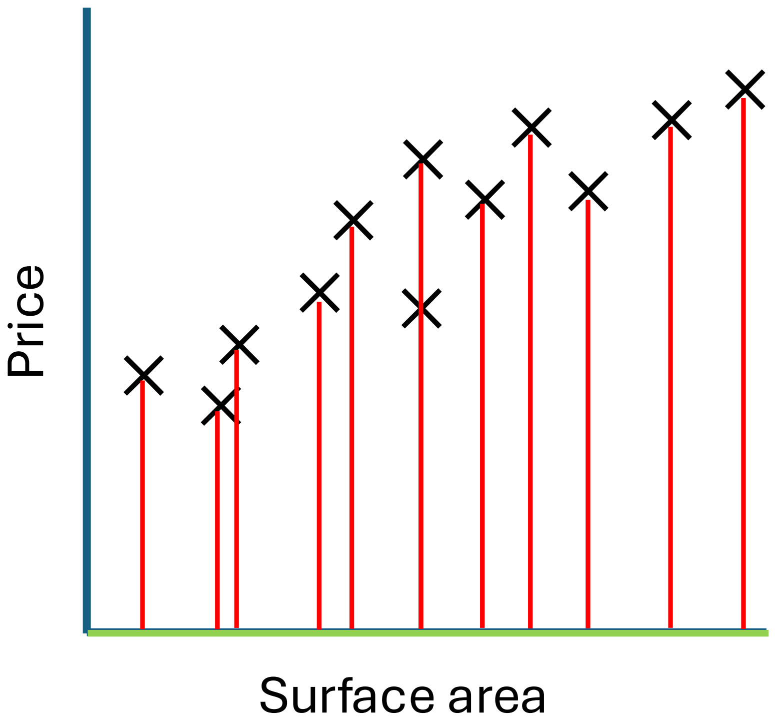


House price model

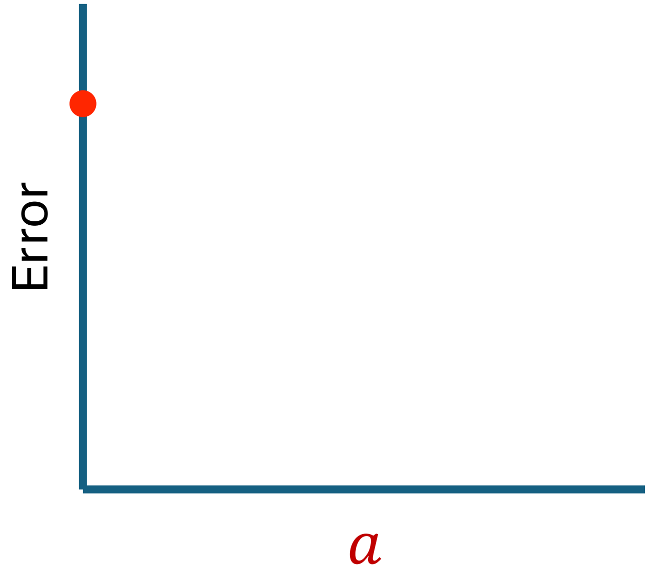
1. Initial guess: $a = 0, b = 0$
2. Compute the error of the model
3. Slightly increase or decrease the value of a such that the error of the new model is lower
4. Do the same for b
5. Repeat step 2 until the error no longer decreases

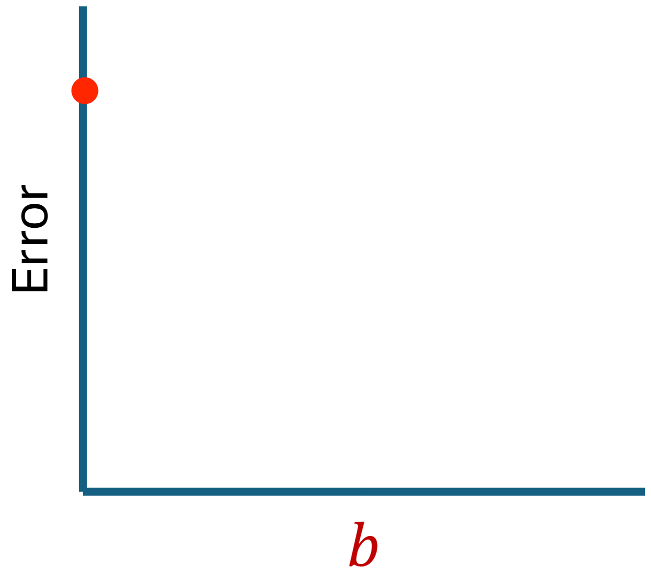


House price model

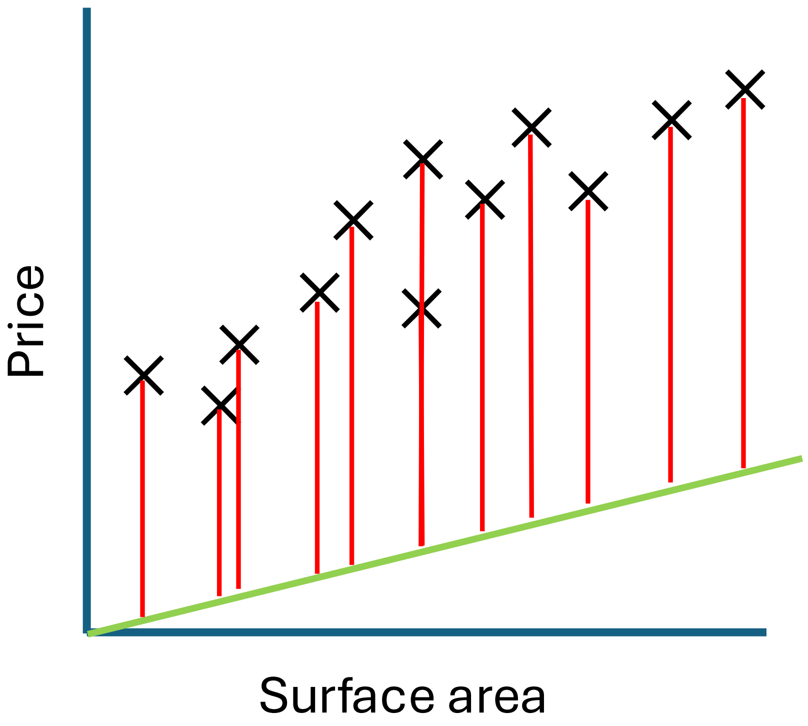


$$\hat{y} = ax + b \qquad Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

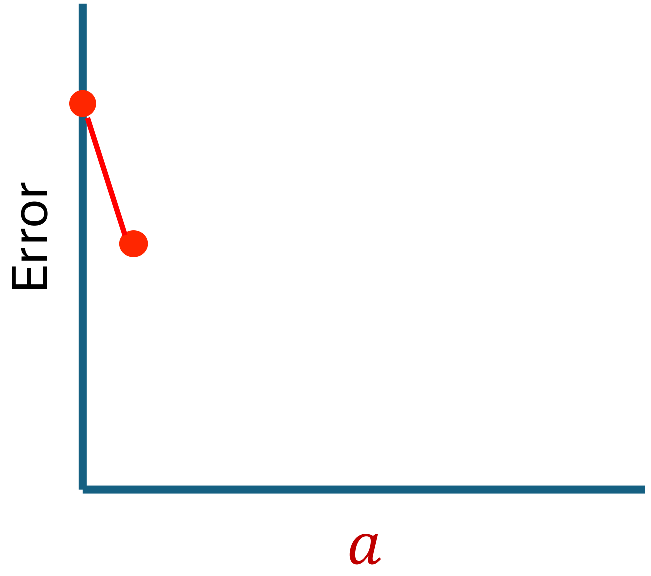


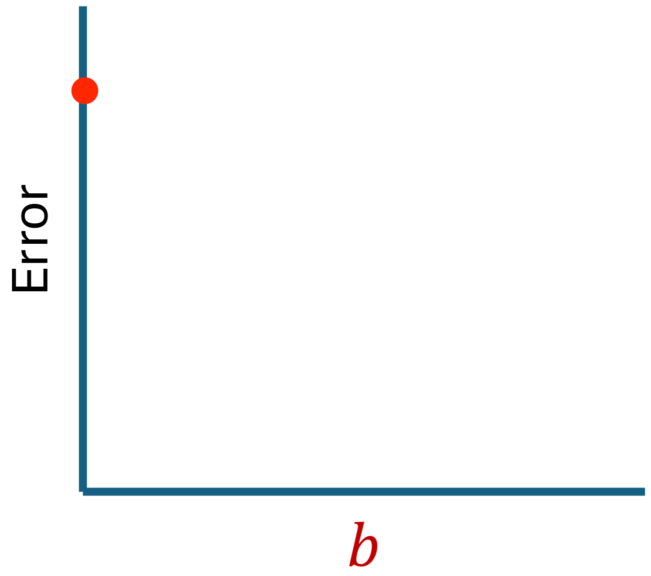


House price model

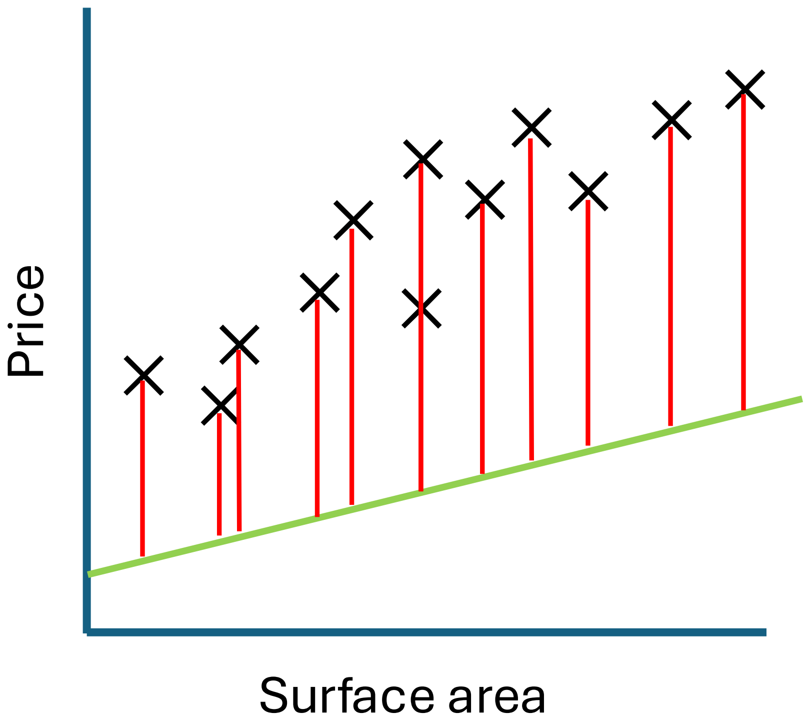


$$\hat{y} = ax + b \qquad Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



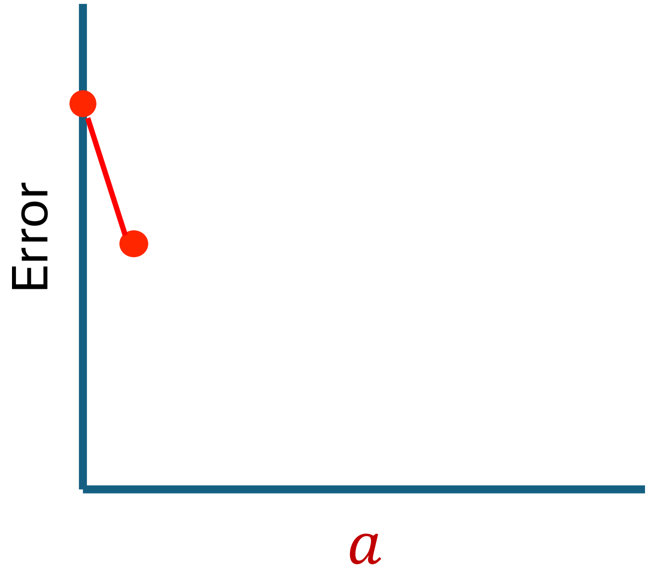


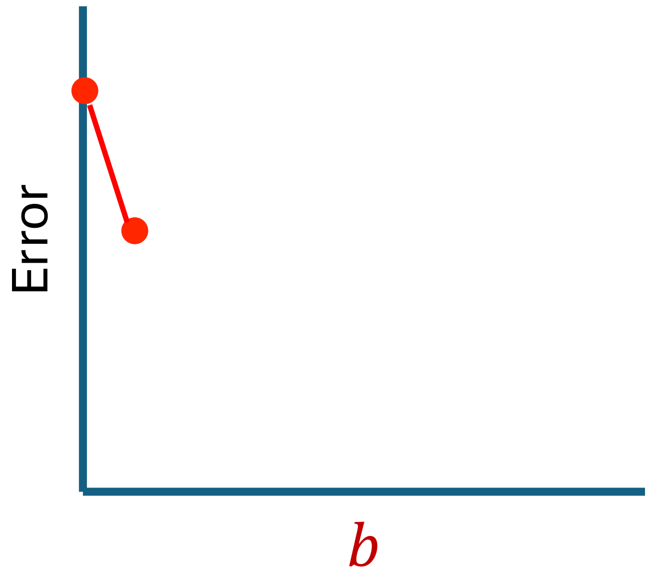
House price model



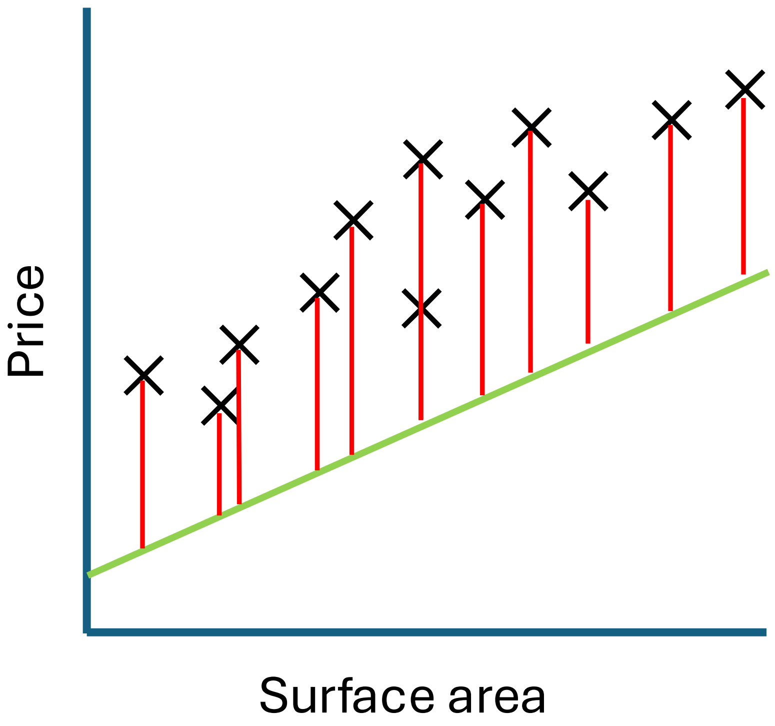
$$\hat{y} = ax + b$$

$$Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

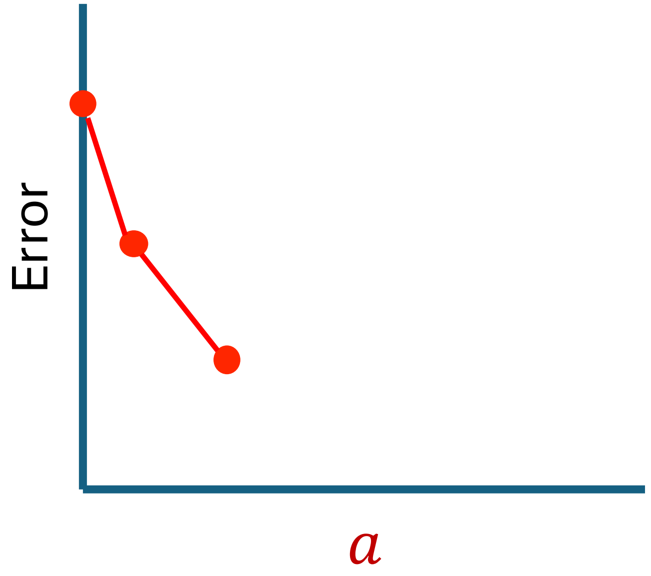


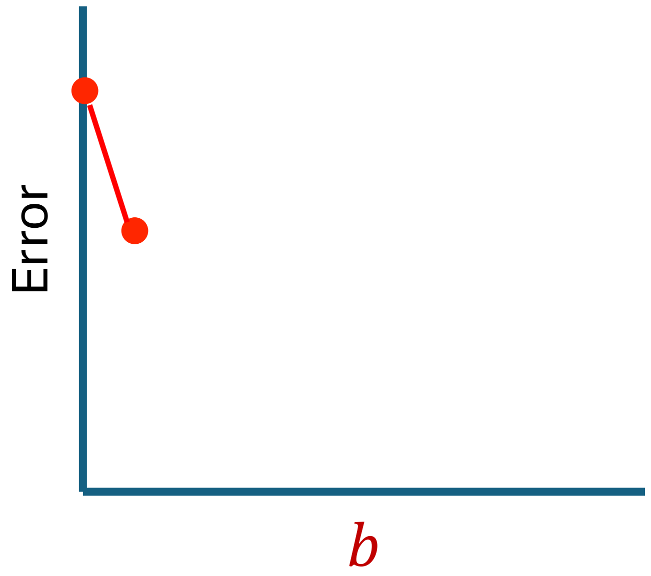


House price model

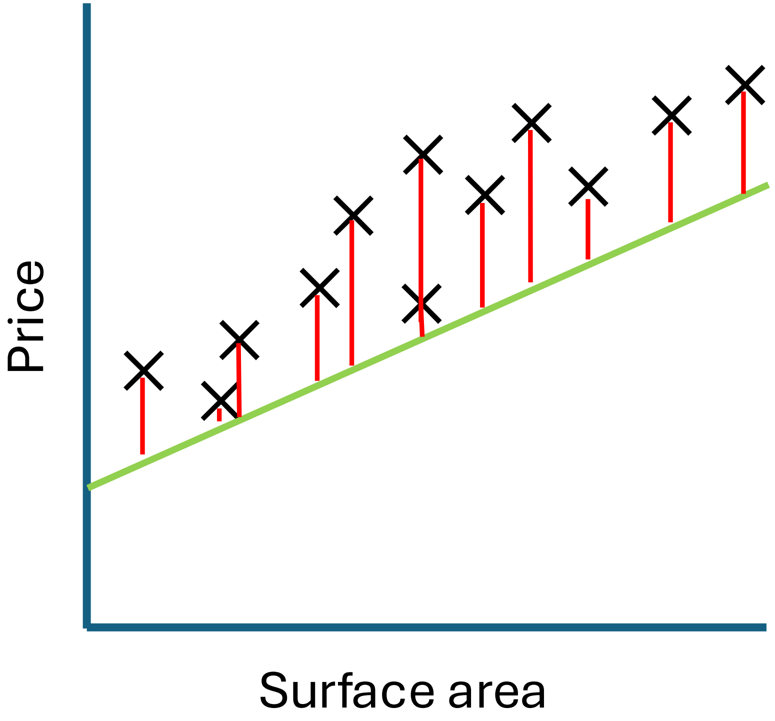


$$\hat{y} = ax + b \qquad Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

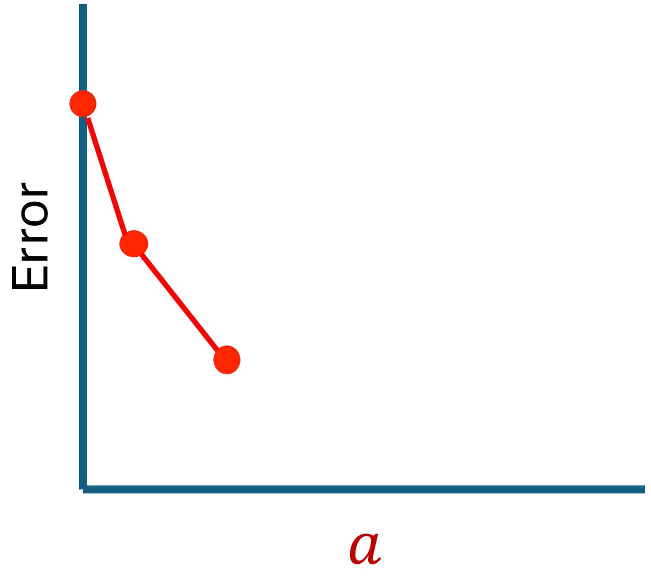


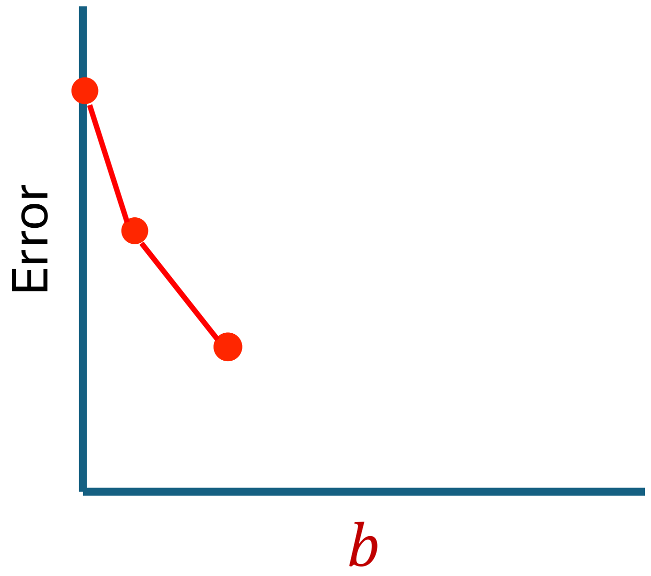


House price model

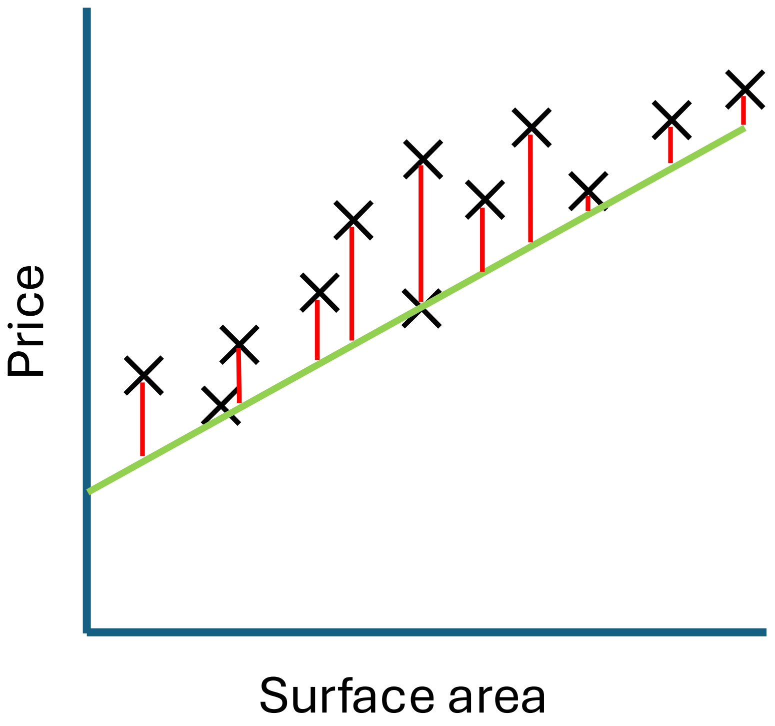


$$\hat{y} = ax + b \qquad Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



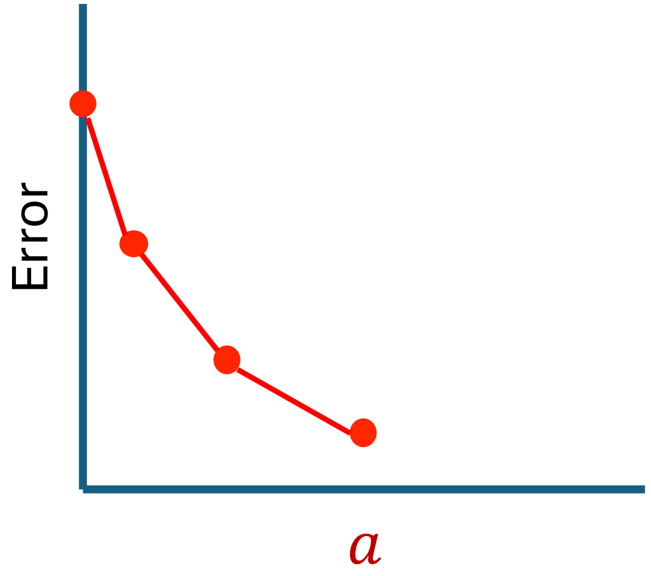


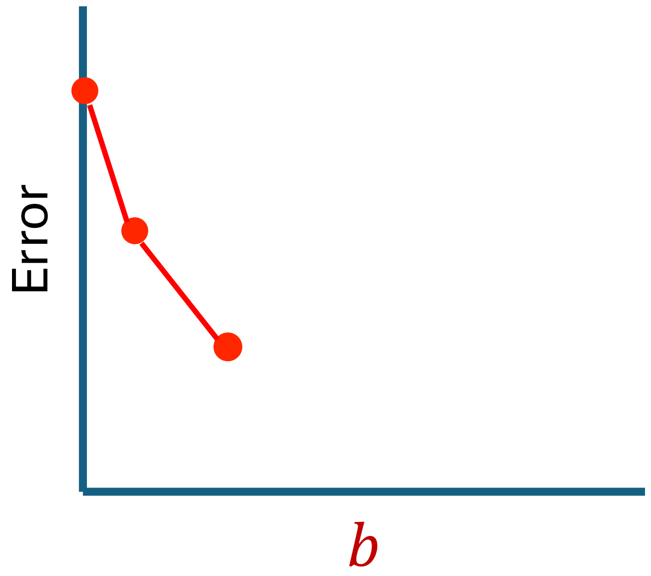
House price model



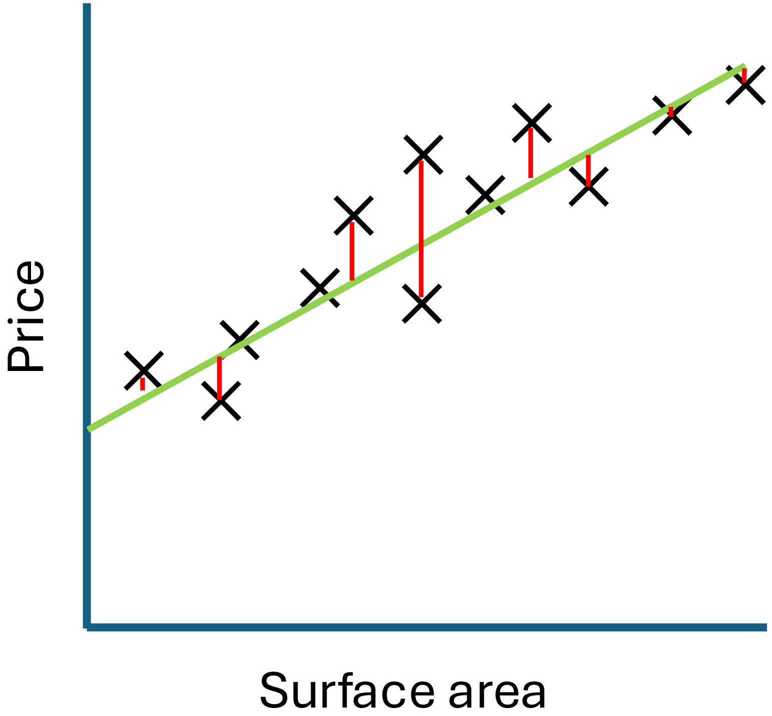
$$\hat{y} = ax + b$$

$$Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



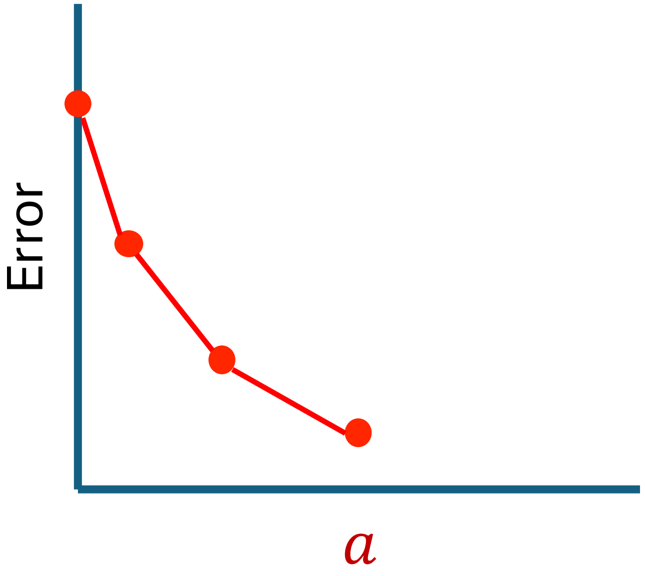


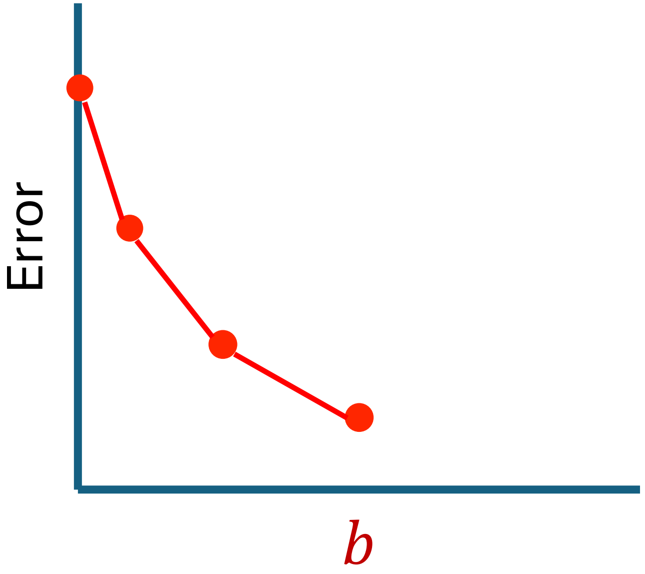
House price model



$\hat{y} = ax + b$

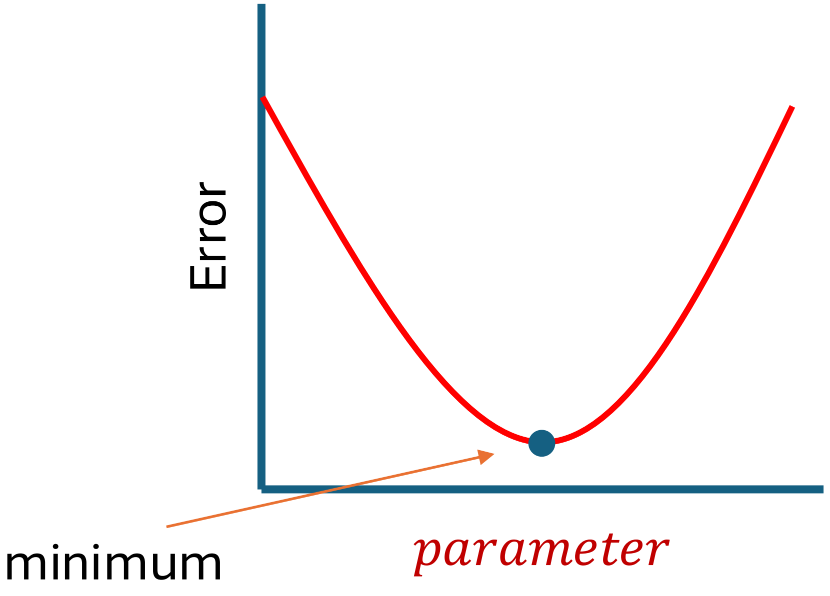
$$Error(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$





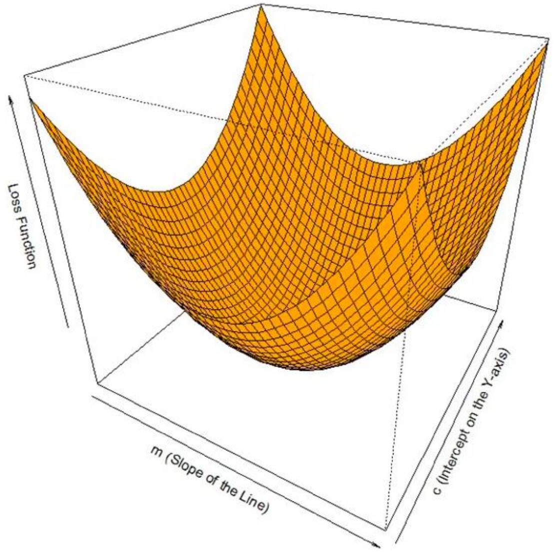
House price model

- Optimization problem: minimize the loss function (MSE)
- Simple linear regression with ordinary least squares (OLS)
- Gradient Descent algorithm



$$\widehat{\alpha} = \bar{y} - (\widehat{\beta}\,\bar{x}),$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$



Parametrization

- Parameters are constants that define a particular transformation
- In a multivariate linear model (for example the price of a house):

$$f(x_1, x_2) = ax_1 + bx_2 + c$$

x_1, x_2 are variables $\quad\quad$ a, b, c are parameters

- f is parametrized by x_1, x_2
- In the following cubic model:

$$g(x) = 2x^3 - 1x^2 + 3x - 4$$

x is a variable $\quad\quad$ $2, -1, 3, -4$ are parameters

- In machine learning, training a model means finding the parameters that minimize an objective function

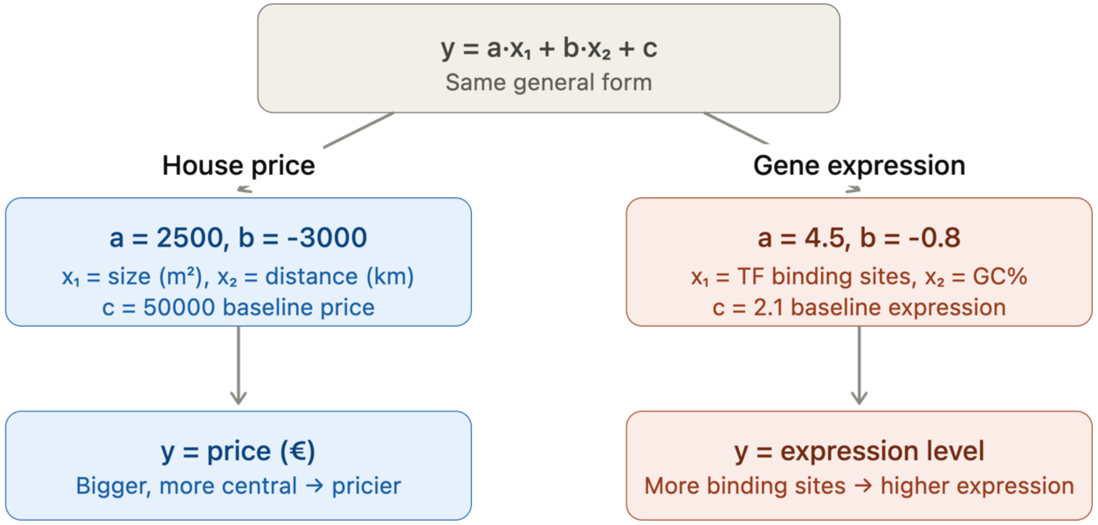
Linear model: $y = ax_1 + bx_2 + c$

- $y^{(1)} = 2500x_1^{(1)} - 3000x_2^{(1)} + 5000$

price = a·(size) + b·(distance to city center) + c
a=2500, b=-3000, c=5000

- $y^{(2)} = 4.5x_1^{(2)} - 0.8x_2^{(2)} + 2.1$

expression = a·(number of transcription factor binding sites) + b·(GC content %) + c
a=4.5, b=-0.8, c=2.1



Same equation, different real-world meaning behind a, b, c

Example: tomato price model

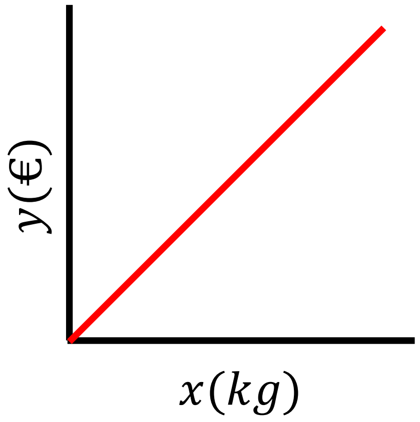
- How is the amount of tomatoes of a certain variety that I buy related to the price that I pay?
- Assumptions:

1. Each kg of tomatoes is worth equal
2. There is no flat rate to pay when buying tomatoes

We can find a by looking at the price tag

$$y = f(x) = ax$$

E.g. $a = \frac{2€}{kg}$



slope $= a$

Example: exponential growth

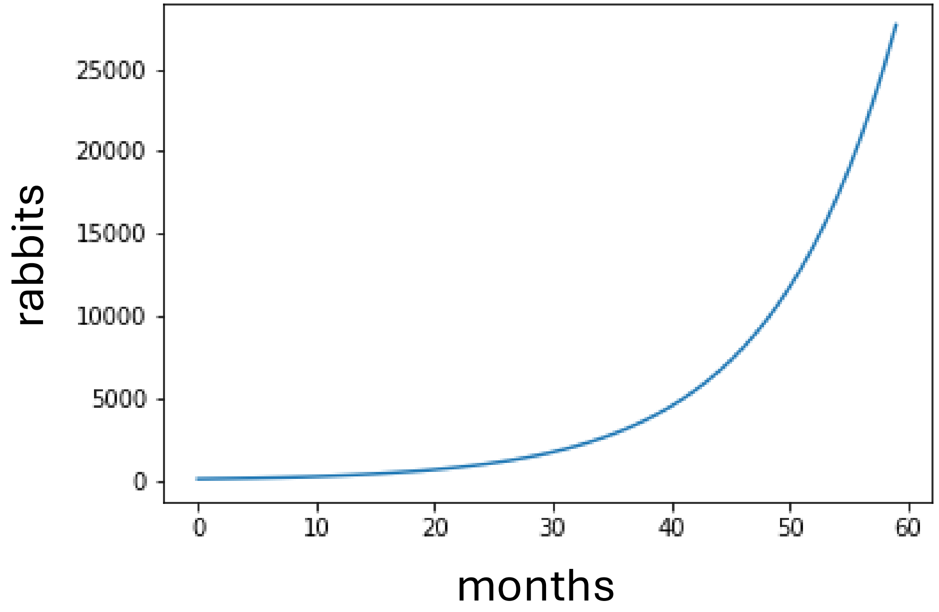
- How does a rabbit population grow with time?
- Suppose we start with a population of 100 rabbits and it increases by 10% each month. The rabbit population population after 5 years is given by

$$P(n) = 100 \cdot (1.1)^{60} = 27680$$

- The general model of exponential growth is

$$P(n) = a(1 + r)^{t},$$

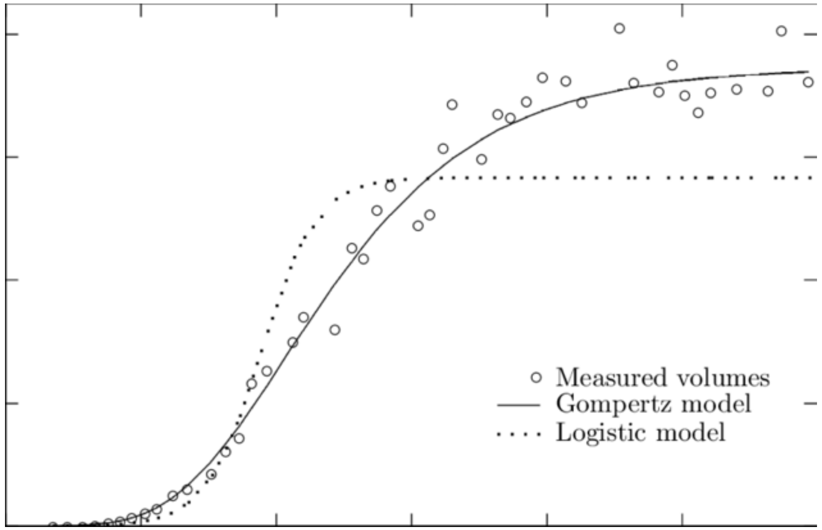
Where a is the initial population, r the growth rate and t the time



Example: logistic growth

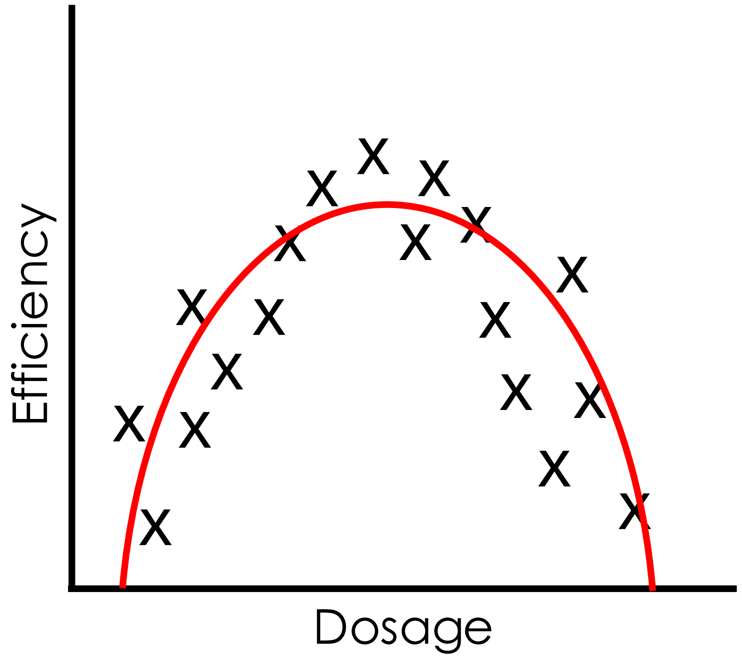
- The population cannot grow indefinitely
- L is the carrying capacity of the system
- k is the logistic growth rate
- x_0 is the sigmoid midpoint
- We find L, k and x_0 by minimizing some error function in our observations

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$



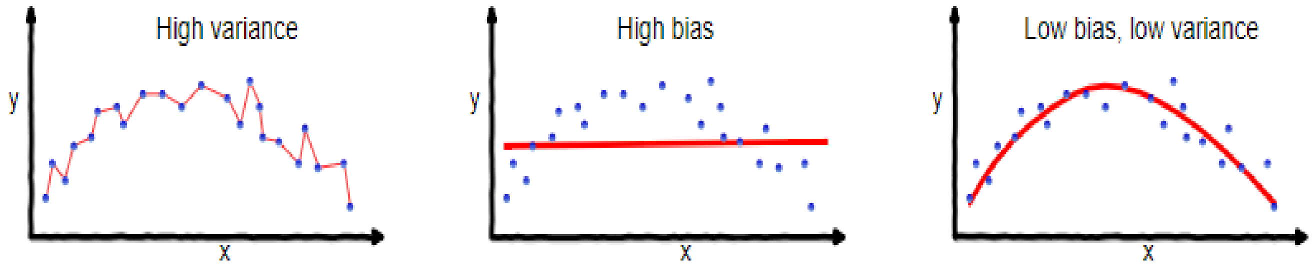
Example: drug efficiency model

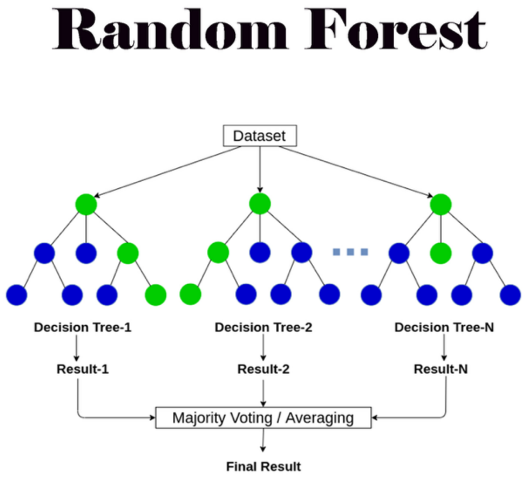
- The effectiveness of a drug may be unnoticeable for small doses, ideal in intermediate doses and harmful in high doses
- This can be modelled by a quadratic function: $E(x) = ax^2 + bx + c$
- Again, a, b and c are the parameters of the model, in this case the coefficients of a quadratic equation, and we find them using the available data



How did we build these models?

- We have some freedom to choose how we want the function that describes our model to be
- The more degrees of freedom (parameters), the less assumptions we do about the structure of our data
- We aim to generalize → reasonable predictions on unseen data



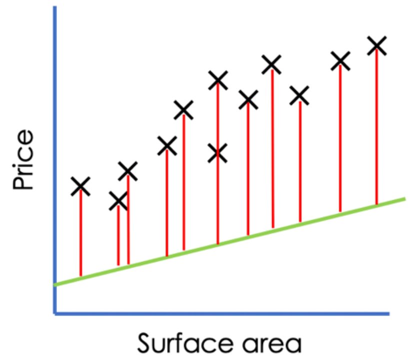


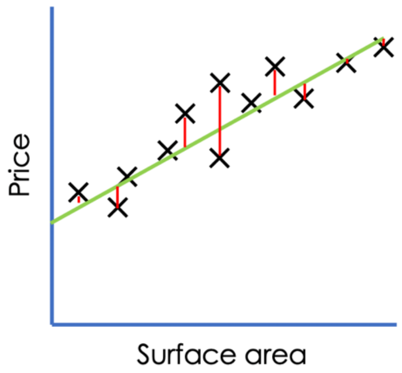
ML models examples: Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, SVMs...

Model training

- Metric that we want to optimize
- Algorithm → procedural method to find the parameters that minimize that metric
- Model → set of parameters & functions that define the transformation from the input data to prediction

$$\hat{y} = ax + b$$





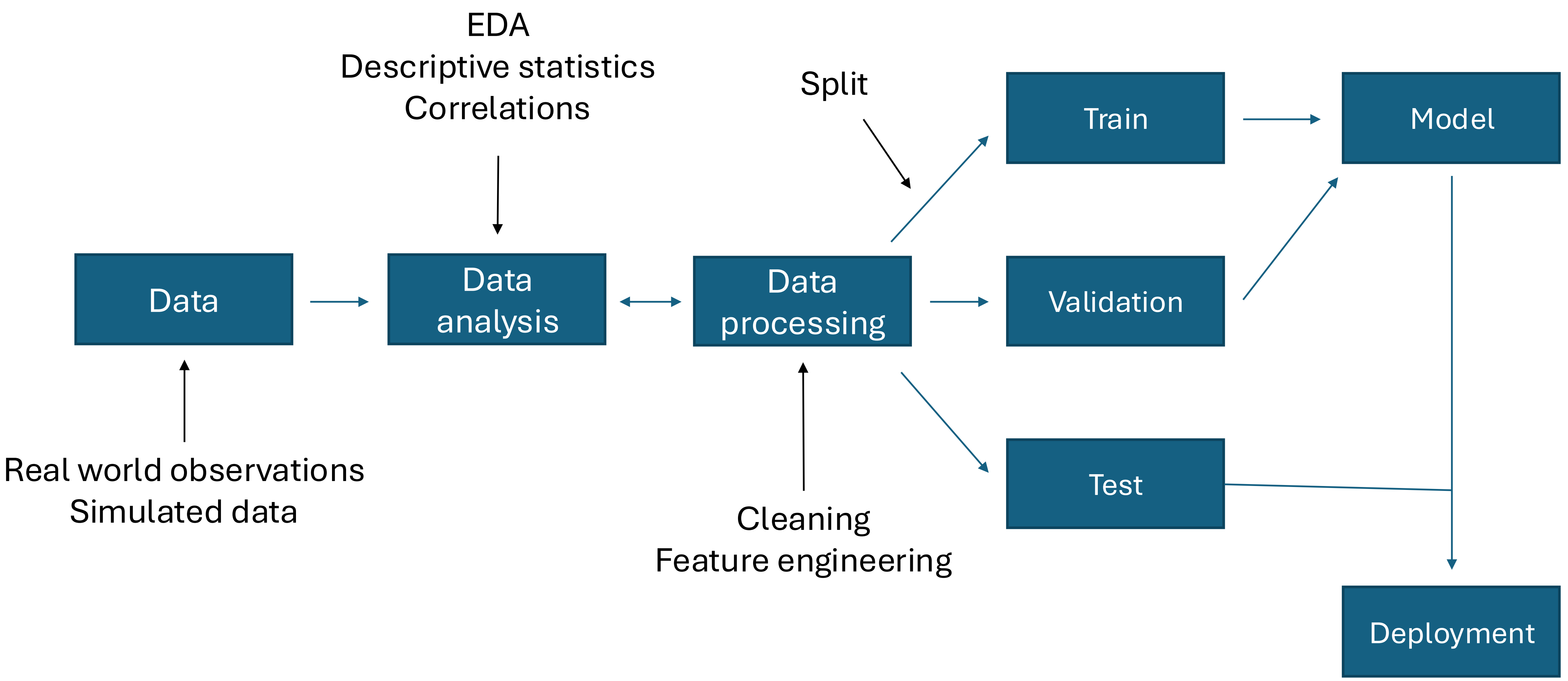
Dog

Supervised learning

UMAP projection

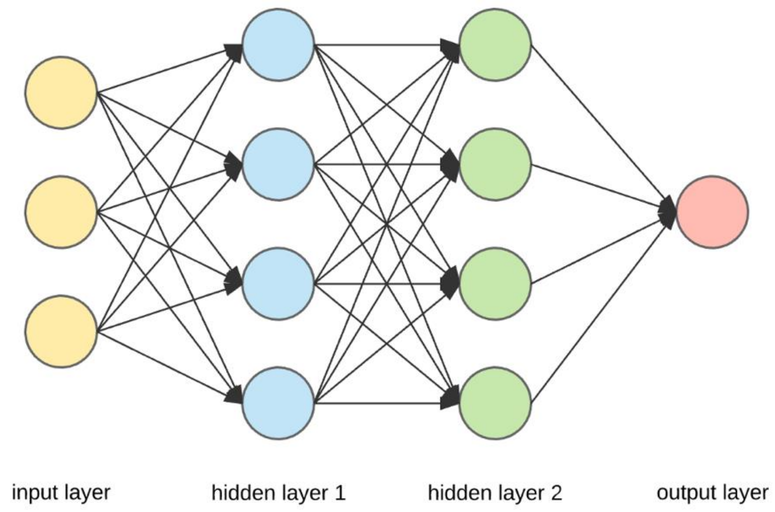
Unsupervised learning

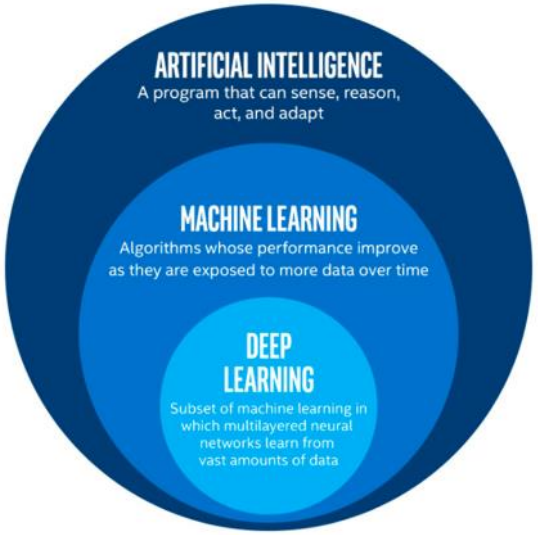
General Machine Learning procedure



Deep Learning

- Sub-field of Machine Learning that uses deep neural networks inspired by the human brain
- Very powerful models when large amounts of data are available





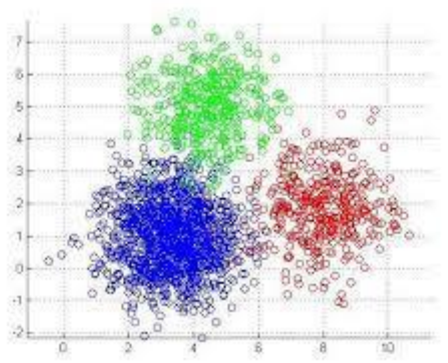

PUB
QUIZ
starts at
19:05
BILLIARD CLUB
Buffalo
SINCE 2017
Ungvár

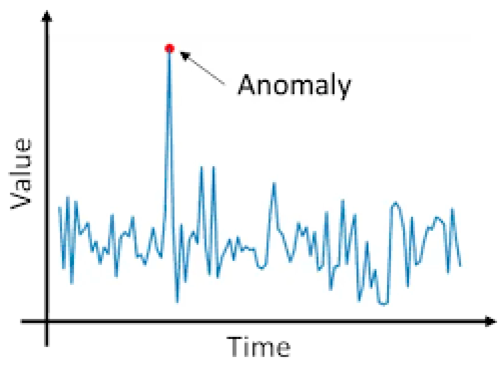
Supervised vs unsupervised learning

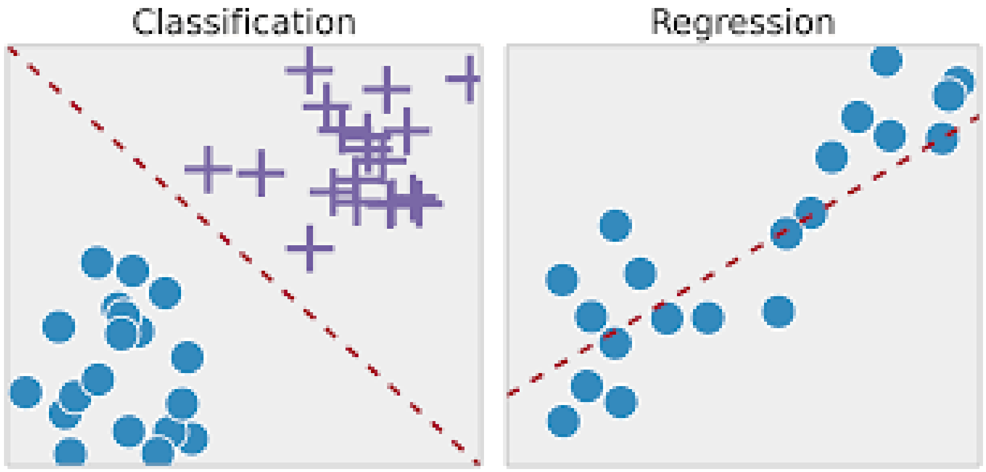
- Supervised: We use labels for our target -> Classification / regression
- Unsupervised: No labels ->

Dimensionality reduction, clustering,

Anomaly detection





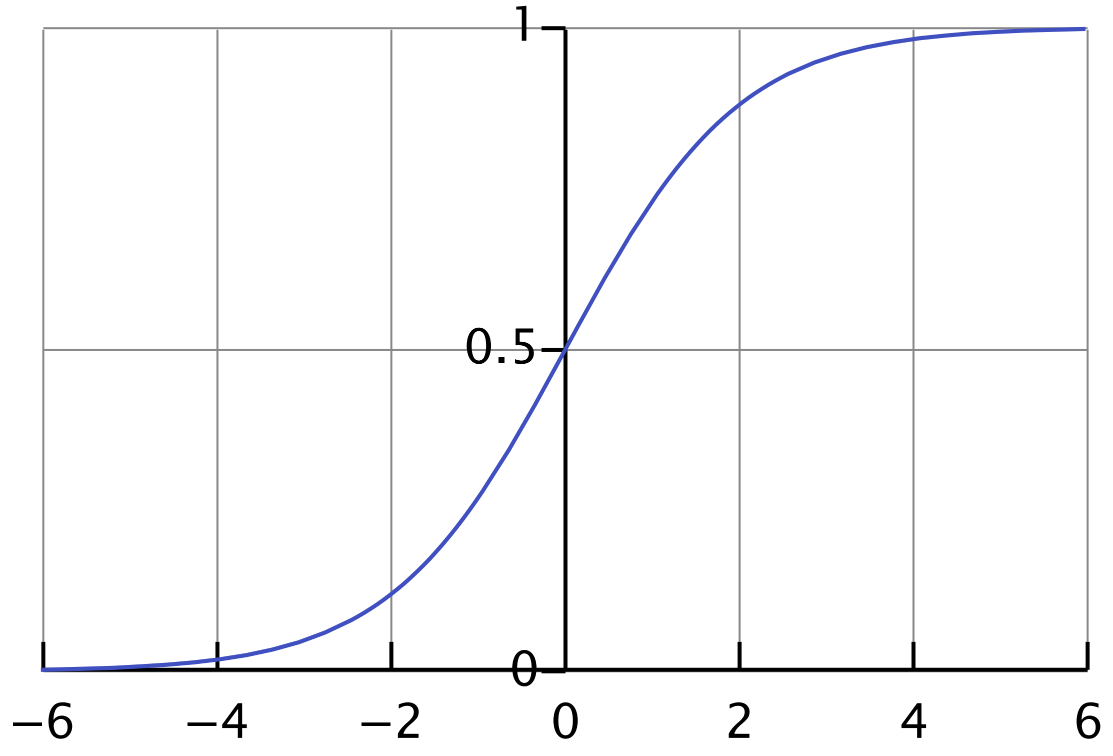


Classification example: Logistic regression

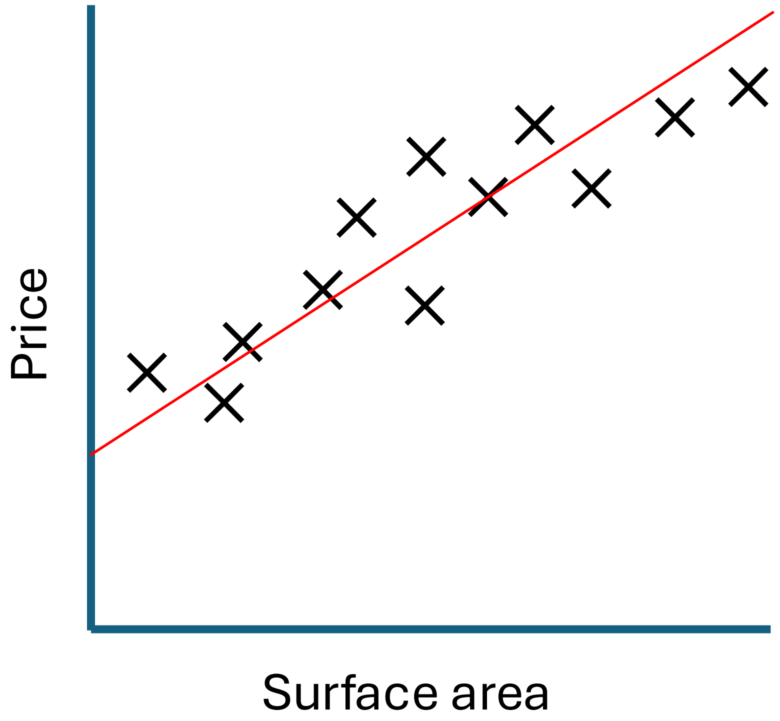
- Logistic function:

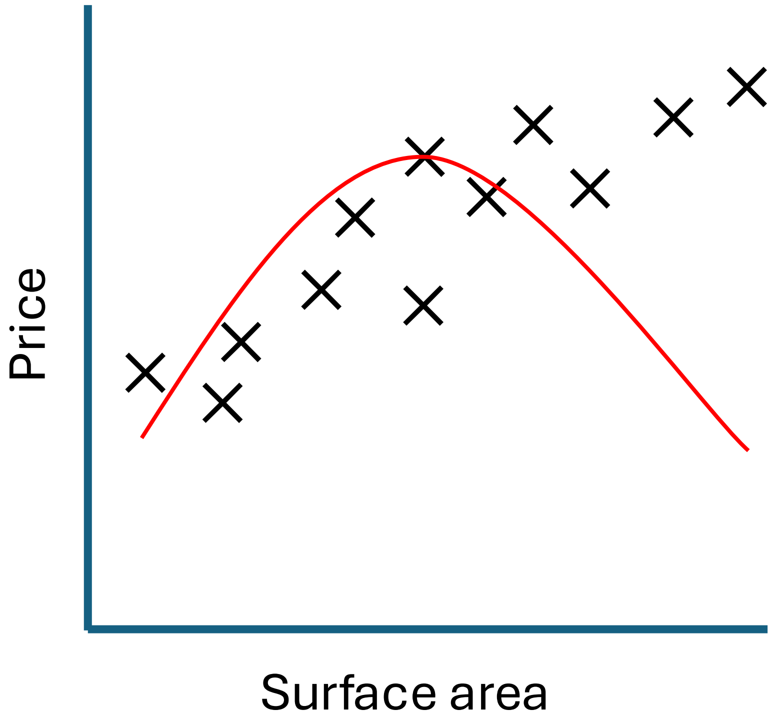
$$f(x) = \frac{1}{1 + e^{-x}}$$

- Output is a probability
- Used for classification



Bias-variance tradeoff

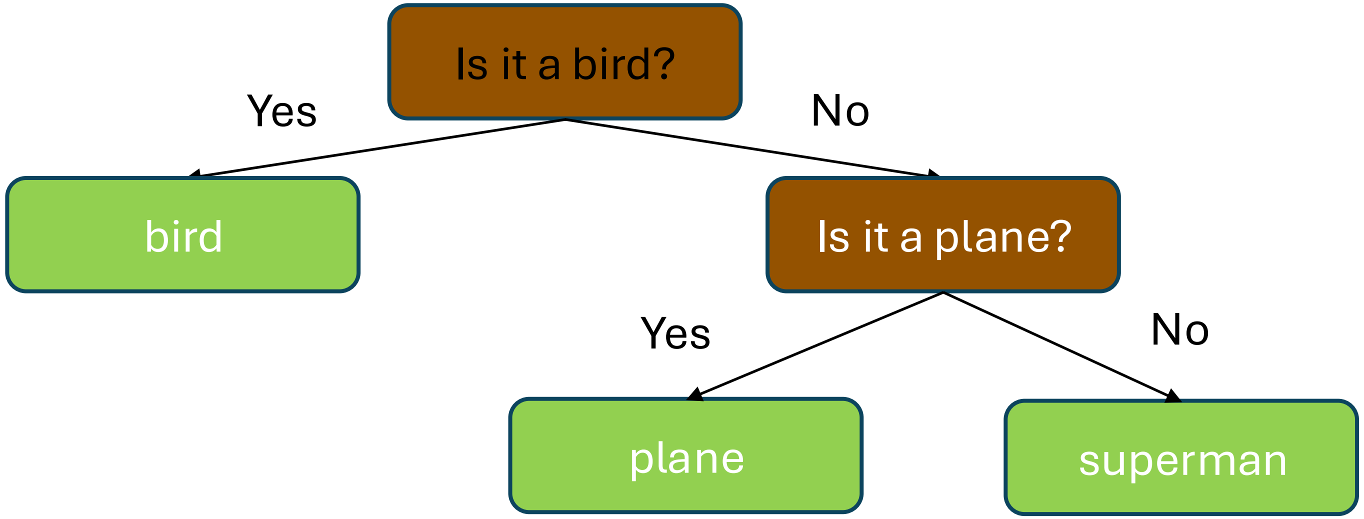




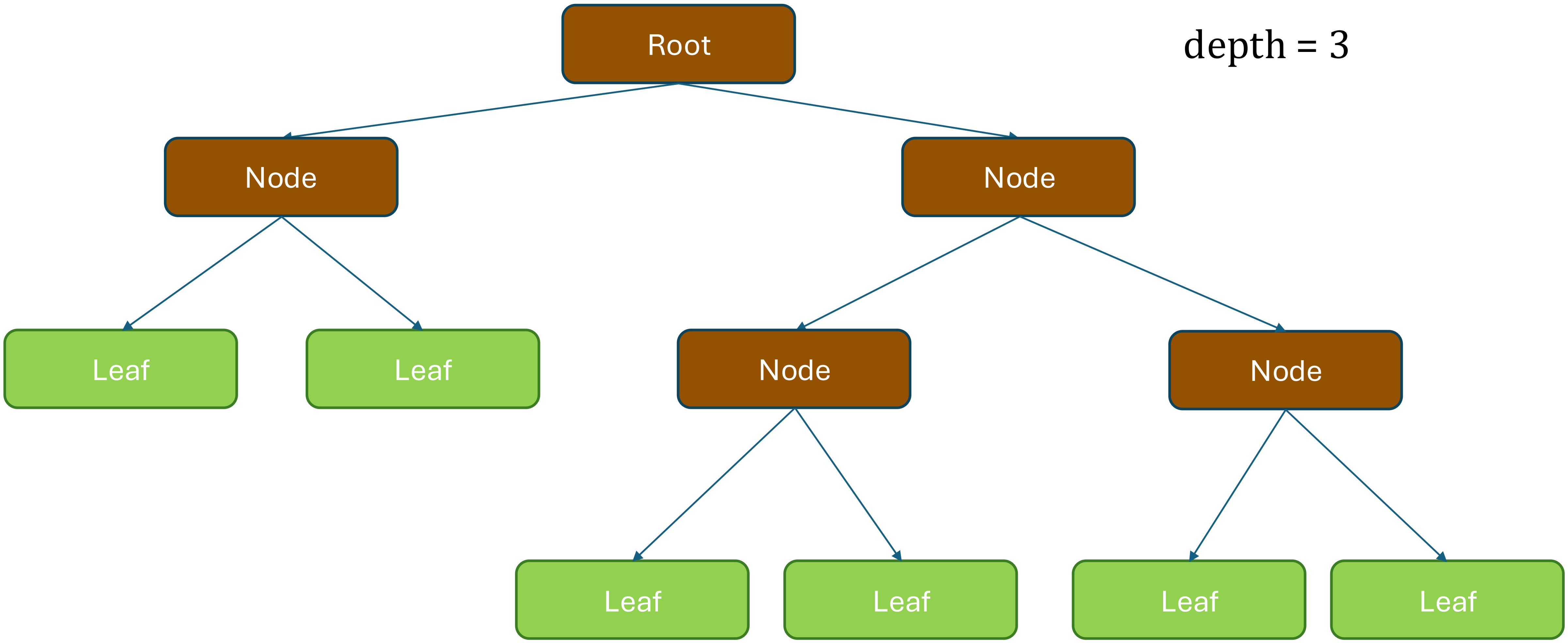
In statistics, the **bias of an estimator** (or **bias function**) is the difference between this estimator's expected value and the true value of the parameter being estimated. An estimator or decision rule with zero bias is called ***unbiased***. In statistics, "bias" is an *objective* property of an estimator

Decision trees

- A decision tree is a decision support hierarchical model that uses a tree-like model of decisions and their possible consequences.
- “*If* condition 1 *and* condition 2 *then* outcome”



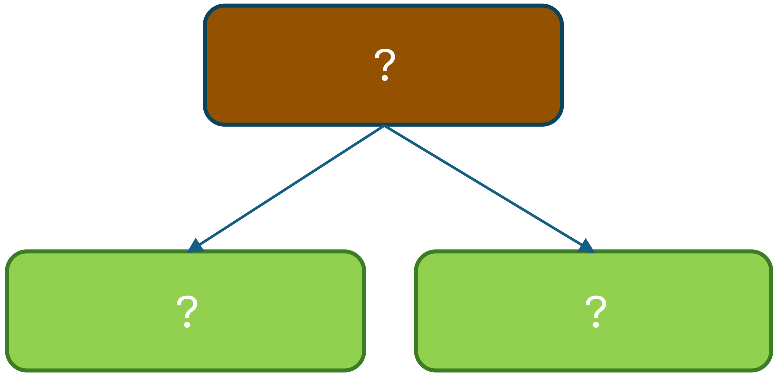
Decision trees



Decision tree learning

Tumor cells

Mean radius	Mean smoothness	Mean concavity	M or B?
14.97	0.08421	0.01947	0
15.7	0.09597	0.06593	1
13.48	0.1016	0.1063	1
14.5	0.1101	0.08842	0
17.46	0.09812	0.1417	1
12.47	0.09965	0.08005	0
11.89	0.1225	0.05929	0
13.61	0.09488	0.08625	1
14.6	0.08682	0.0839	1
11.26	0.0802	0.09274	0



Gini impurity

$$b_i = 1 - p_0^2 - p_1^2$$

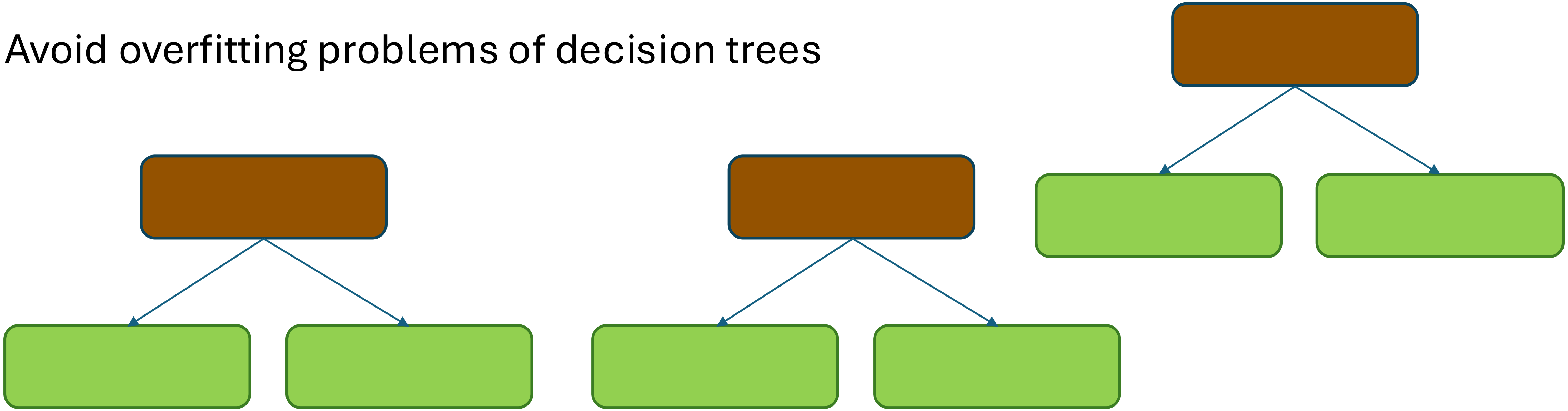
$$Gini(N) = \alpha b_1 + \beta b_2$$

Decision tree learning

- Easy to build, use and interpret
- Require little data preparation
- Perform well on large datasets
- Tend to overfit training data
- Generalization problems
- Limit the number of elements on each leaf
- Prune the trees

Random forests

- Ensemble learning method for classification and regression
- For classification tasks, the output is the majority voting of trees
- For regression tasks, the output is the mean of the prediction of all individual trees
- Avoid overfitting problems of decision trees



Random forests

- Create a bootstrapped dataset of the same size as the original dataset by selecting random samples from the dataset
- Build *n* decision trees, selecting at each step *m* variables
- Evaluate the random forest with the samples that didn’t make it to the bootstrapped dataset
- Repeat with a different *m*

PUB QUIZ

starts at
19:05

BILLIARD CLUB
SINCE Buffalo® 2017
Ungvár