Pattern Recognition Project 1

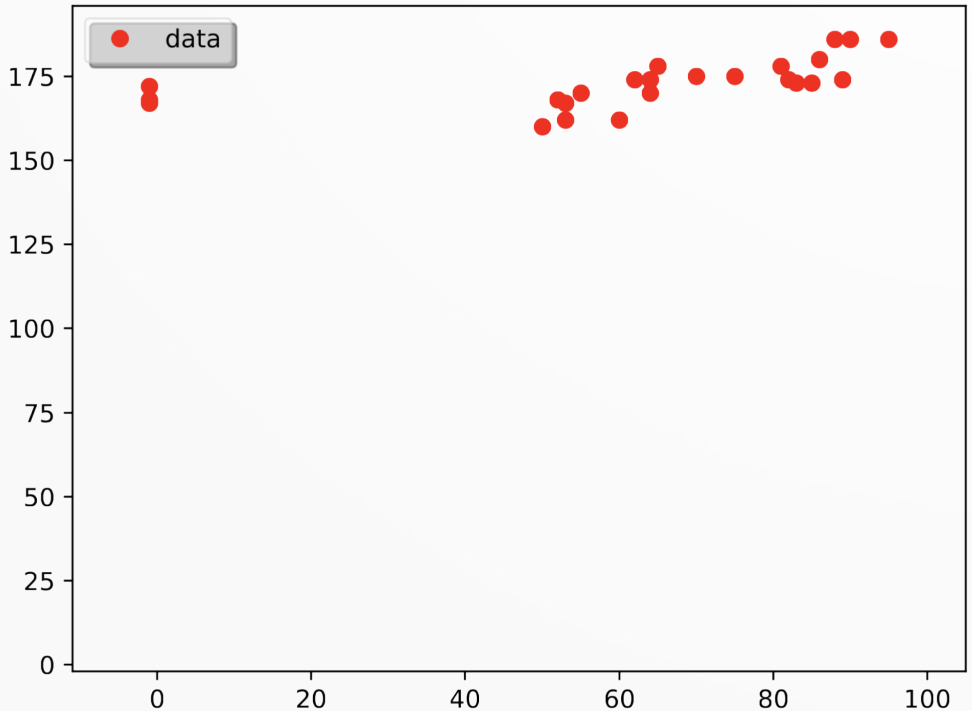
Lukas Drexler, Leif Van Holland, Reza Jahangiri, Mark Springer, Maximilian Thiessen

Task 1

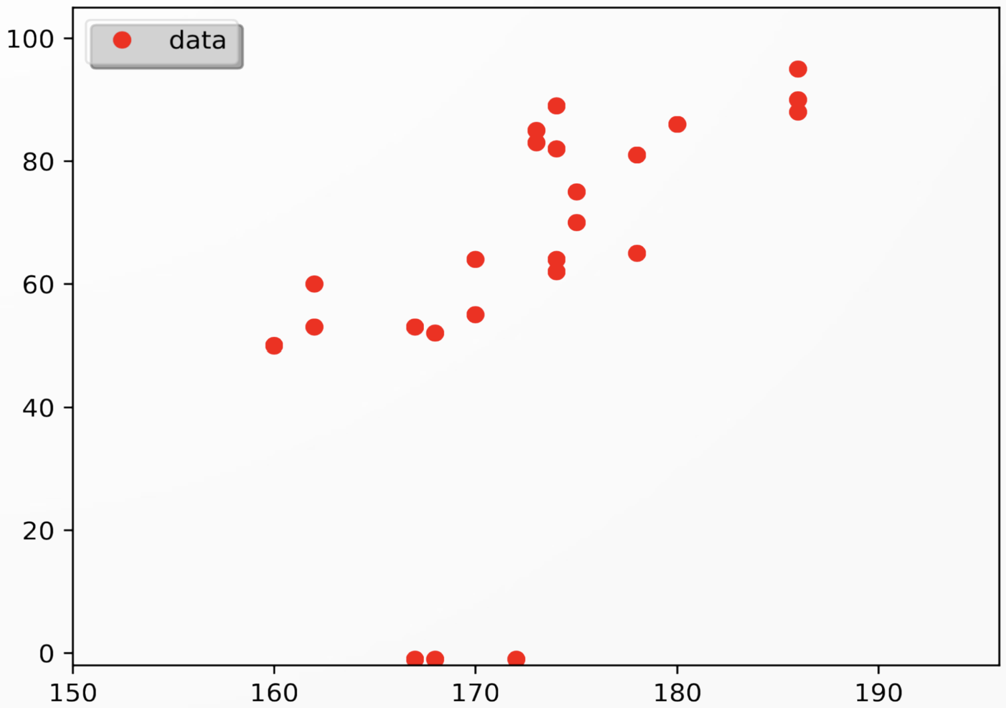
Task 1 – Outliers

Task: plot the data without outliers

weight vs. height



height vs. weight

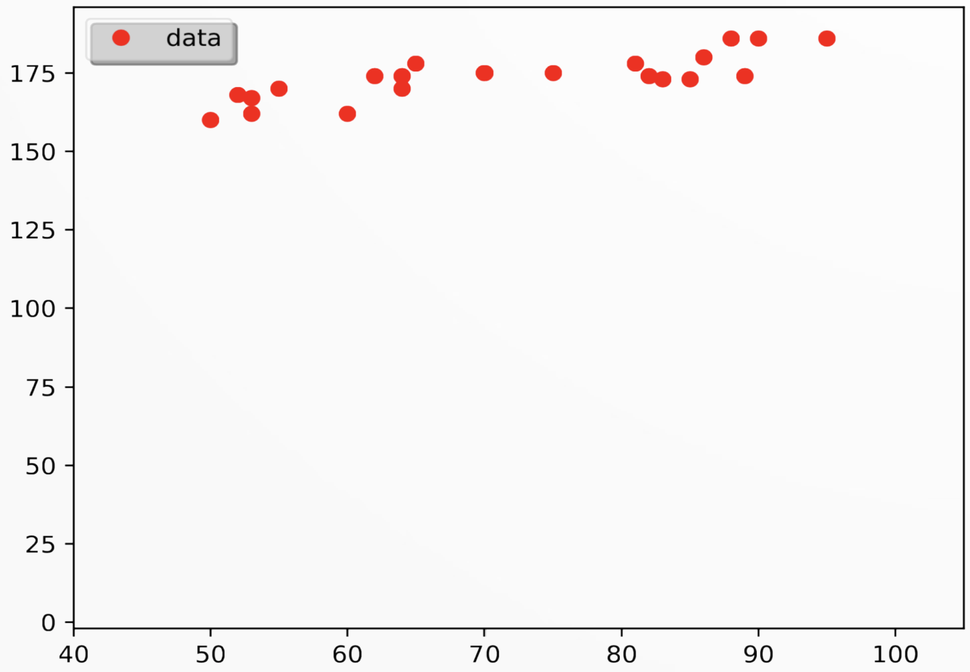


Task 1 – Outliers (cont.)

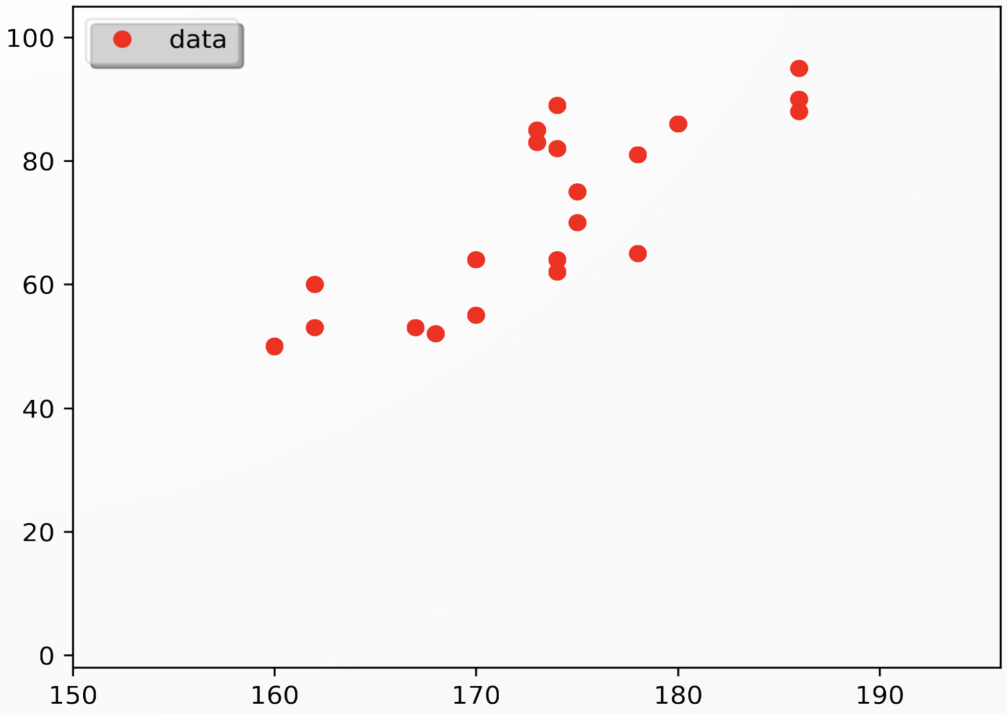
Use **np.all** to filter out lines with outliers

```
X = X[np.all(X > 0, axis=1)]
```

weight vs. height



height vs. weight



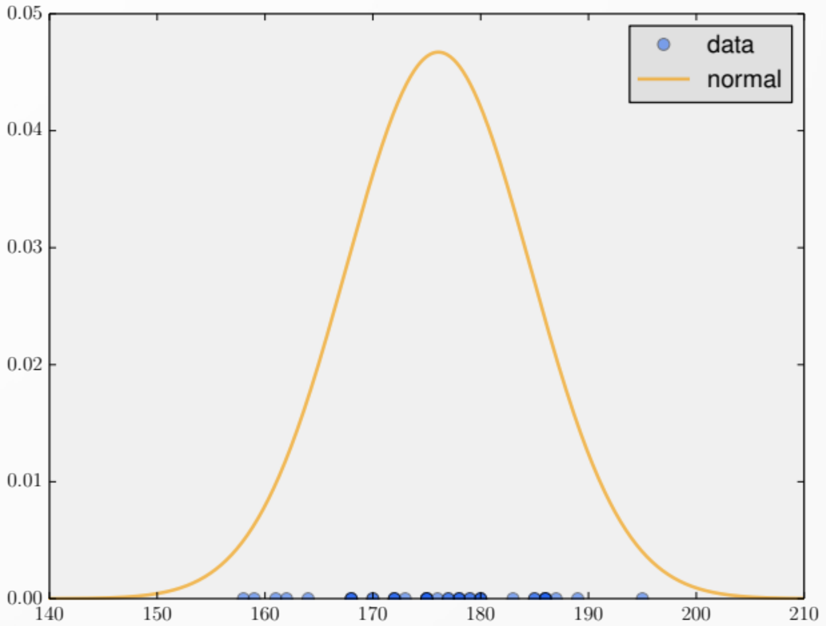
Task 2

Task 2 – 1D Gaussian

Given 1D array of data

Compute **mean**
and **standard deviation**

Plot data and **normal distribution**
characterizing its density



Task 2 – 1D Gaussian (cont.)

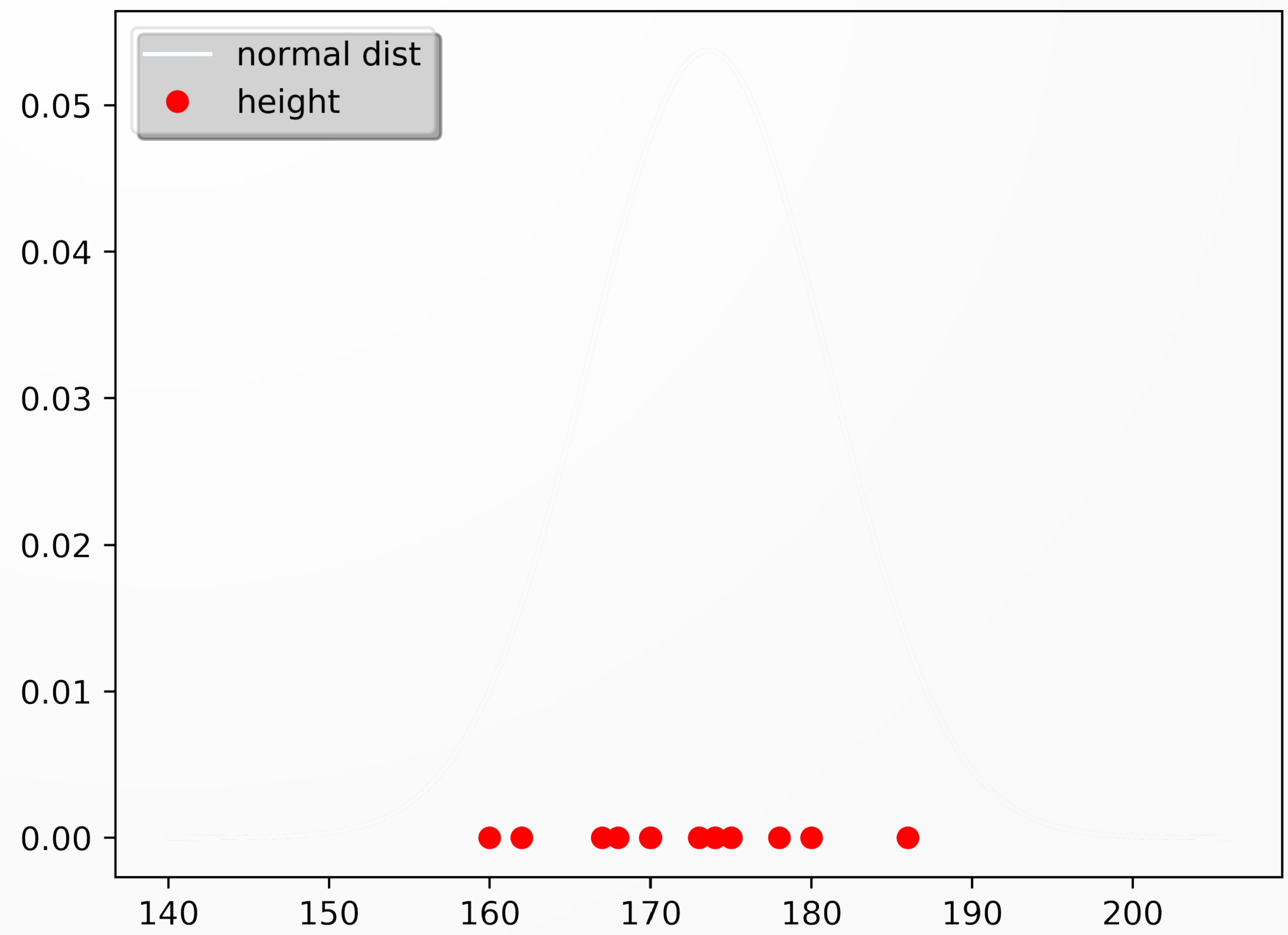
mean: $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

standard deviation:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$$

normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Task 2 – 1D Gaussian (cont.)

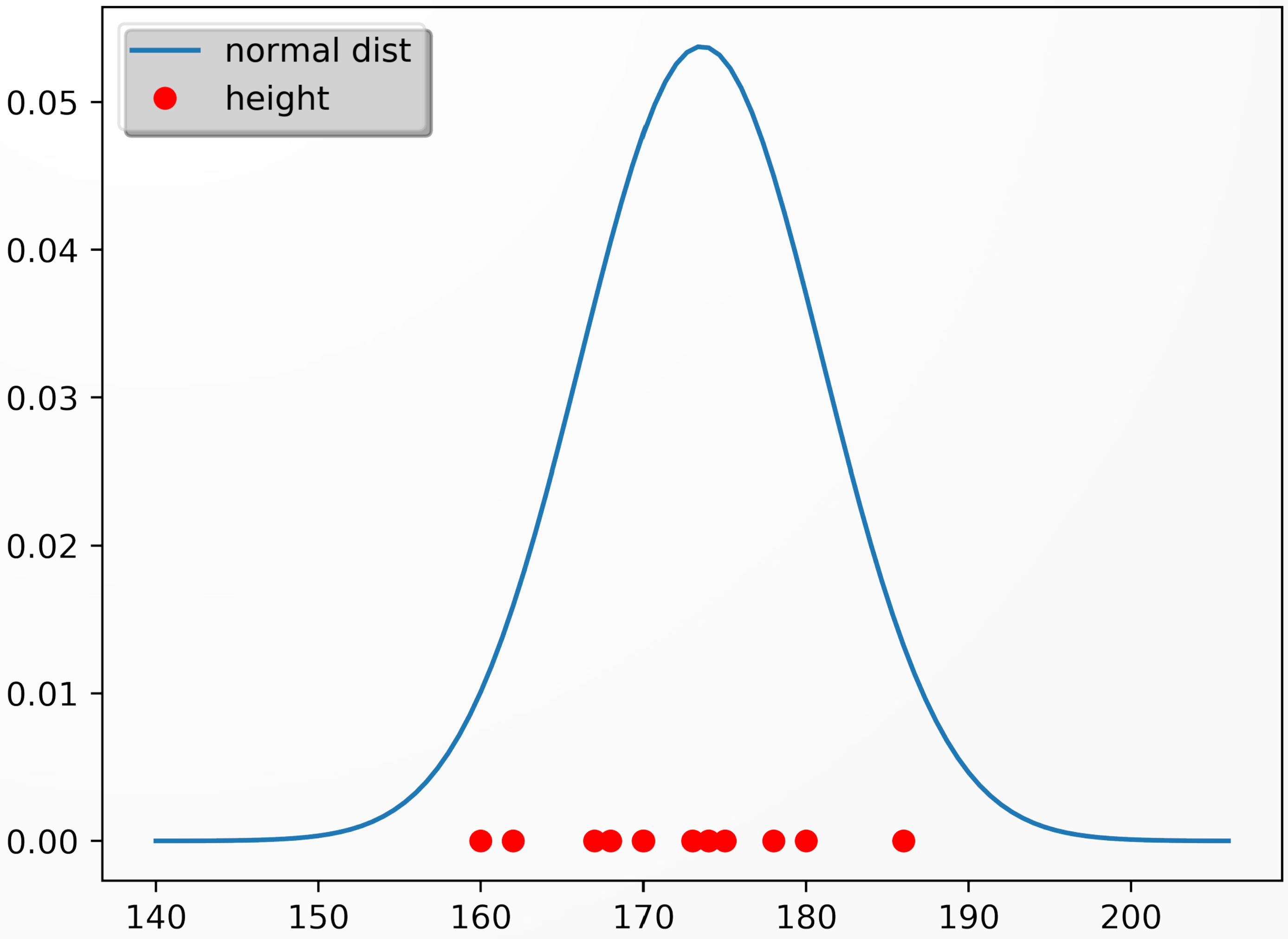
mean: $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

standard deviation:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2}$$

normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



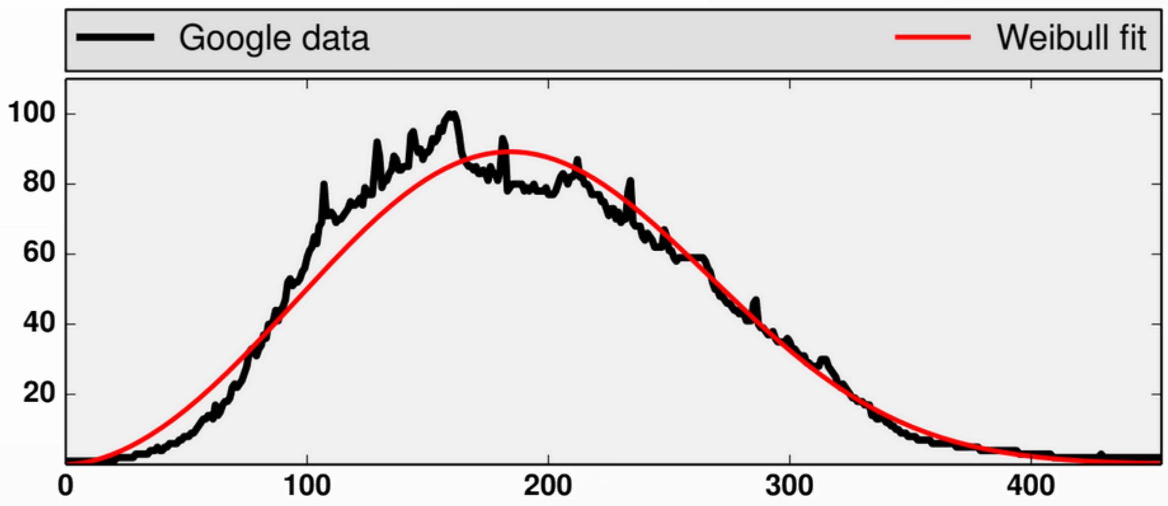
Task 3

Task 3

Fit Weibull distribution to given data

$$f(x \mid \kappa, \alpha) = \frac{\kappa}{\alpha}\left(\frac{x}{\alpha}\right)^{\kappa-1} e^{\left(\frac{x}{\alpha}\right)^{\kappa}}$$

Estimate parameters α and κ by **maximum likelihood**



Task 3 – Using the histogram

For a dataset $D = \{d_i\}_{i=1}^{n}$ consider its histogram $h(x_j) = h_j$

$$L(\alpha, \kappa \mid D) = N(\log \kappa - \kappa \log \alpha) + (\kappa - 1) \sum_i \log d_i - \sum_i \left(\frac{d_i}{\alpha} \right)^{\kappa}$$

Using the histogram:

$$L(\alpha, \kappa \mid D) = N(\log \kappa - \kappa \log \alpha) + (\kappa - 1) \sum_j x_j \cdot \log h_j - \sum_j \left(\frac{x_j \cdot h_j}{\alpha} \right)^{\kappa}$$

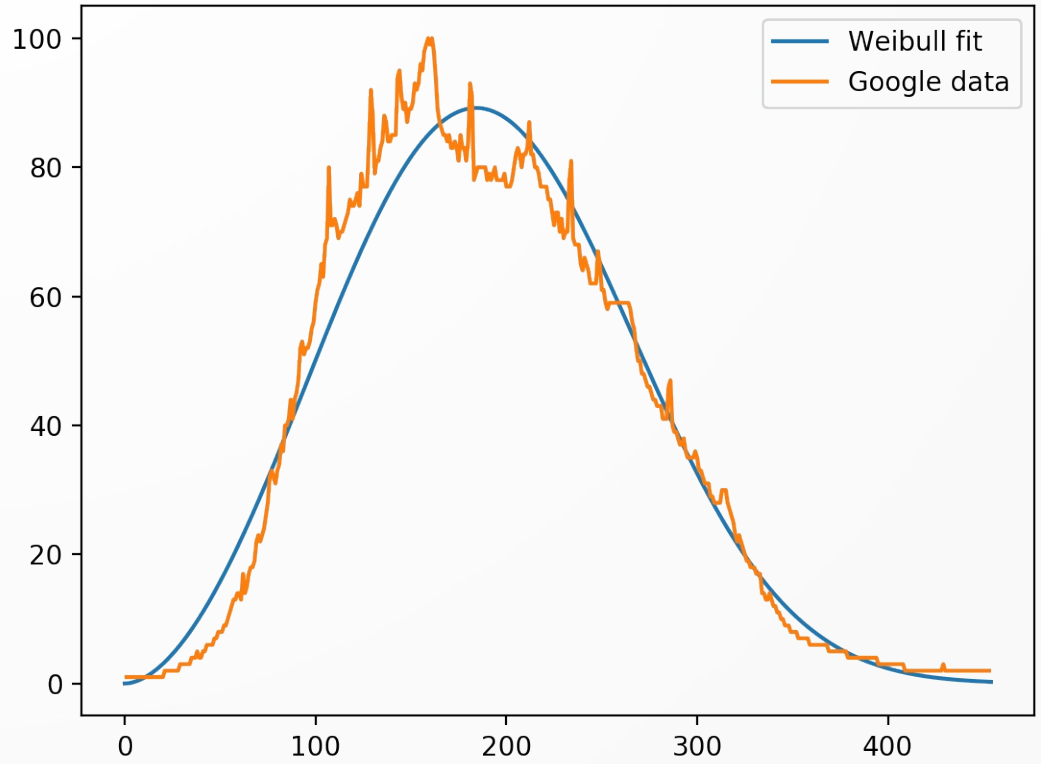
Likewise for partial derivatives

Task 3 – Newton's method

Estimate parameters using Newton's method for partial derivatives of L

$$\kappa_{20} \approx 2.80856$$

$$\alpha_{20} \approx 215.42857$$



Task 3 – Solving via `odeint`

`odeint` solves ODEs in the form $\frac{dy}{dt} = g(t, y)$ iteratively

consider ODE $x'(t) = \nabla L(x(t))$ for a function $x : \mathbb{R} \to \mathbb{R}^2$

with evenly spaced values $t_0, t_1, ..., t_n \in \mathbb{R}$

interpret x as $x(t_i) = (\kappa_i, \alpha_i)$

Task 3 – Solving via `odeint`

`odeint` integrates x numerically, starting at t_0

for given $x(t_0)$, step size h we have

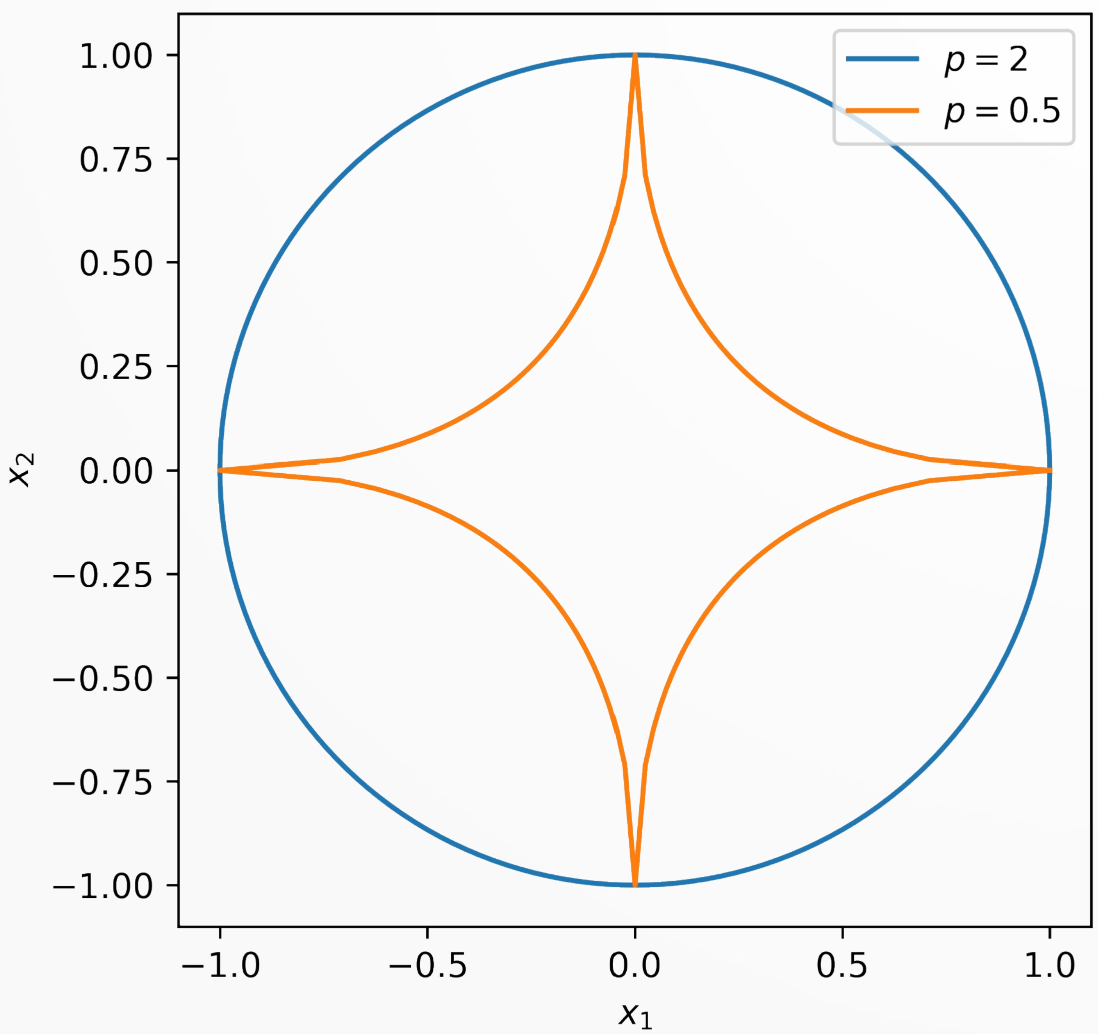
$$x(t_{i+1}) = x(t_i) + h \cdot \nabla L(x(t_i))$$

method eventually settles in a local optimum

for $\kappa_0 = 5,\ \alpha_0 = 100$ we found same optimum as before

Task 4

Task 4 – (not) a norm



Consider $\|\cdot\|_{0.5} \colon \mathbb{C}^n \mapsto \mathbb{R}$

defined like a **p-norm** for

$$p = \frac{1}{2}$$

$$\|(x_1, \ldots, x_n)^T\|_{0.5} := \left(\sum_{i=0}^{n} |x_i|^{\frac{1}{2}} \right)^2$$

Task 4 – (not) a norm

$$\|(x_1, ..., x_n)^T\|_{0.5} := \left(\sum_{i=0}^{n} |x_i|^{\frac{1}{2}}\right)^2$$

for it to be a norm, following **axioms** have to hold:

$\forall x \in \mathbb{C}^n : \ \|x\|_{0.5} = x \Rightarrow x = 0$ (definiteness)

$\forall x \in \mathbb{C}, \alpha \in \mathbb{R} : \ \|\alpha x\|_{0.5} = |\alpha| \cdot \|x\|_{0.5}$ (abs. homogeneity)

$\forall x, y \in \mathbb{C}^n : \ \|x + y\|_{0.5} \leq \|x\|_{0.5} + \|y\|_{0.5}$ (triangle inequality)

Task 4 – (not) a norm

$$\|(x_1, ..., x_n)^T\|_{0.5} := \left(\sum_{i=0}^{n} |x_i|^{\frac{1}{2}}\right)^2$$

for it to be a norm, following **axioms** have to hold:

$\forall x \in \mathbb{C}^n : \; \|x\|_{0.5} = x \Rightarrow x = 0$ (definiteness)

$\forall x \in \mathbb{C}, \alpha \in \mathbb{R} : \; \|\alpha x\|_{0.5} = |\alpha| \cdot \|x\|_{0.5}$ (abs. homogeneity)

$\forall x, y \in \mathbb{C}^n : \; \|x + y\|_{0.5} \leq \|x\|_{0.5} + \|y\|_{0.5}$ (triangle inequality)

Task 4 – (not) a norm

$$\|(x_1, \ldots, x_n)^T\|_{0.5} := \left(\sum_{i=0}^{n} |x_i|^{\frac{1}{2}}\right)^2$$

$$\boxed{\forall x, y \in \mathbb{C}^n : \ \|x + y\|_{0.5} \leq \|x\|_{0.5} + \|y\|_{0.5} \quad \text{(triangle inequality)}}$$

consider unit vectors $\mathrm{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Task 4 – (not) a norm

$$\|(x_1, \dots, x_n)^T\|_{0.5} := \left(\sum_{i=0}^{n} |x_i|^{\frac{1}{2}}\right)^2$$

$\forall x, y \in \mathbb{C}^n : \ \|x + y\|_{0.5} \leq \|x\|_{0.5} + \|y\|_{0.5}$ (triangle inequality)

consider unit vectors $\mathrm{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$$\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_{0.5} = (\sqrt{1} + \sqrt{1})^2 = 4 > 2 = \left\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_{0.5} + \left\| \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\|_{0.5}$$

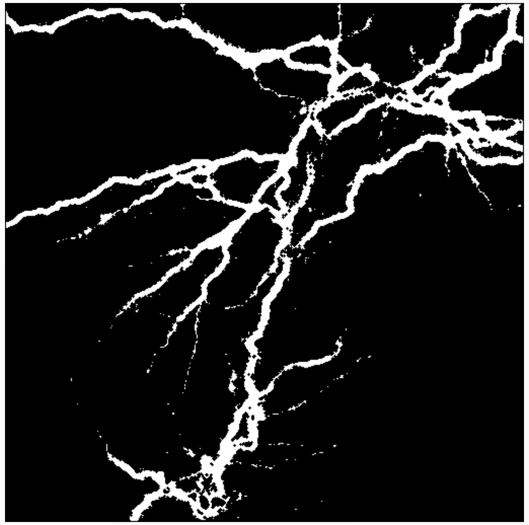
Task 5

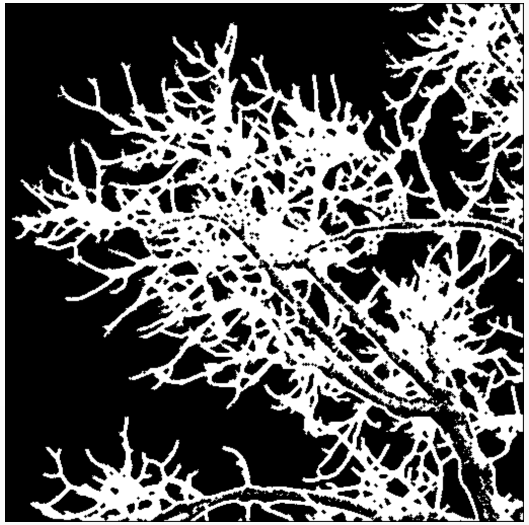
Task 5 – Box counting dimension

calculate box counting dimension

3 steps

- binarize image
- count boxes containing white pixels
- linear regression





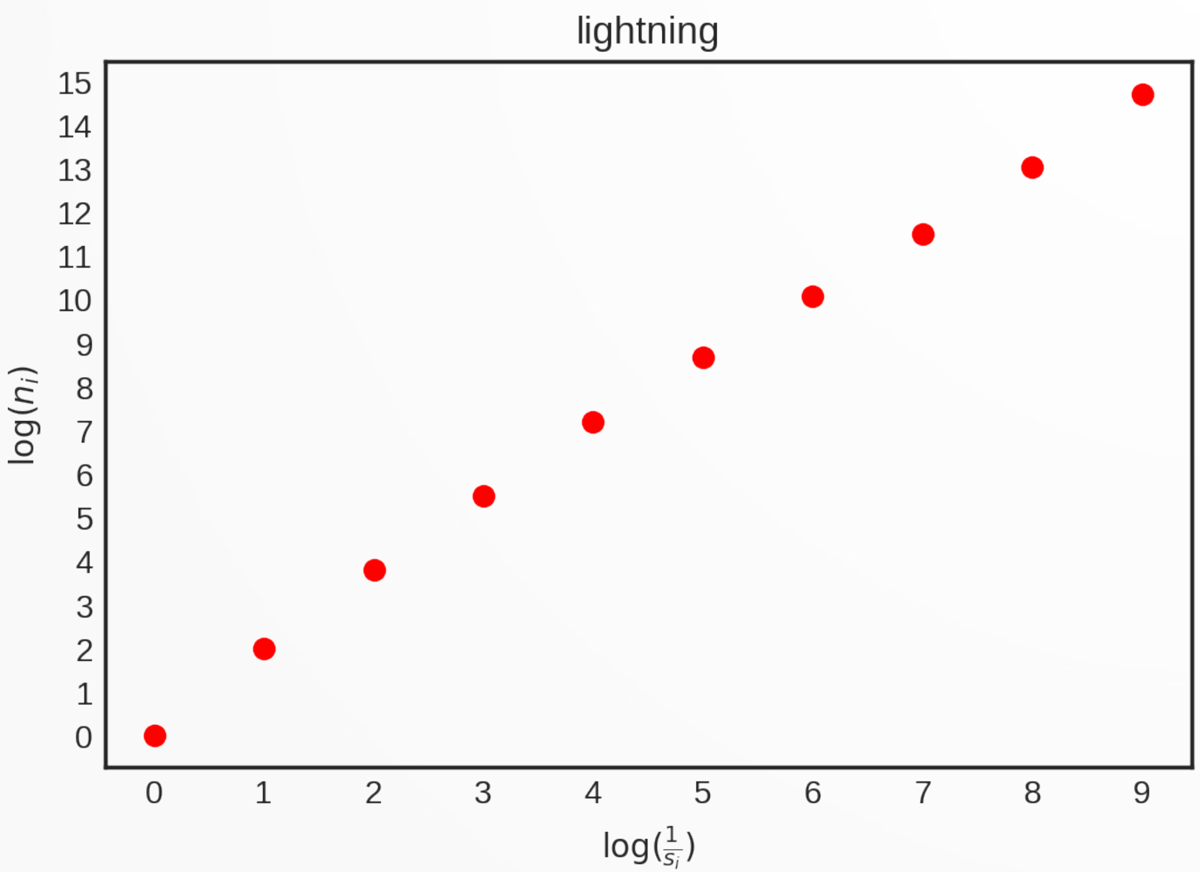
Task 5 – Box counting

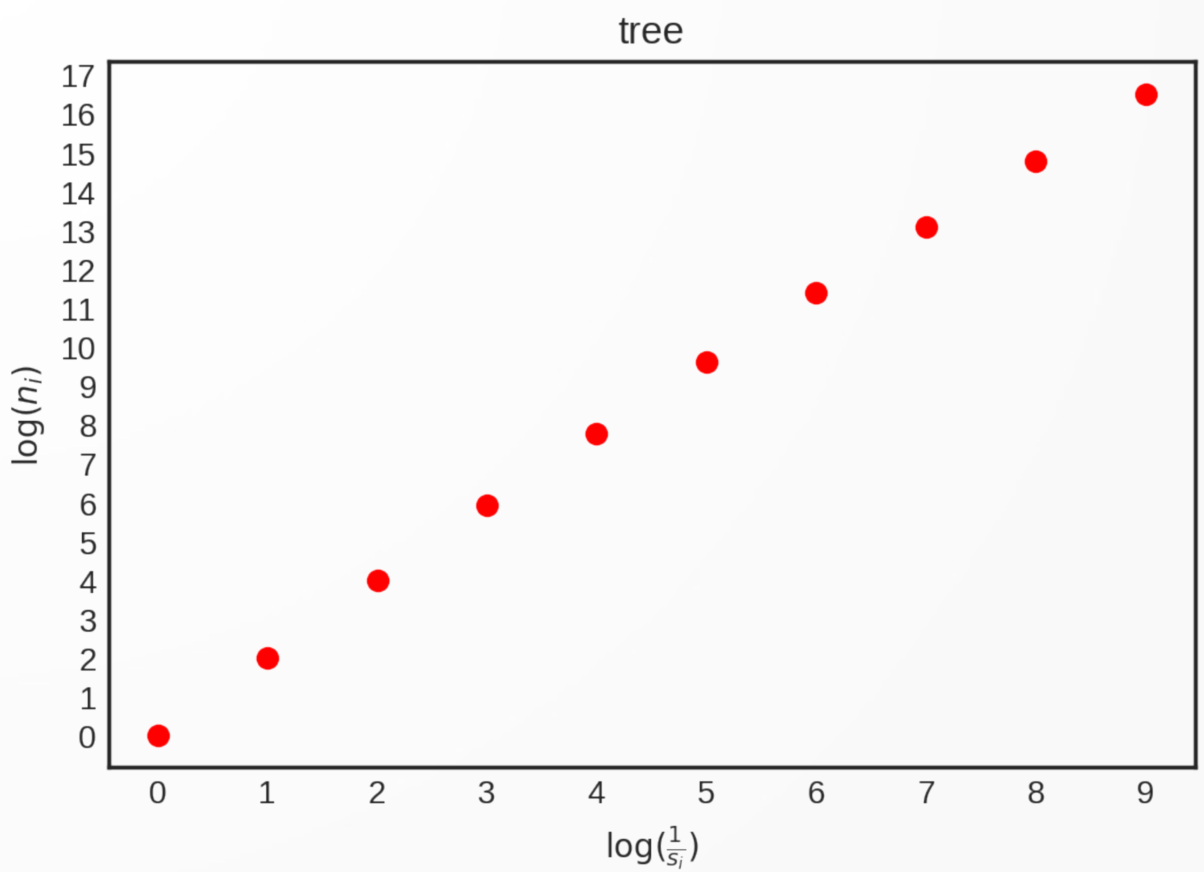
split matrix into four equally sized submatrices

count number of submatrices containing at least one white pixel

recursively split those submatrices

Task 5 – Box counting





Task 5 – Linear regression

regression by least-squares

- find a,b, such that

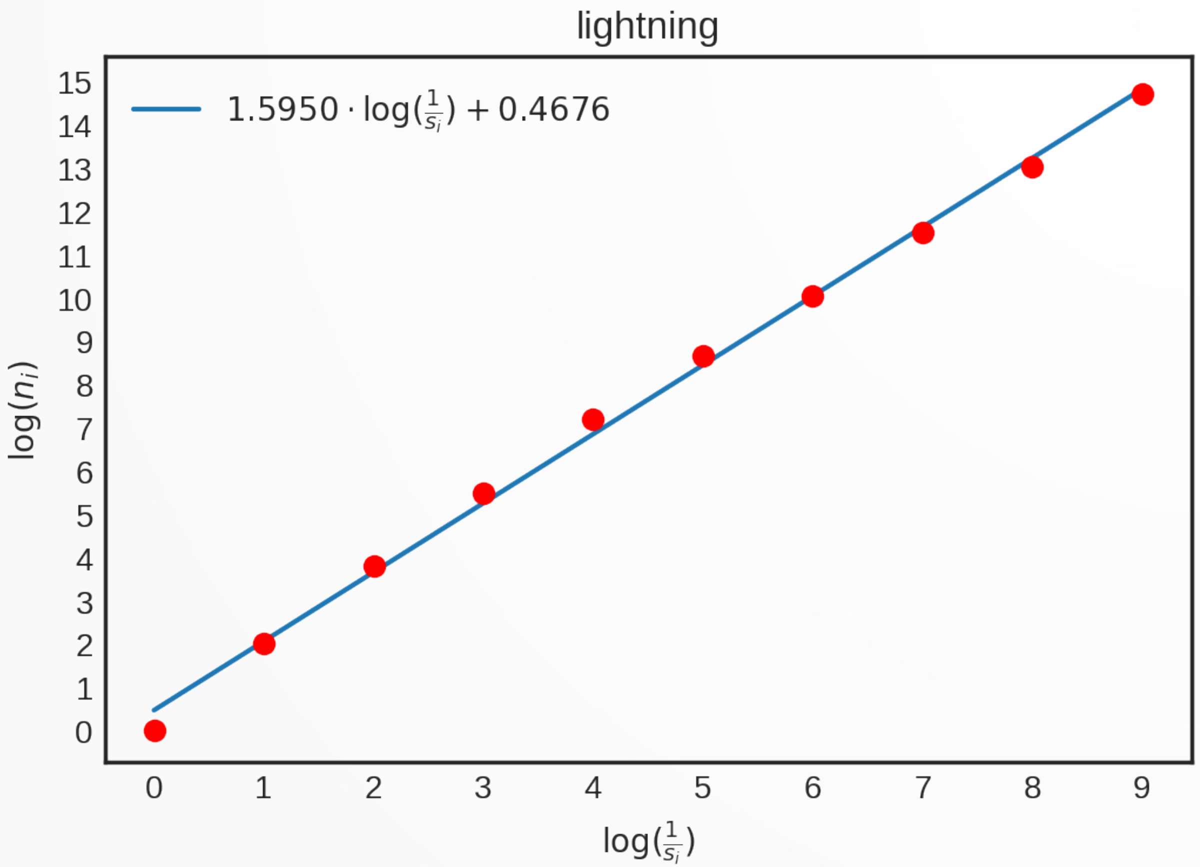
$$\sum_i \left(y_i - (ax_i + b)\right)^2$$

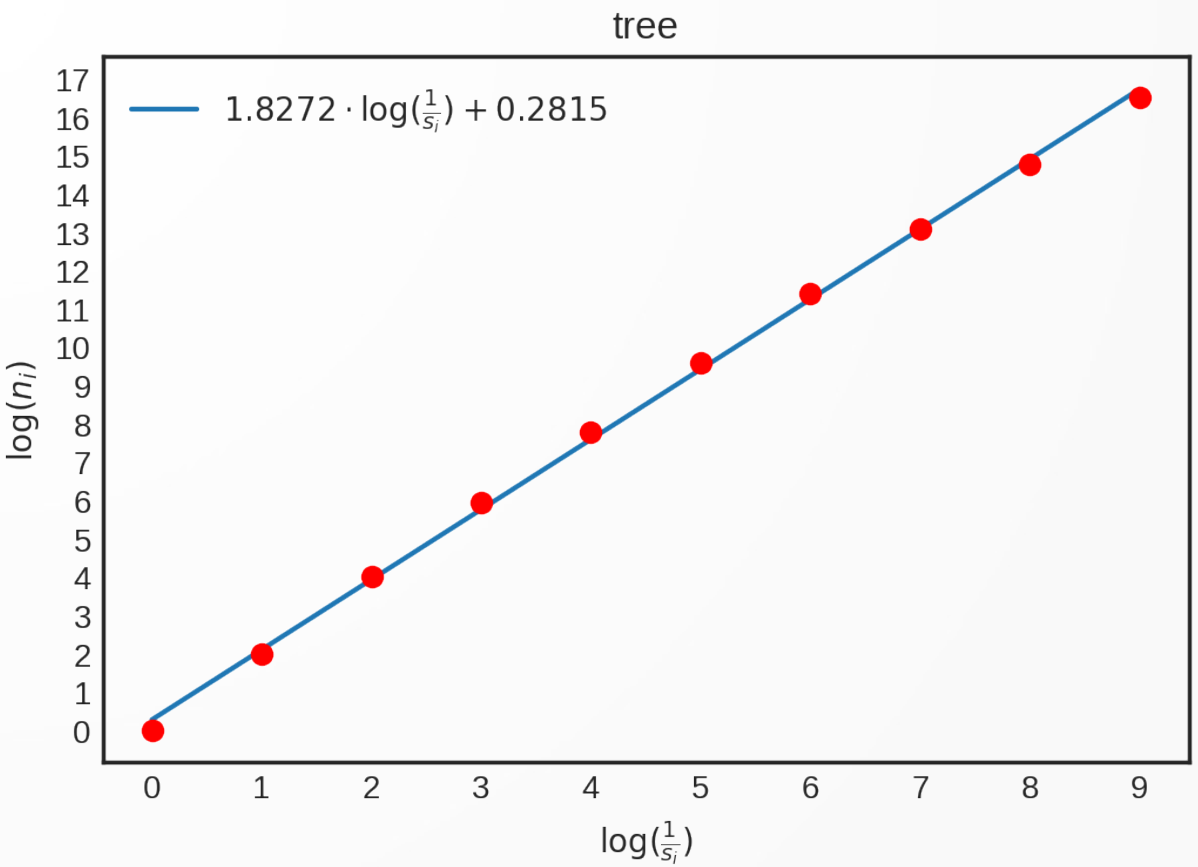
is minimized

matrix-vector formulation (as shown in lecture)

$$w = \begin{bmatrix} a \\ b \end{bmatrix} = \left(X^T X\right)^{-1} X^T y$$

Task 5 – Linear regression





dimension $\Leftrightarrow$ slope

- lightning: D = 1.595
- tree: D = 1.8272

Thank you for your attention!