Machine Learning for Exploration Geophysics

Th2: Basic Principles of Supervised Learning

10. - 12. March 2020

Hamburg

Outline

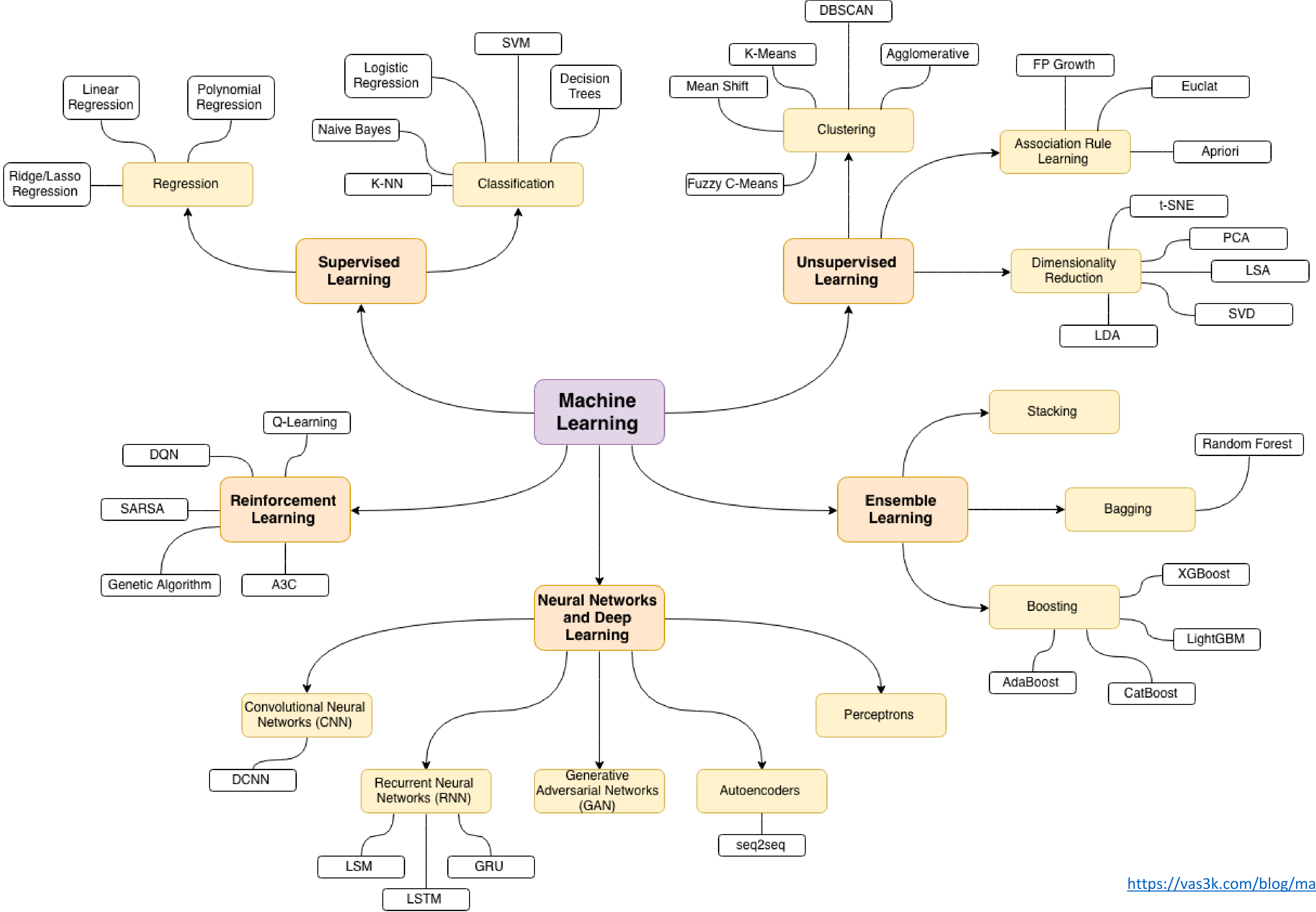
- Three components of supervised learning
- Regression
  - Linear regression
  - Polynomial regression
  - Gradient descent
  - Regularization
  - Training / CV / Test
  - k-Fold Cross-Validation
  - Hyperparameter tuning

Three components of supervised learning

Goal of supervised learning algorithms: predict results based on incoming data

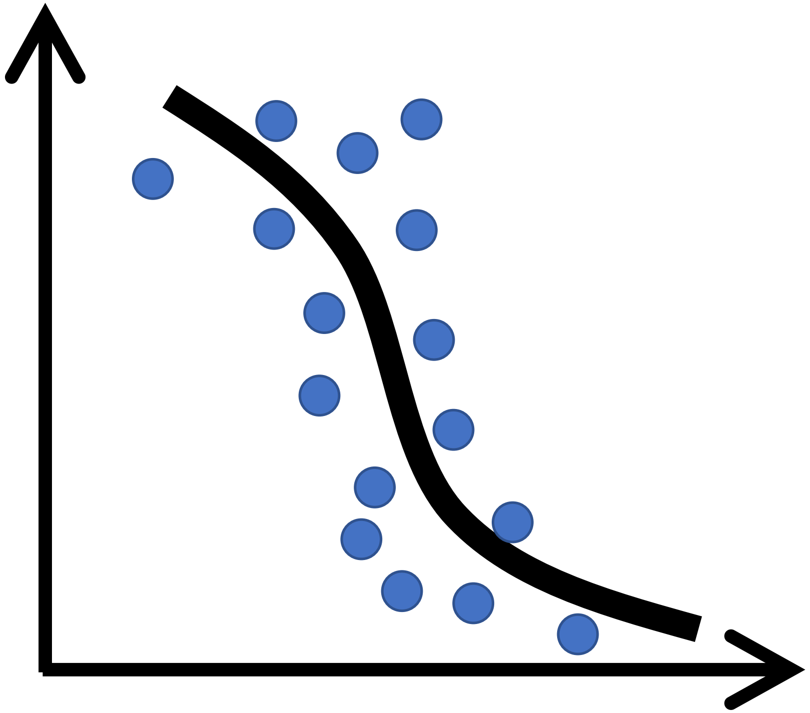
Three components to teach the machine:

- Data
  - The more diverse the data, the better the result!
  - Two main ways to get the data: manual and automatic
- Features
  - Feature selecting is time consuming (AutoML)
  - Subjective selection is the main source of errors
- Algorithms
  - Any problem can be solved!
  - But, the precision, performance, and size of the final model depend on the algorithm
  - “Garbage in - Garbage out”

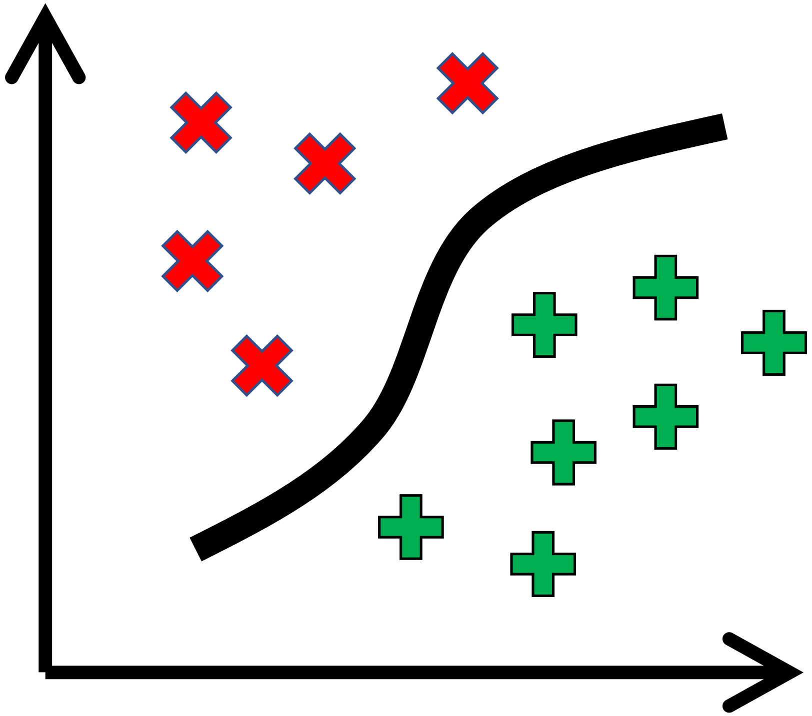


https://vas3k.com/blog/machine_learning/

Regression vs Classification



Regression



Classification

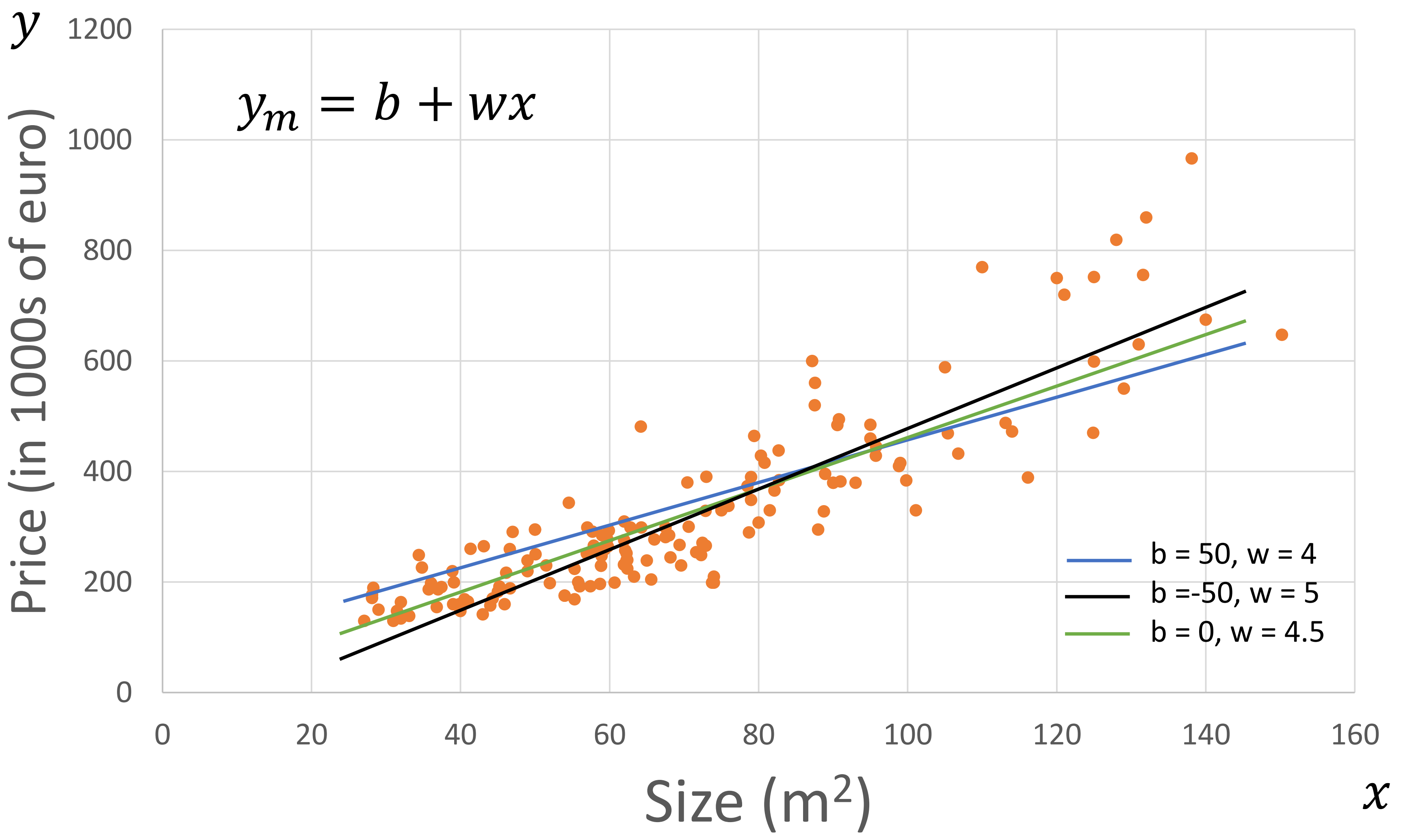
Housing Prices (Berlin Neukölln)

Size in m^2 (x)	Price in 1000's of euro (y)
41,34	260,505
54,52	343,476
28,12	171,532
34,81	226,265
…	…

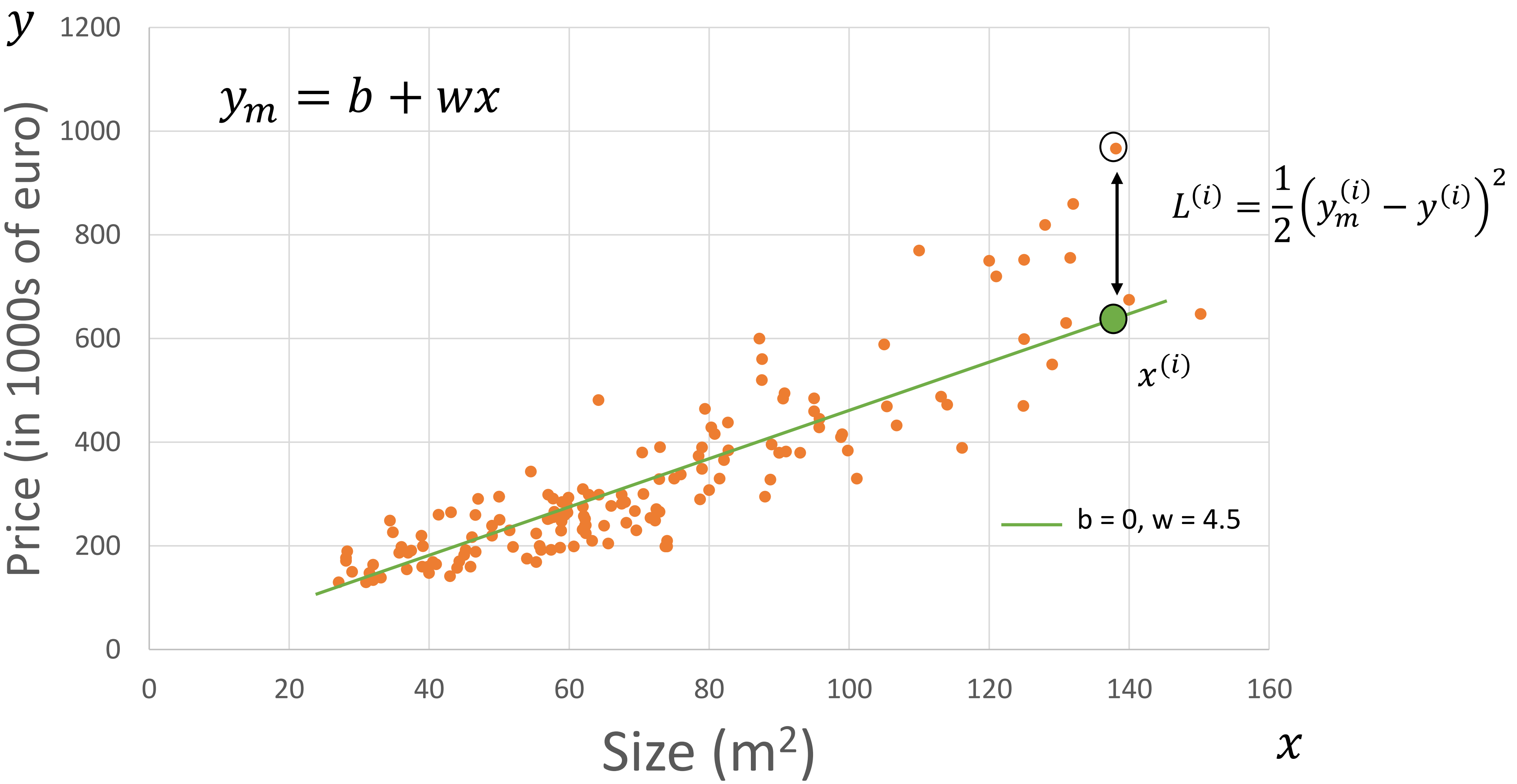
- Linear regression:

$$y_m = b + wx$$

Housing Prices (Berlin Neukölln)



Housing Prices (Berlin Neukölln)



- Linear regression:

$$F(x) = b + wx$$

- Parameters:

$$b, w$$

- Cost Function:

$$J(b, w) = \frac{1}{2N}\sum_{i=1}^{N}\left(F\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Goal:

$$\underset{b,w}{\text{minimize}}\left(J(b, w)\right)$$

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N)} \end{bmatrix} = \underset{\boldsymbol{X}}{\begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \dots & \dots \\ 1 & x^{(N)} \end{bmatrix}} \underset{\boldsymbol{w}}{\begin{bmatrix} b \\ w \end{bmatrix}} + \boldsymbol{\epsilon}$$

$$\boxed{\boldsymbol{w} = \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}}$$

- Linear regression:

$$F(x) = b + wx$$

- Parameters:

$$b, w$$

- Cost Function:

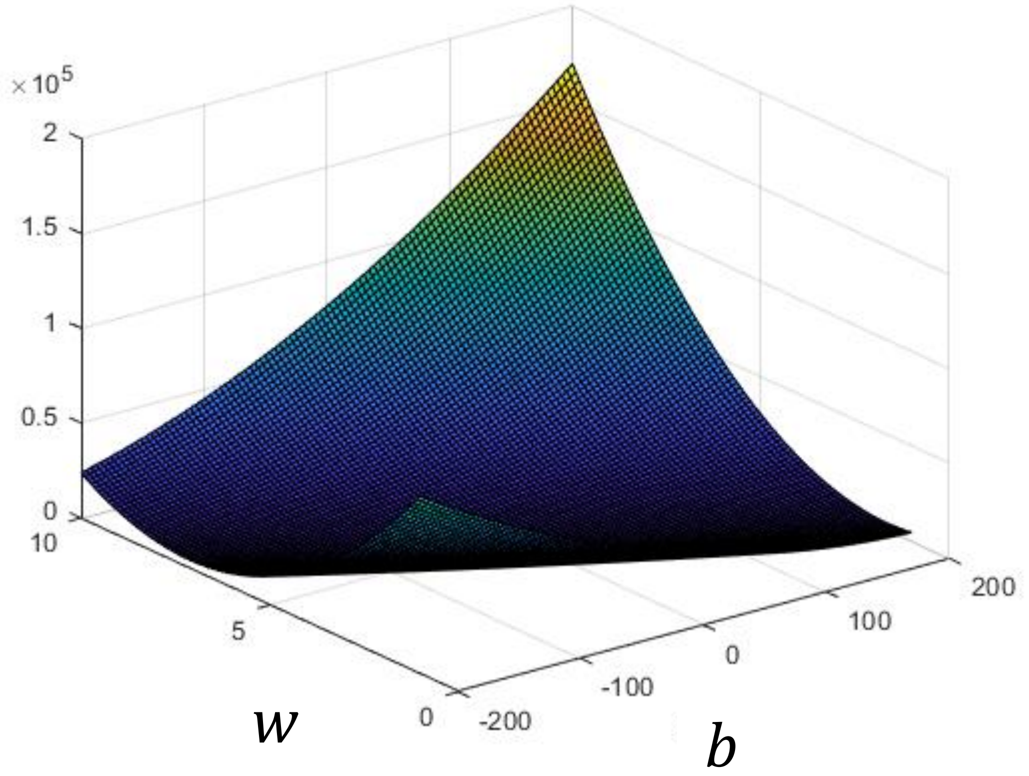
$$J(b,w) = \frac{1}{2N}\sum_{i=1}^{N}\left(F\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Goal:

$$\underset{b,w}{\text{minimize}}\left(J(b,w)\right)$$

$$J(b,w)$$

(function of the parameters b, w)

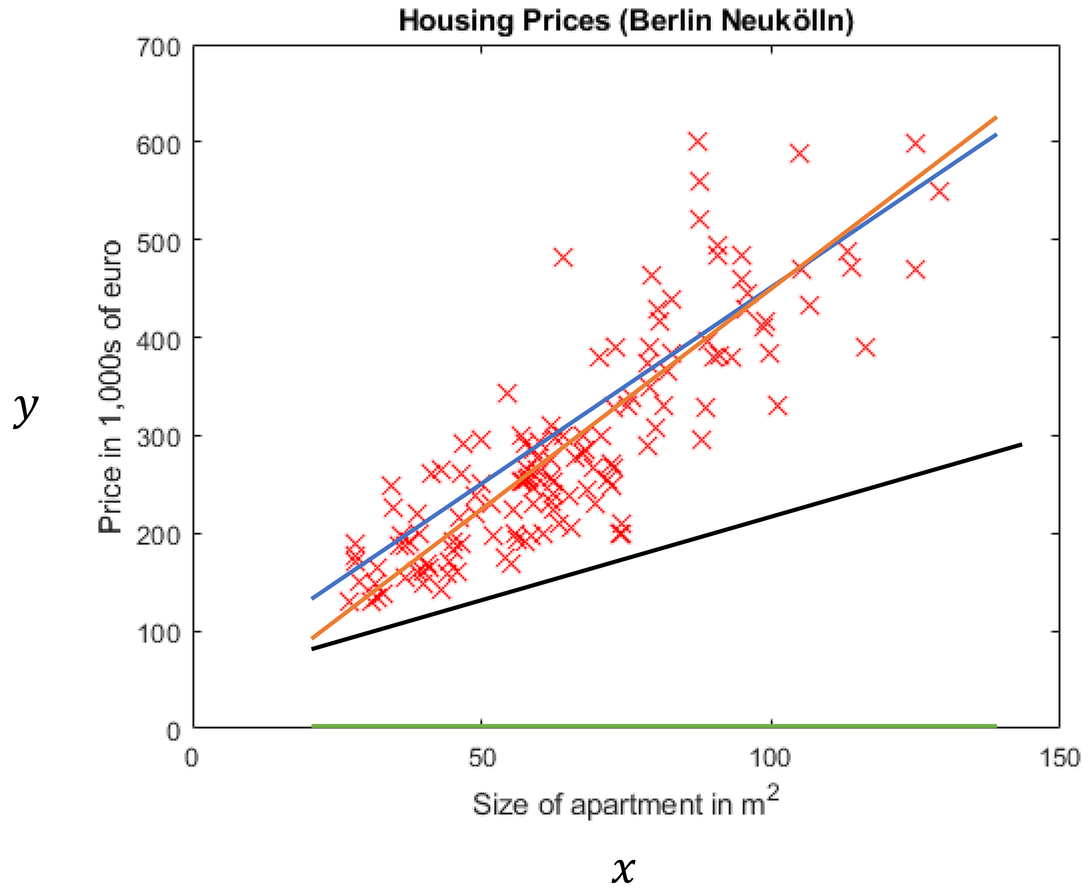


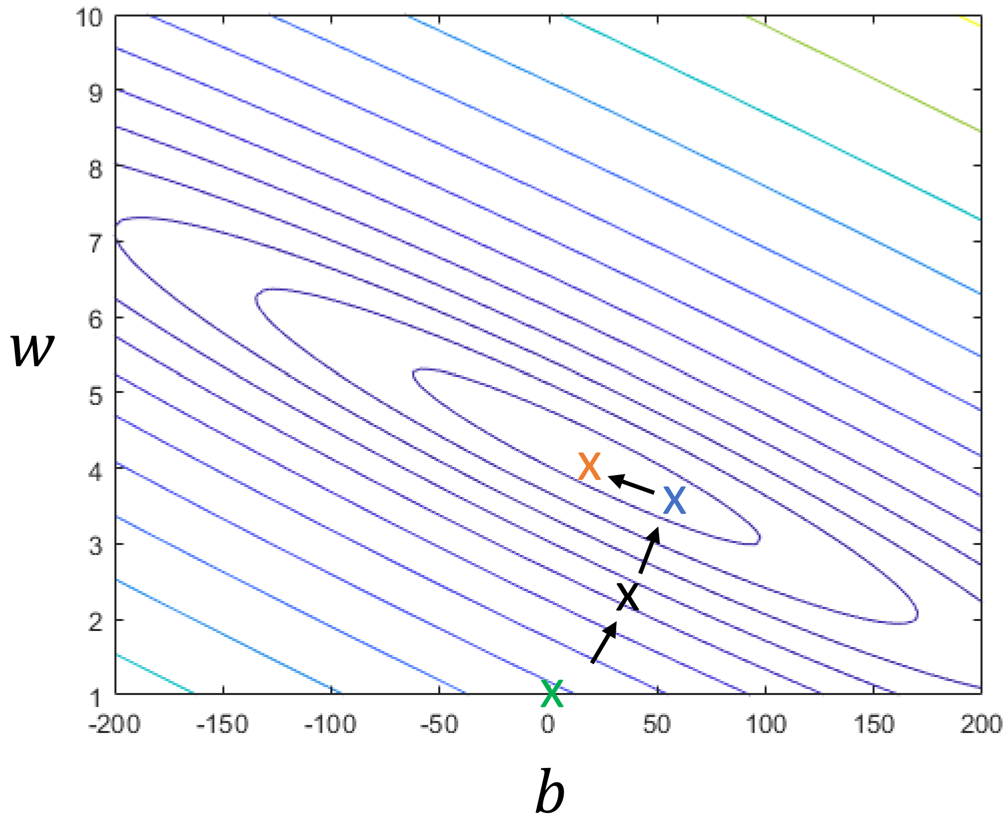
$$F(x)$$

(for fixed b, w, this is function of x)

$$J(b, w)$$

(function of the parameters b, w)





Gradient descent

Want: $\min_{b,w} J(b,w)$

1. Start with a guess:

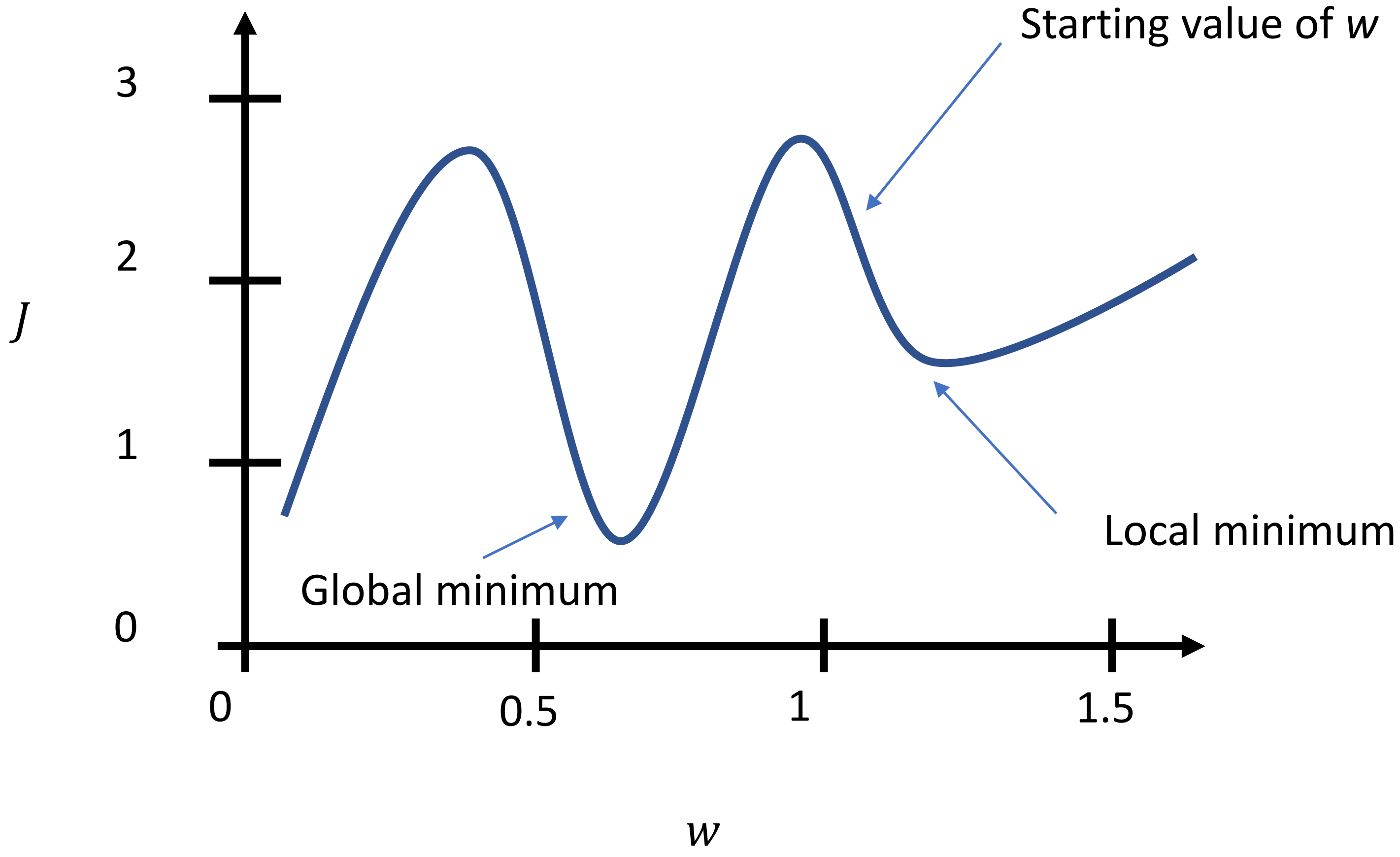
$$b = b_0, \qquad w = w_0$$

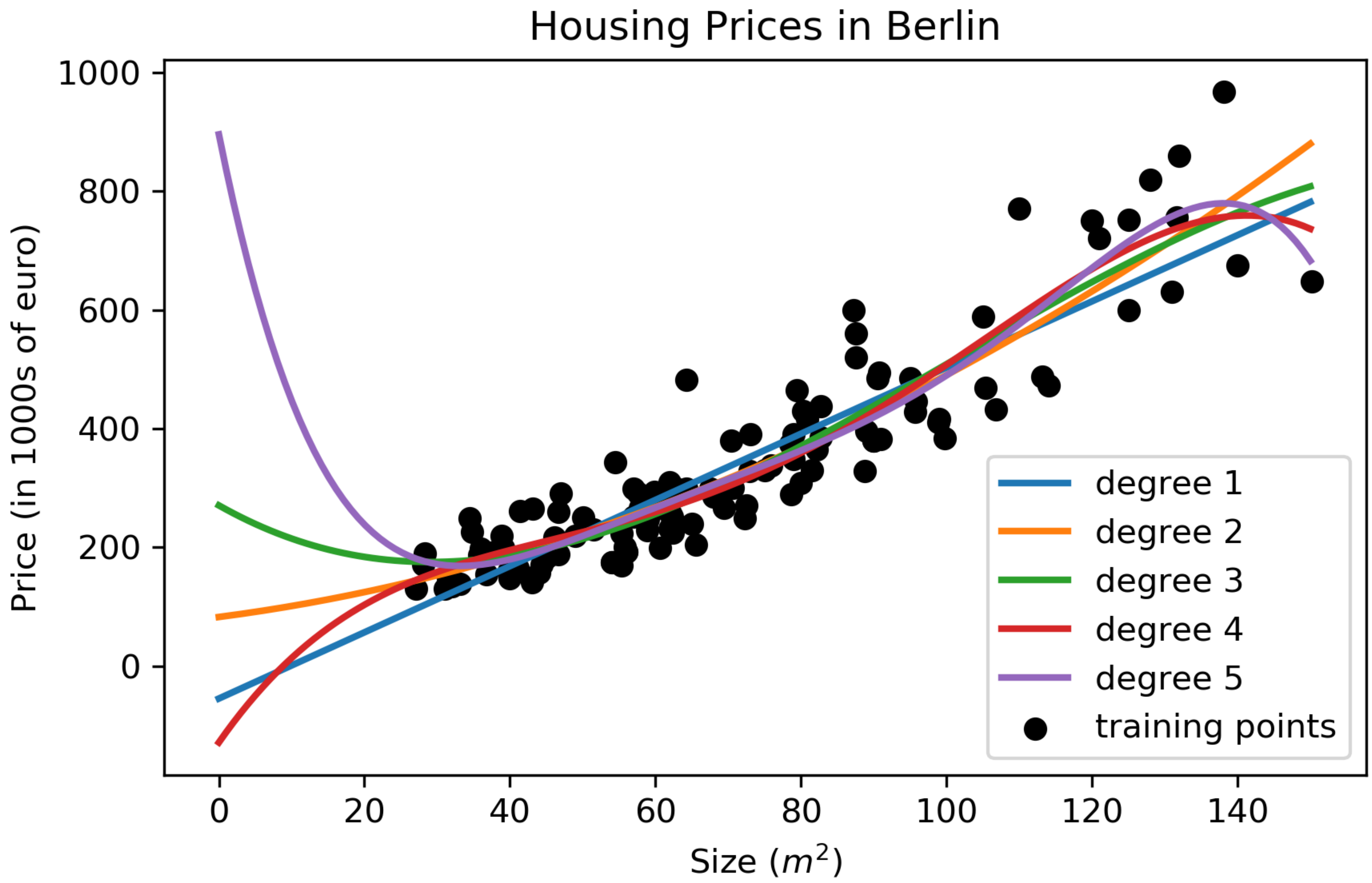
2. Calculate:

$$db = \frac{\partial J}{\partial b}, \qquad dw = \frac{\partial J}{\partial w}$$

3. Update:

$$b_{i+1} = b_i - \alpha db, \qquad w_{i+1} = w_i - \alpha dw$$


Starting value of w
3
2
J
1
0
Local minimum
Global minimum
0
0.5
1
1.5
w


Housing Prices in Berlin
1000
800
600
400
200
0
Price (in 1000s of euro)
0
20
40
60
80
100
120
140
Size (m^2)
degree 1
degree 2
degree 3
degree 4
degree 5
training points

Regression with regularization

- Ridge

  Linear least squares with L_2 regularization:

$$J = \frac{1}{2N} |y - Xw|_2^2 + \alpha \sum_i w_i^2$$

- Lasso

  Linear least squares with L_1 regularization:

$$J = \frac{1}{2N} |y - Xw|_2^2 + \alpha \sum_i |w_i|$$

- ElasticNet

  Linear regression with combined L_1 and L_2 regularization:

$$J = \frac{1}{2N} |y - Xw|_2^2 + \alpha \gamma \sum_i |w_i| + \alpha (1 - \gamma) \frac{1}{2} \sum_i w_i^2$$

Training / Test / CV

60% - training	20% - CV	20% - test

- **Training set** is a subset of data that is used to train the model
- **Test set** is a small subset of data that are <u>not</u> used to train the model
- **Cross validation dataset** is created in addition to the training and test sets to select model parameters (“hyperparameters”)

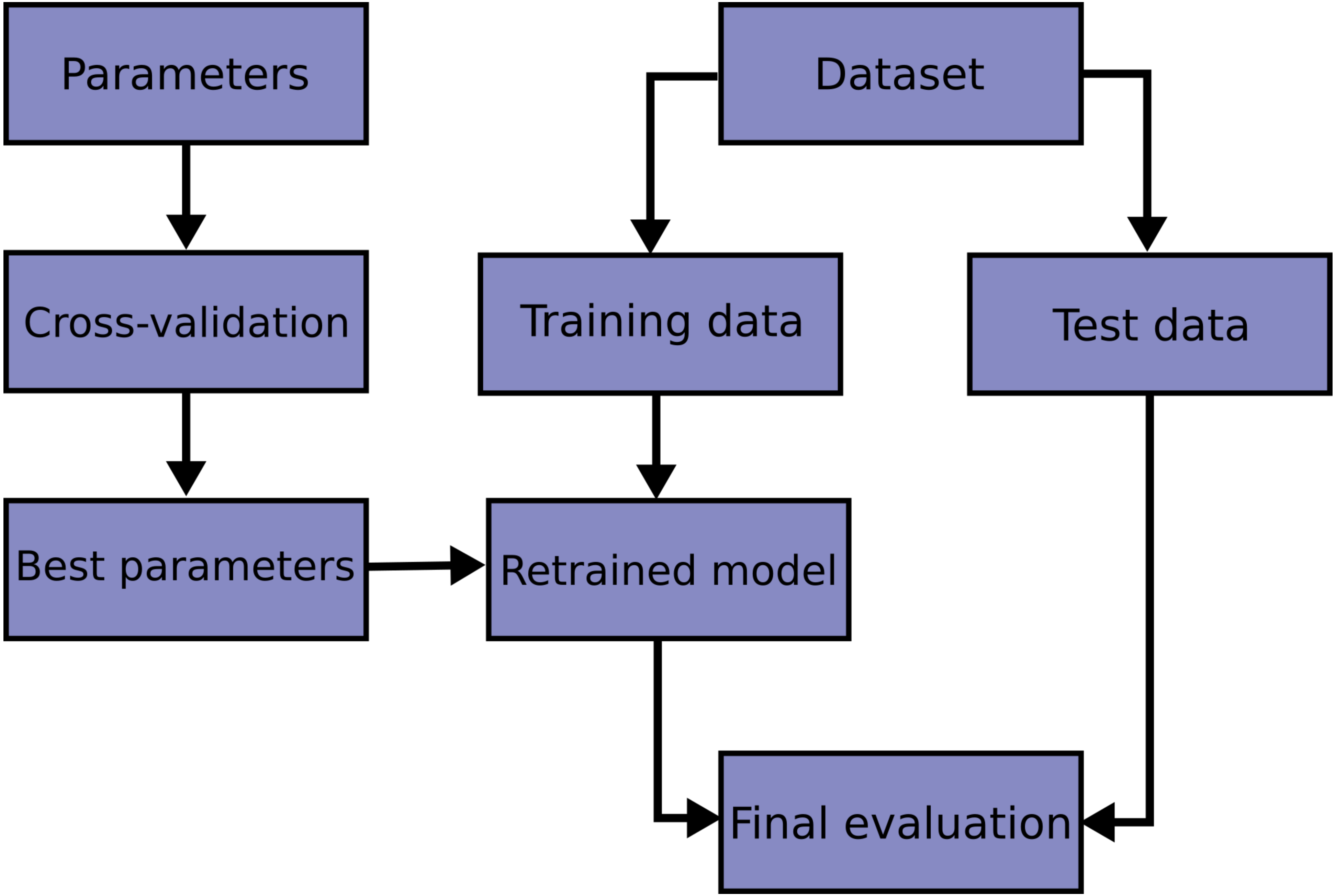


Figure source: https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation

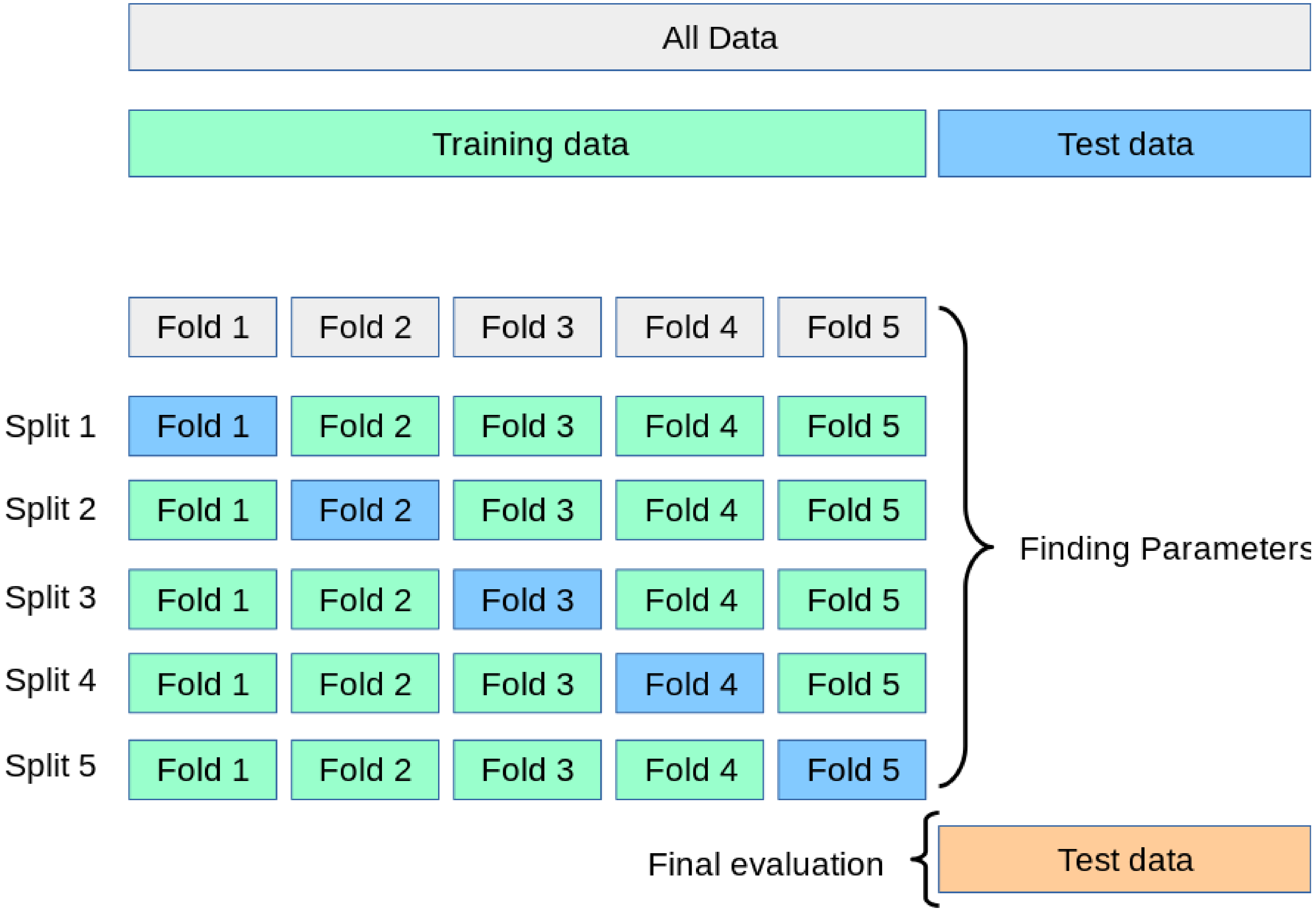


Figure source: https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation

Hyperparameters

Example: polynomial regression with combined L_1 and L_2 regularization:

$$J = \frac{1}{2N} |y - Xw|_2^2 + \alpha \gamma \sum_i |w_i| + \alpha (1 - \gamma) \frac{1}{2} \sum_i w_i^2$$

Hyperparameters:

- α
- γ
- degree