EVML3

REGRESSION
HANDS-ON

JEROEN VEEN

HAN_UNIVERSITY
OF APPLIED SCIENCES

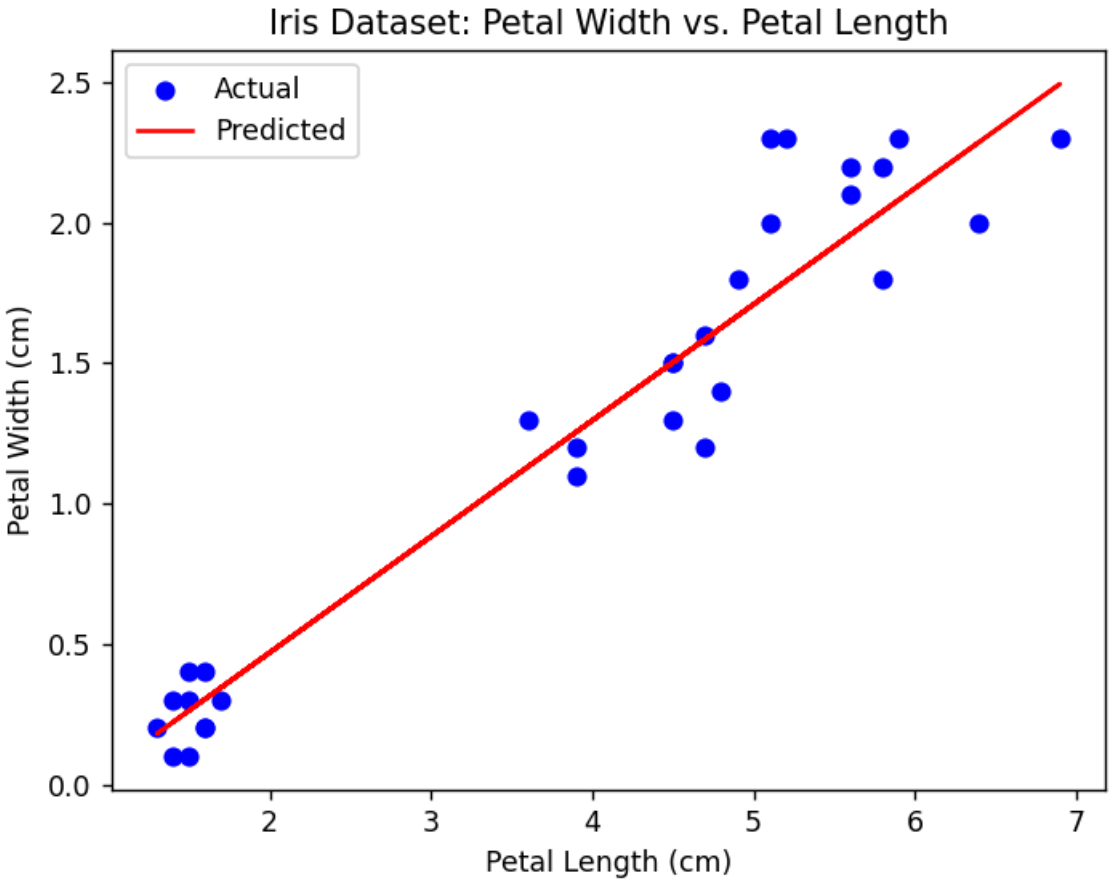
CONTENTS

- Linear regression exercise
- Polynomial regression exercise

SCIKIT LEARN INTERFACE DESIGN

- Estimators
  - estimation performed by the fit() method
  - dataset as a parameter (or two for supervised learning)
  - Any other parameter is considered a hyperparameter
- Transformers
  - performed by the transform() method
  - fit_transform() is equivalent to calling fit() and then transform() (but sometimes fit_transform() is optimized and runs much faster
- Predictors
  - prediction method performed by predict() method
  - quality of the predictions measured by score() method

LINEAR REGRESSION EXAMPLE

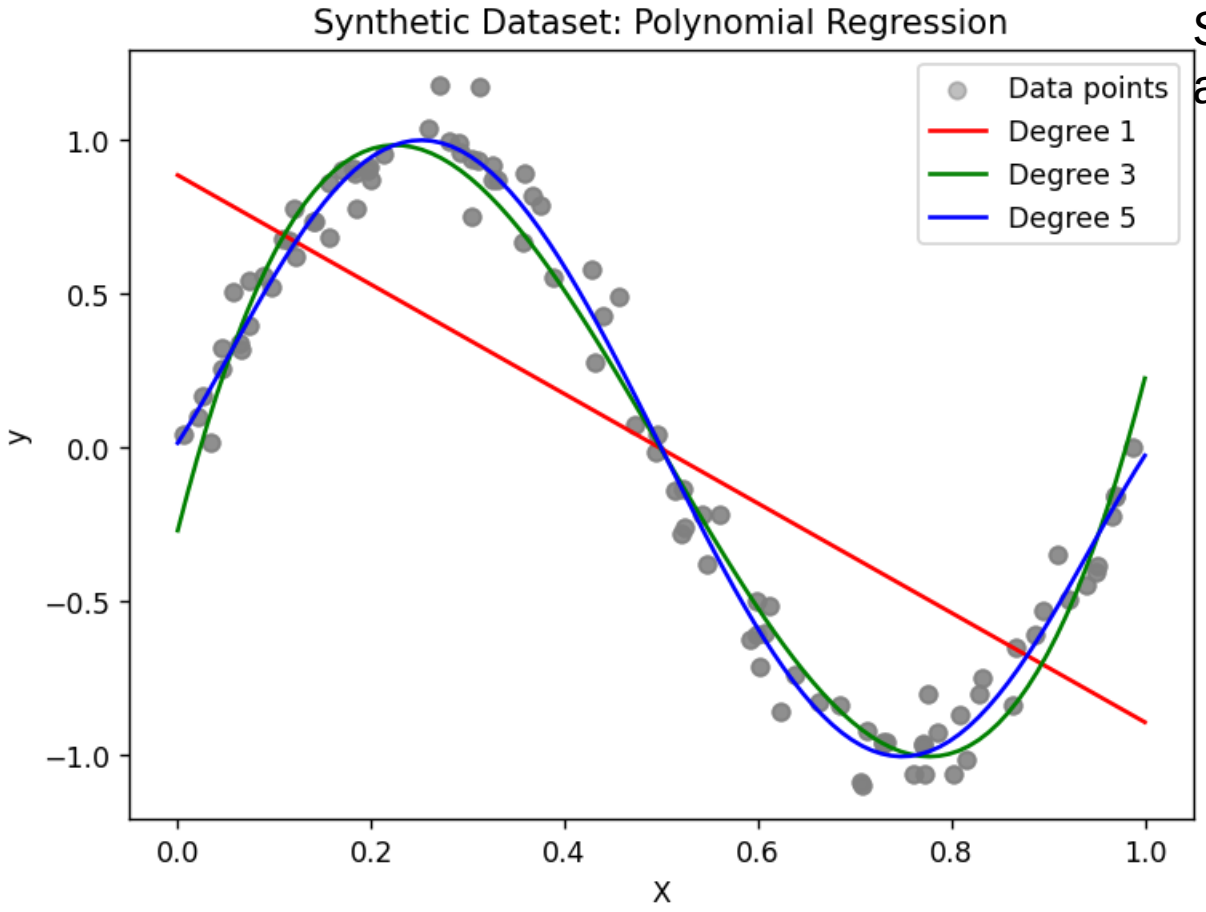


See Regression_01.py
and Regression_04.py

Mean squared error: 0.0405
R-squared score: 0.9268
Model coefficient: 0.4159
Model intercept: -0.3630

Petal length is a good predictor of petal width, explaining about 93% of the variation

POLYNOMIAL REGRESSION EXAMPLE



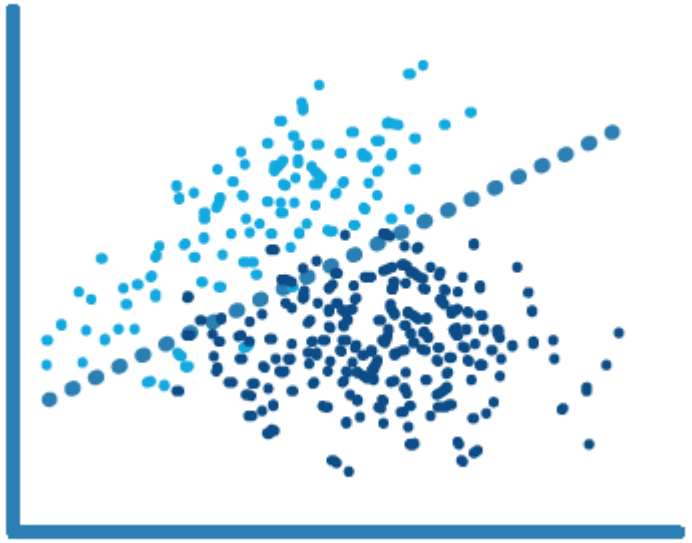
See Regression_02.py and Regression_05.py

POLYNOMIAL REGRESSION

- Exercise 1: Can you improve Regression_01.py or Regression_04.py by polynomial regression?
- Try out generating your own regression example: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_regression.html
- See also Géron, page 129 (114 in new ed.)

LOGISTIC REGRESSION

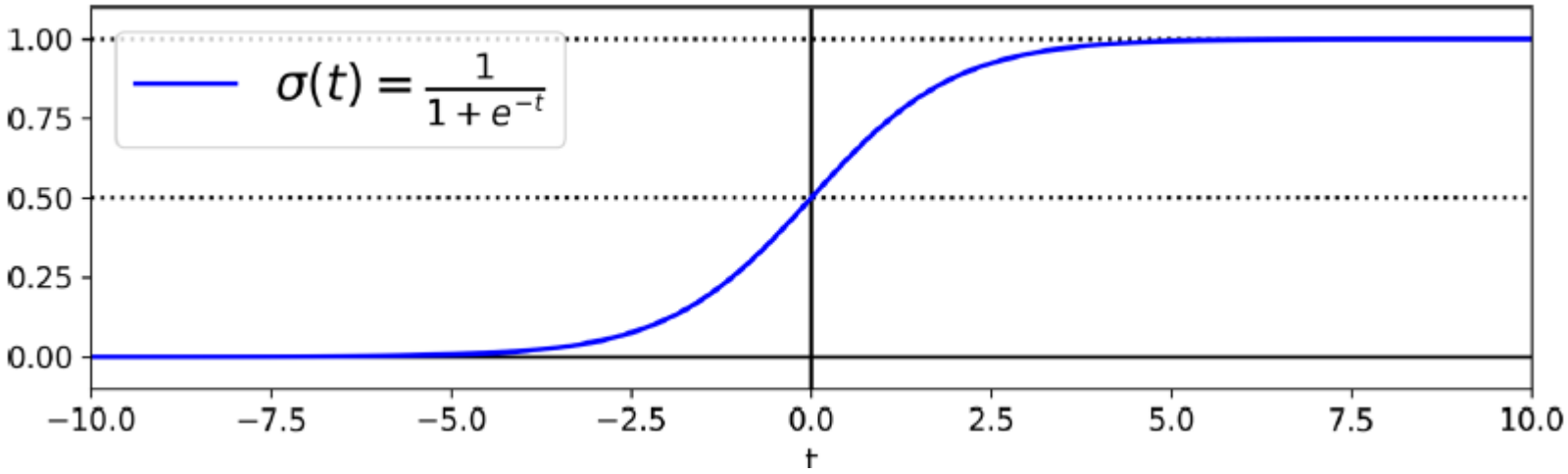
- Estimate the probability that an instance belongs to a particular class
- Binary classifier
- Baseline for evaluating more complex classification methods



Source: Mathworks, Applying Supervised Learning

ESTIMATING PROBABILITY

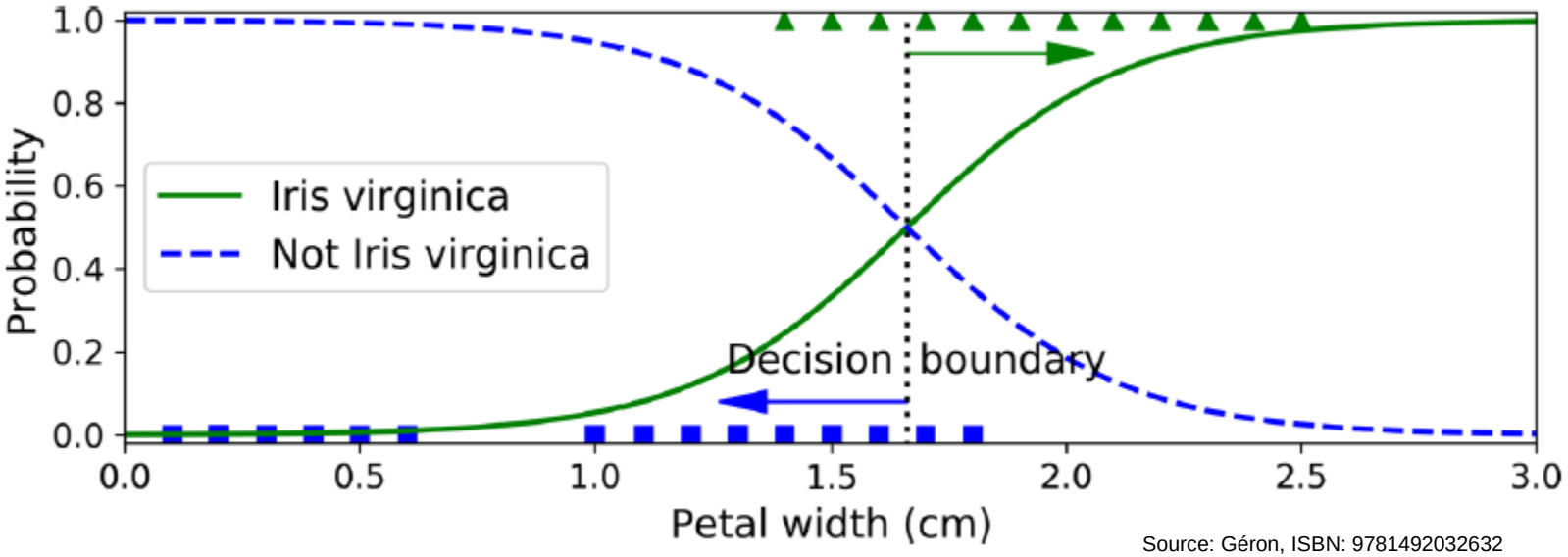
- Logistic functions maps prediction result to probability
- Sigmoid function



https://developers.google.com/machine-learning/crash-course/logistic-regression/calculating-a-probability

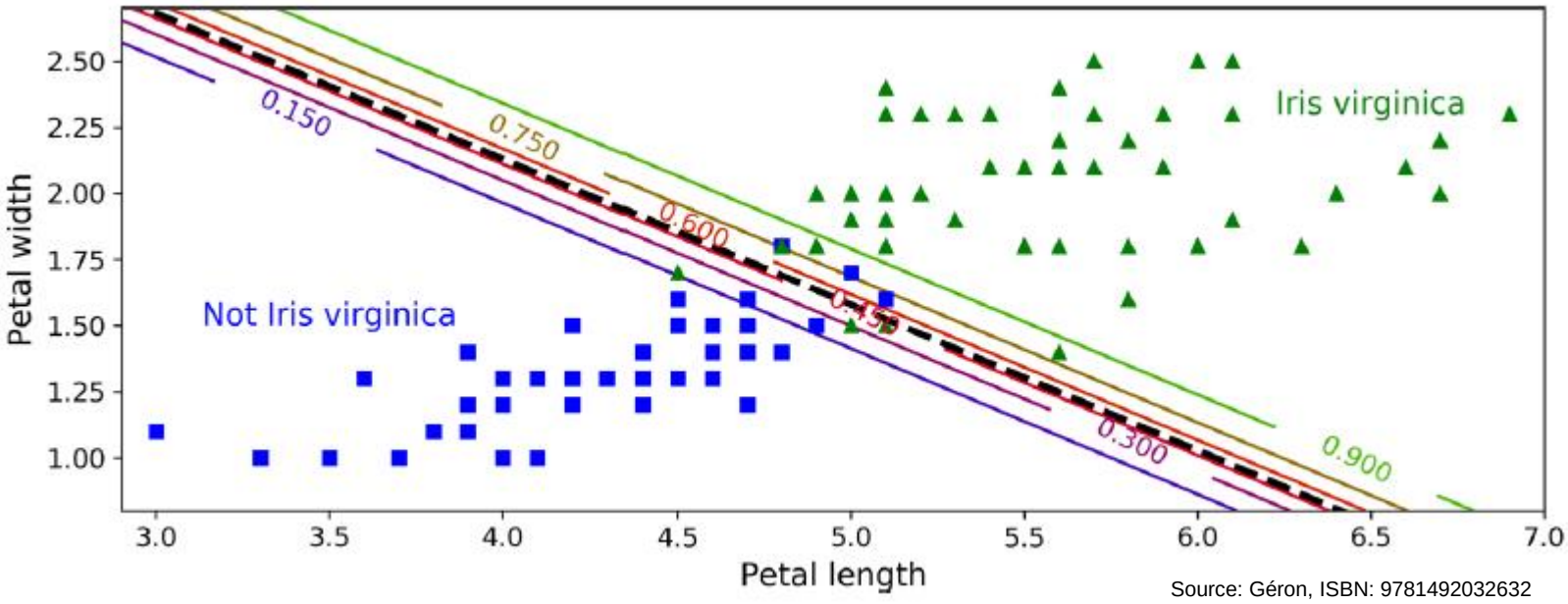
DECISION BOUNDARY

- Aka classification threshold
- Both probabilities are equal to 50% ?



Source: Géron, ISBN: 9781492032632

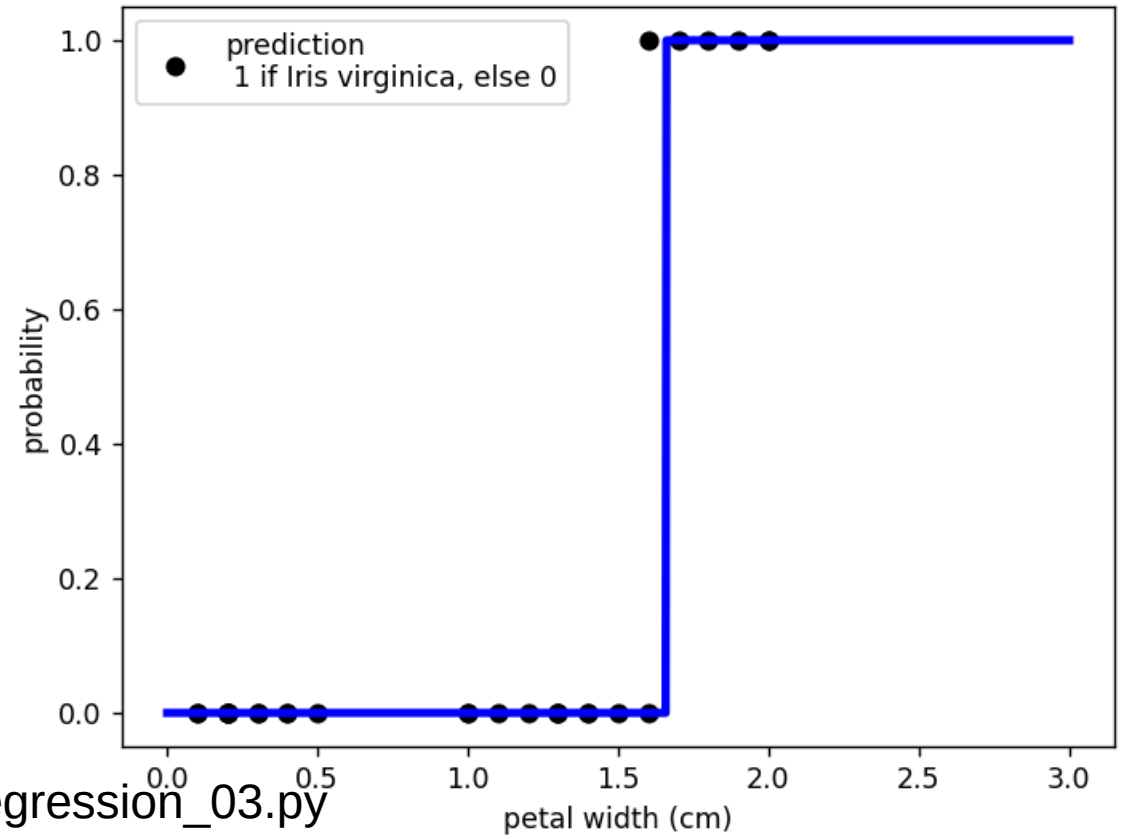
LINEAR DECISION BOUNDARY



Source: Géron, ISBN: 9781492032632

- Logistic Regression models can be regularized

LOGISTIC REGRESSION EXAMPLE



See Regression_03.py

LEANING CURVES IN REGRESSION

- Exercise 02: Plot the learning curves for polynomial regression and experiment with various degrees
- Can you interpret the curves?
- Build on Regression_01.py
  and see Géron, page 130-134