Regression

Course: Analytics, Machine Learning, and the Digital Economy

- **Lecturer Radjabova Dilnora**

Learning Algorithm Trade-off

- There are always two aspects to consider when designing a learning algorithm:
  - Try to fit the data well
  - Be as robust as possible

Learning Algorithm Trade-off

- The predictor that you have generated using your training data must also work well on new data.

Learning Algorithm Trade-off (cont.)

- When we create predictors, usually the simpler the predictor is, the more robust it tends to be in the sense of begin able to be estimated reliably.

Learning Algorithm Trade-off (cont.)

- On the other hand, the simple models do not fit the training data aggressively.

Learning Algorithm Trade-off (cont.)

- Training Error vs. Testing Error:
  - Training error → reflects whether the data fits well
  - Testing error → reflects whether the predictor actually works on new data

Learning Algorithm Trade-off (cont.)

- Bias vs. Variance:
  - Bias → how good the predictor is, on average; tends to be smaller with more complicated models
  - Variance → tends to be higher for more complex models

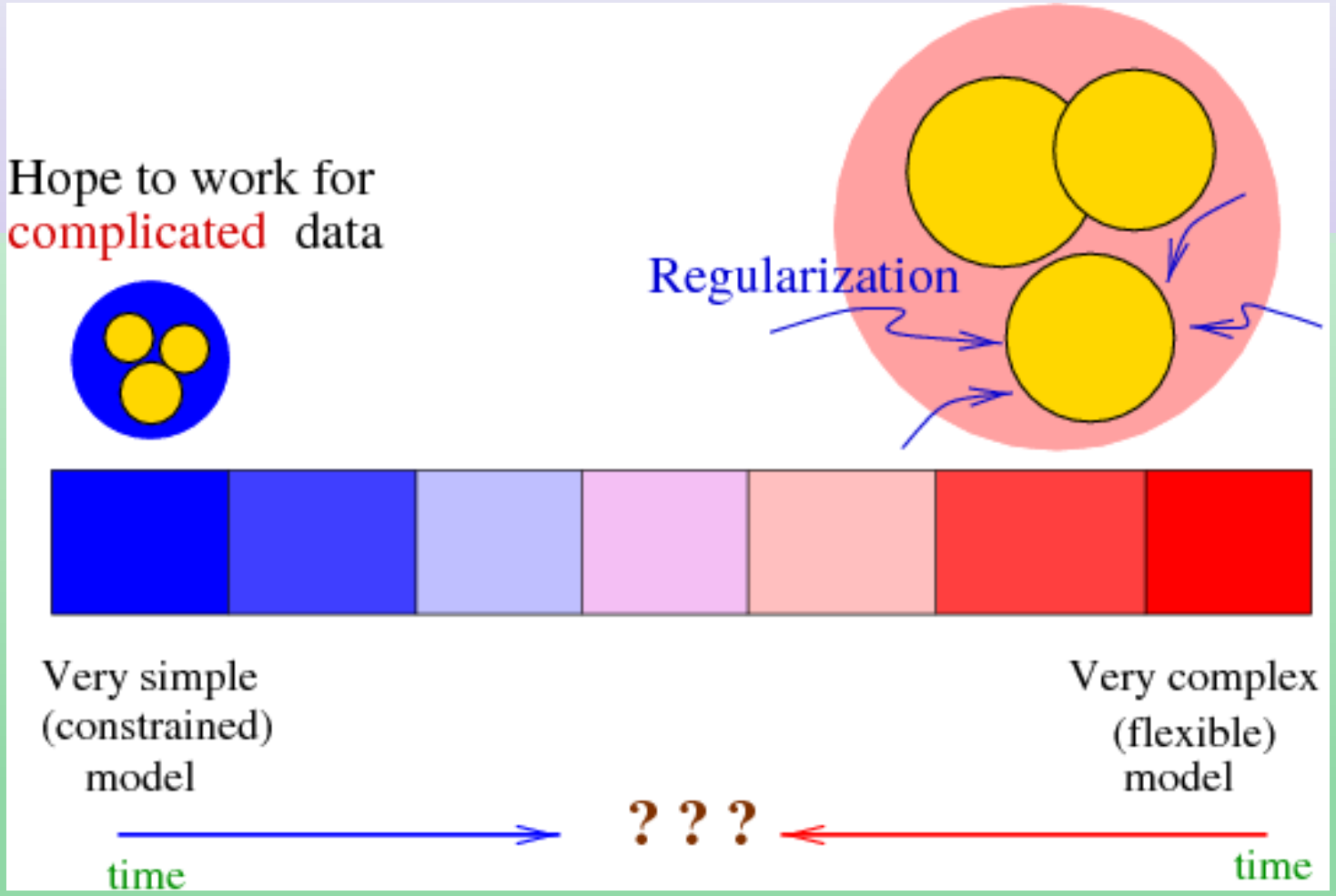
Learning Algorithm Trade-off (cont.)

- Fitting vs. Over-fitting:
  - If you try to fit the data too aggressively, then you may over-fit the training data. This means that the predictors works very well on the training data, but is substantially worse on the unseen test data.

Learning Algorithm Trade-off (cont.)

- Empirical Risk vs. Model Complexity:
  - Empirical risk → error rate based on the training data
  - Increase model complexity = decrease empirical risk but less robust (higher variance)

Learning Spectrum



Supervised vs. Unsupervised Learning

- Supervised Learning:
  - All the predictors, $\boldsymbol{X_i}$, and the response, Y_i, are observed.
    - Many regression and classification methods

Supervised vs. Unsupervised Learning

- Unsupervised Learning:
  - Here, only the $\boldsymbol{X_i}$'s are observed (not Y_i's).
  - We need to use the $\boldsymbol{X_i}$'s to guess what Y would have been, and then build a model form there.
    - Clustering and principal components analysis

Terminology

- **Notation**
  - Input X: *feature, predictor,* or *independent variable*
  - Output Y: *response, dependent variable*

Terminology

- **Categorization**
  - Supervised learning vs. unsupervised learning
    - *Key question*: Is *Y* available in the training data?
  - Regression vs. Classification
    - *Key question*: Is *Y* quantitative or qualitative?

Terminology (cont.)

- **Quantitative:**
  - Measurements or counts, recorded as numerical values (e.g. height, temperature, etc.)

Terminology (cont.)

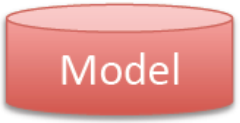
- **Qualitative:** group or categories
  - Ordinal: possesses a natural ordering (e.g. shirt sizes)
  - Nominal: just name the categories (e.g. marital status, gender, etc.)

Terminology (cont.)

Feature X | Label Y

Training Samples

x_1	x_2	x_3	…	x_p	Y
3	5	2	…	1	A
4	2	3	…	2	B
…	…	…	…	…	…
4	2	3	…	3	A



Feature X | Label Y (unknown)

Testing

x_1	x_2	x_3	…	x_p	Y
5	5	2	…	1	?
2	2	1	…	2	?
…	…	…	…	…	…
1	3	2	…	4	?

Supervised Learning

Training

This is a dog

This is a dog

This is a cat

What is this?

This is a dog!

Testing

Classification: Categorical output

Regression: Continuous output

Supervised Learning: Regression vs. Classification

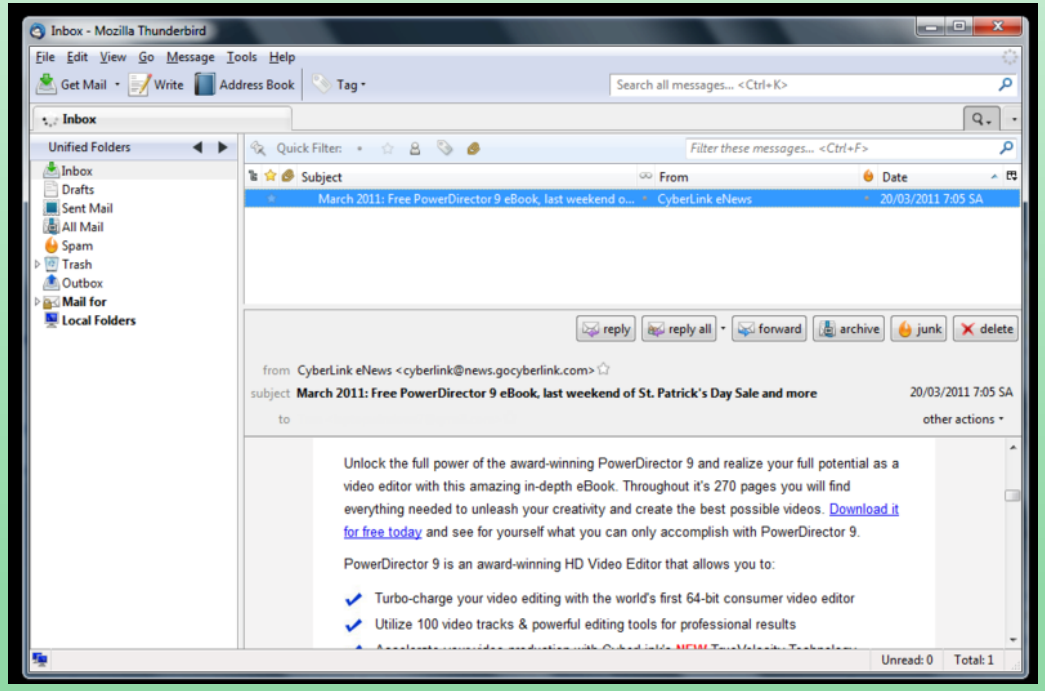
- Regression
  - Covers situations where *Y* is continuous (quantitative)
  - E.g. predicting the value of the Dow in 6 months, predicting the value of a given house based on various inputs, etc.

Supervised Learning: Regression vs. Classification

- Classification
  - Covers situations where Y is categorical (qualitative)
  - E.g. Will the Dow be up or down in 6 months? Is this email spam or not?

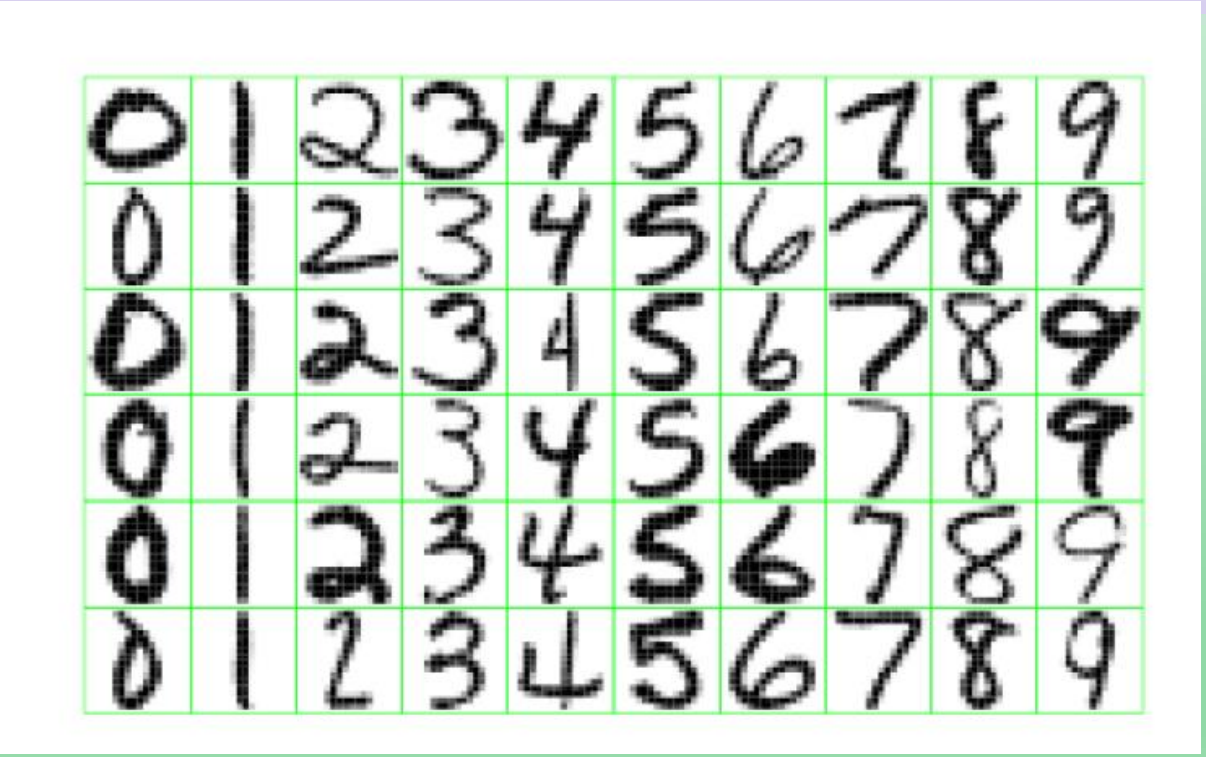
Supervised Learning: Examples

- Email Spam:
  - predict whether an email is a junk email (i.e. spam)



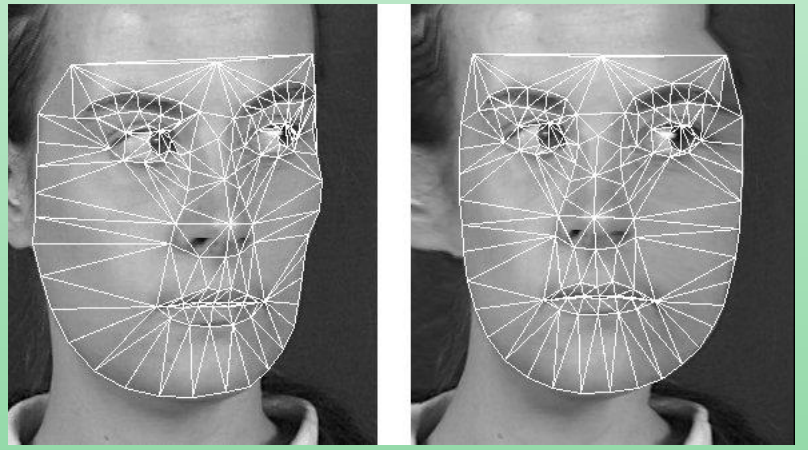
Supervised Learning: Examples

- Handwritten Digit Recognition:
  - Identify single digits 0~9 based on images



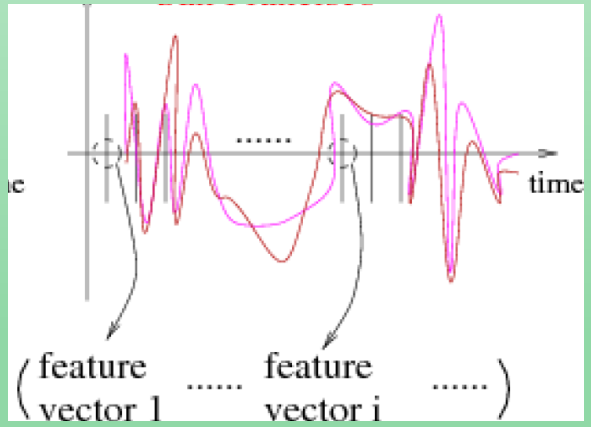
Supervised Learning: Examples

- Face Detection/Recognition:
  - Identify human faces

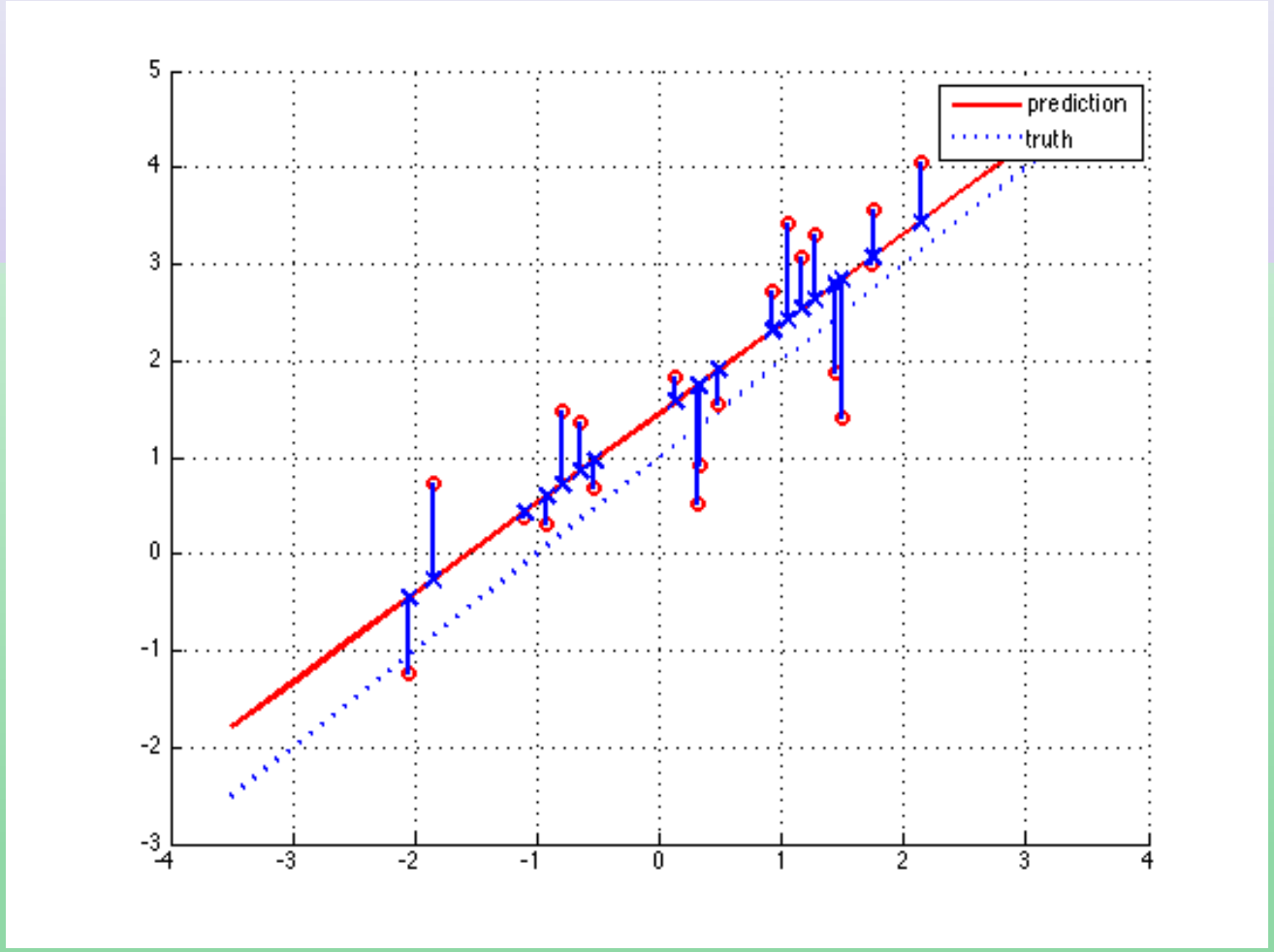


Supervised Learning: Examples

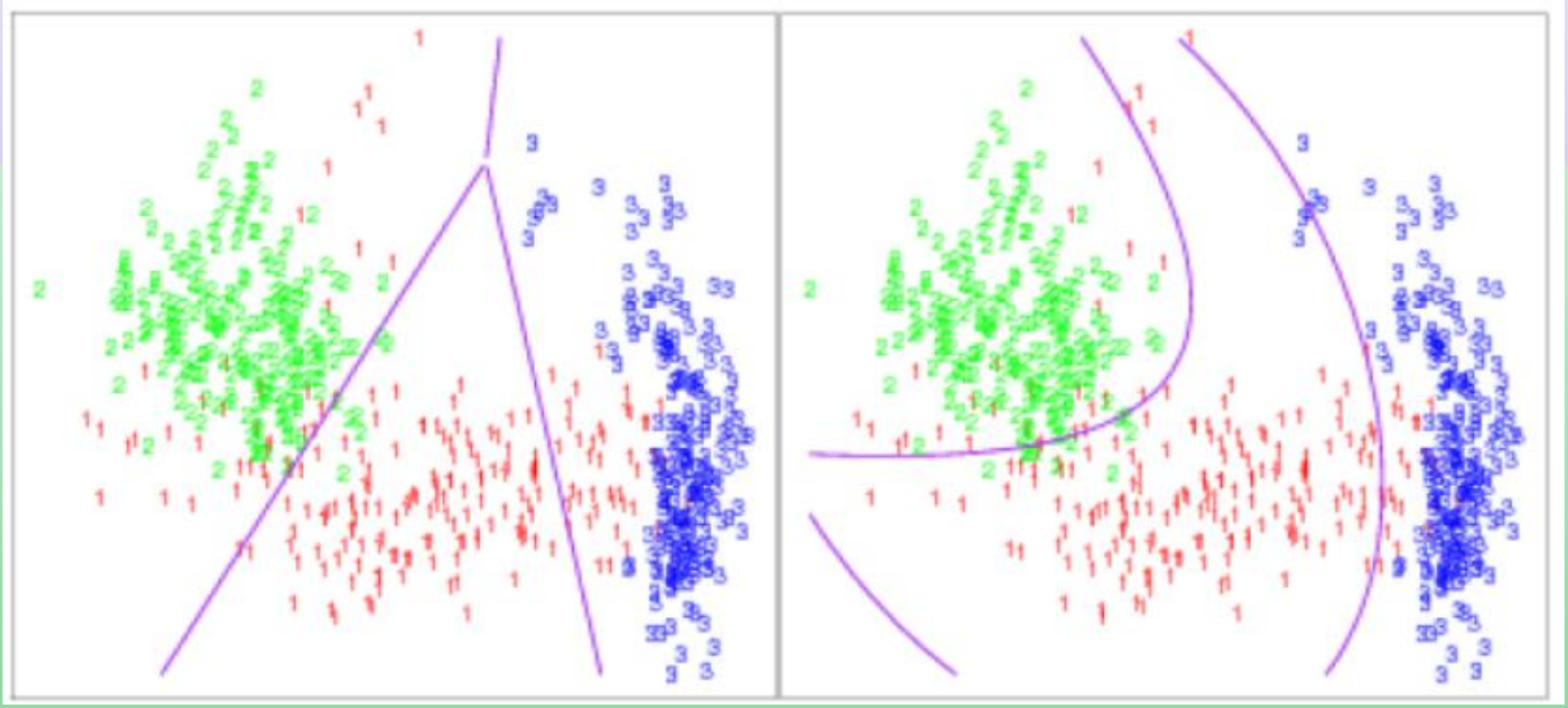
- Speech Recognition:
  - Identify words spoken according to speech signals
    - Automatic voice recognition systems used by airline companies, automatic stock price reporting, etc.



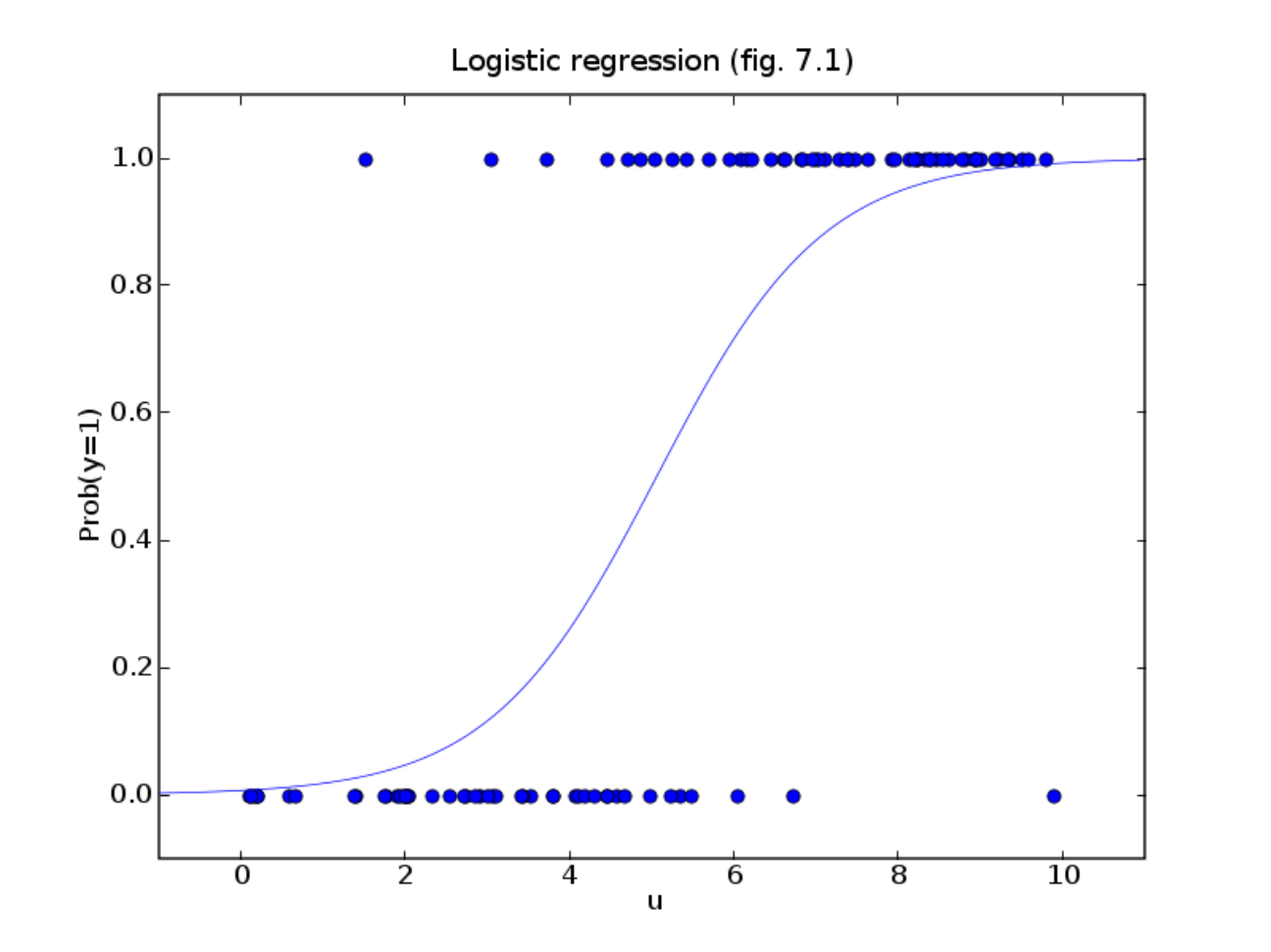
Supervised Learning: Linear Regression



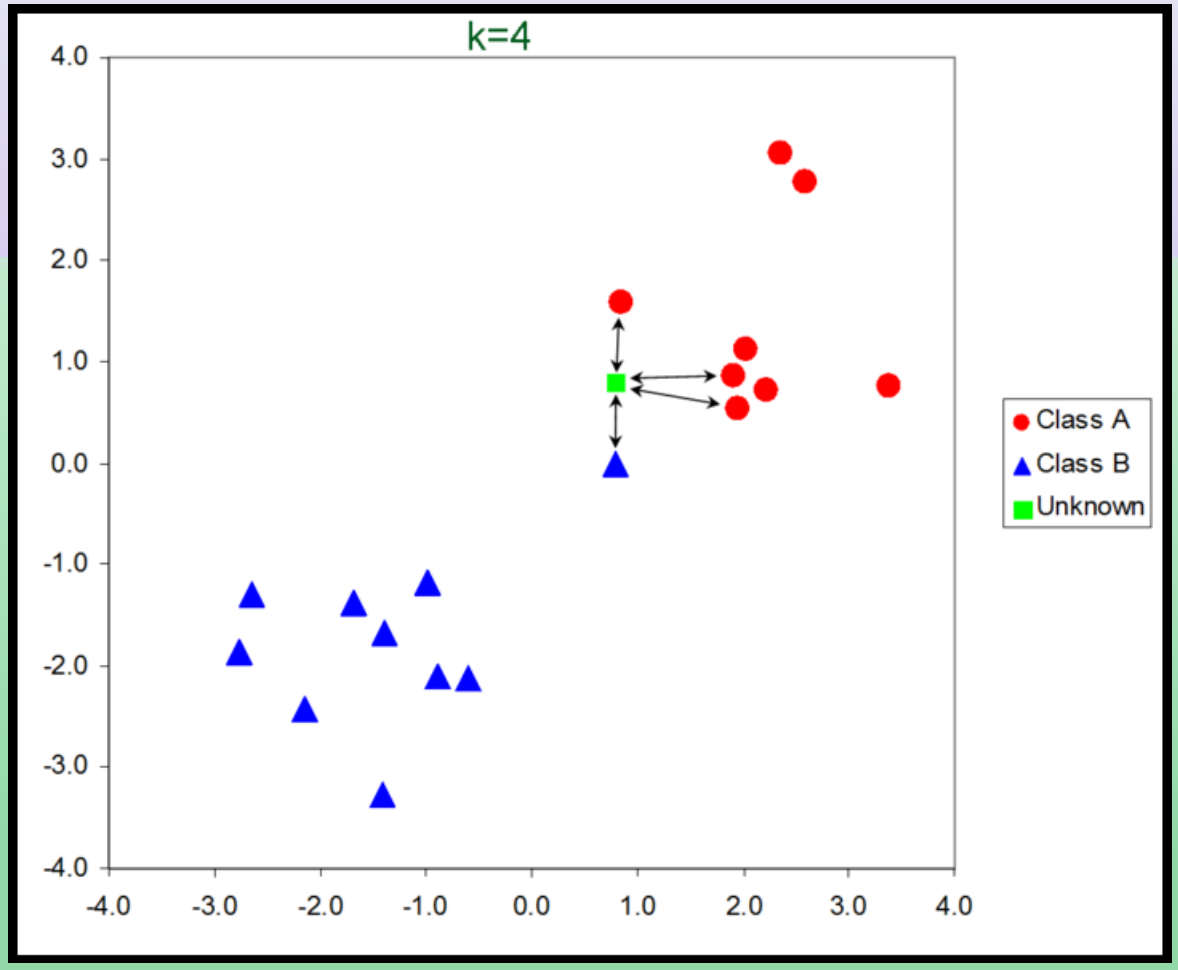
Supervised Learning: Linear/Quadratic Discriminant Analysis



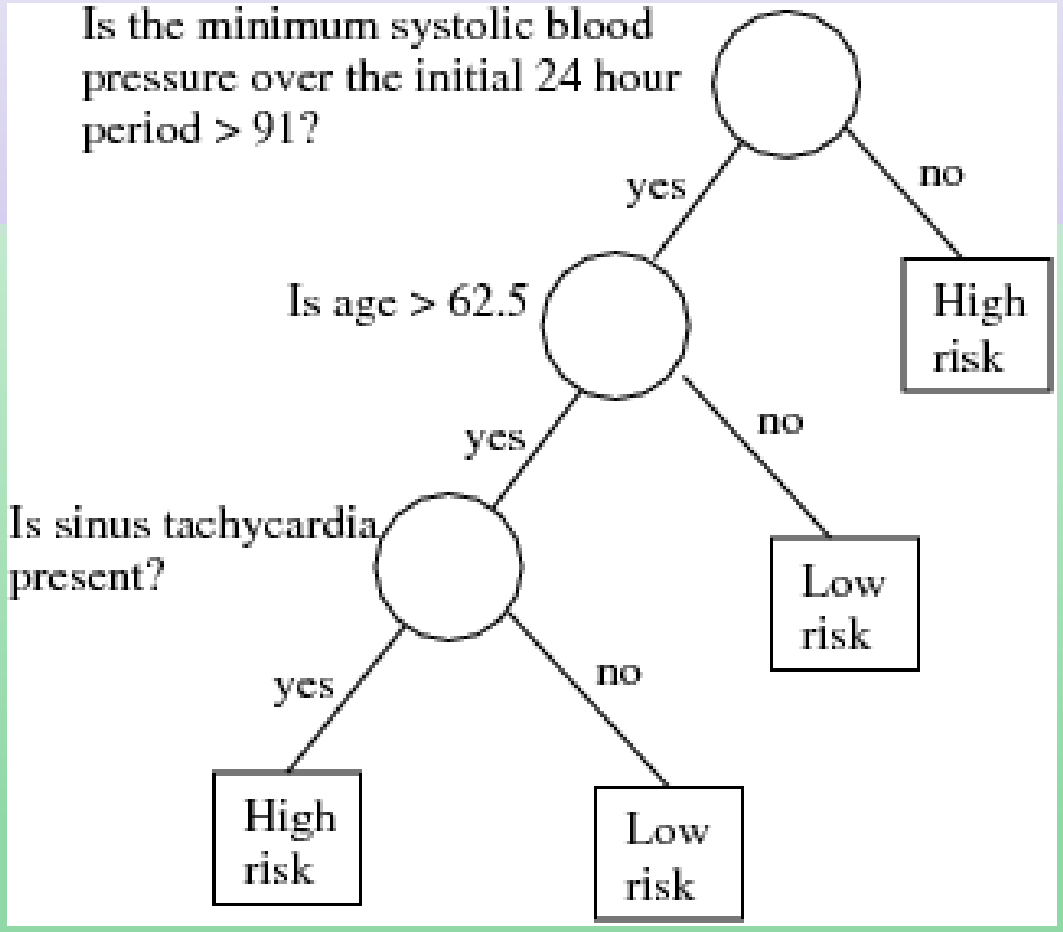
Supervised Learning: Logistic Regression



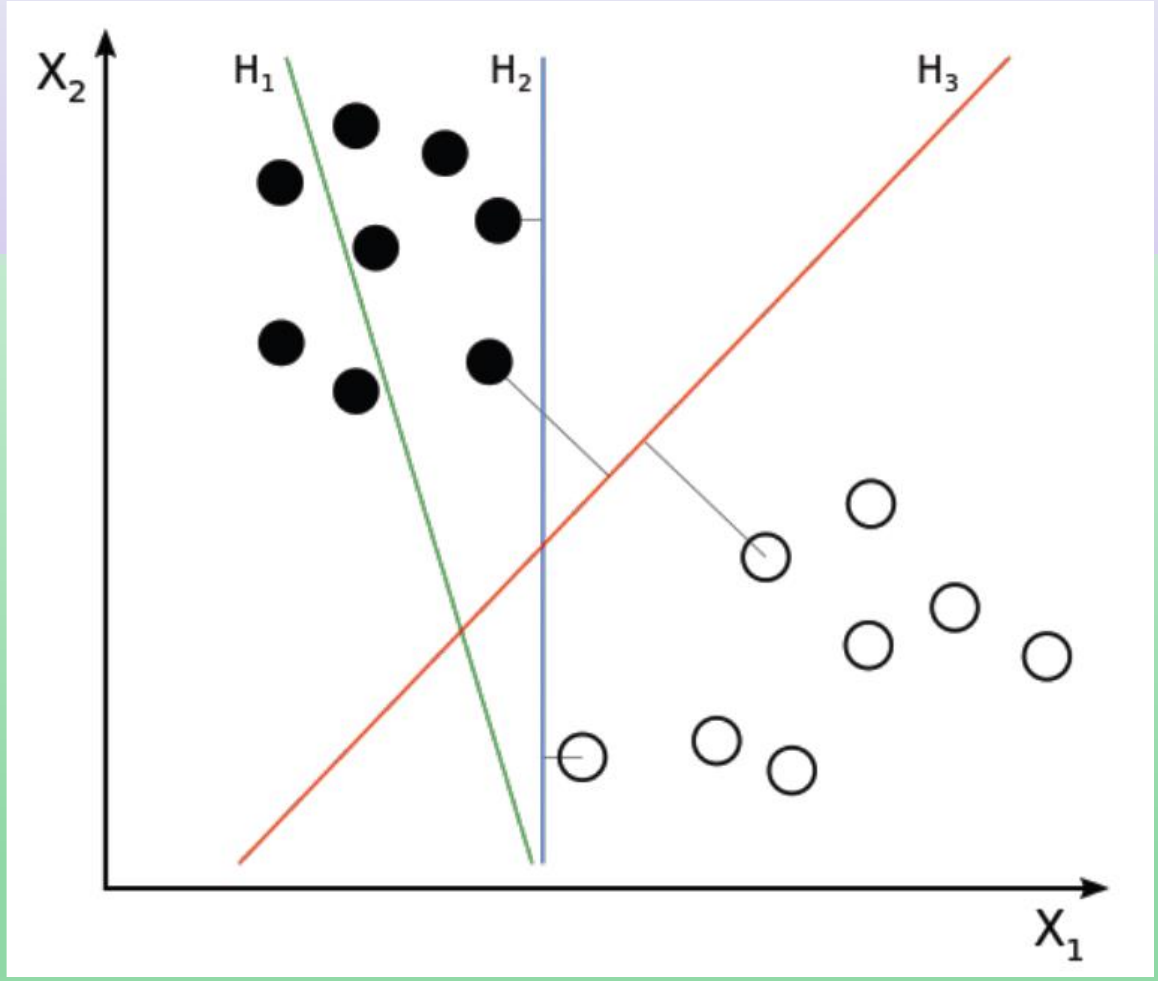
Supervised Learning: K Nearest Neighbors



Supervised Learning: Decision Trees / CART



Supervised Learning: Support Vector Machines



Reference

- 1. Shalev-Shwartz and Ben-David. Understanding Machine Learning: From Theory to Algorithms (Cambridge University Press, 2014)
- 2. Daum´e. A Course in Machine Learning.
- 3. The Art of Statistics: How to Learn from Data by David Shpigelter
- 4. Learning From Data – January 1, 2012 by Yaser S. Abu-Mostafa (Author), Malik Magdon-Ismail (Author), Hsuan-Tien Lin (Author)
- 5. Statistics: The Art and Science of Learning from Data by Alan Agresti
- 6. Learning From Data: An Introduction To Statistical Reasoning by M.Glenber.
- 7. Statistics: Learning from Data (with JMP Printed Access Card) by Rocky Pek
- 8. The Elements of Statistical Learning by Gerim Garold
- 9. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition
- by Aurélien Géron (Author)