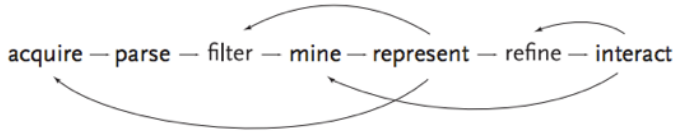
GENERAL ASSEMBLY

INTRO TO DATA SCIENCE
REVIEW

KEY OBJECTIVES

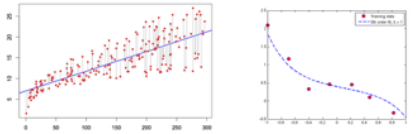
- CONSOLIDATE KNOWLEDGE OF THE THEORY STUDIED SO FAR
- ACTIVATE AND DISCOVER LINKS BETWEEN THEORY AND LAB
- PRACTICE THE TECHNIQUES LEARNED
- ACQUIRE CONFIDENCE AND INDEPENDENCE IN FACING NEW

REVIEW

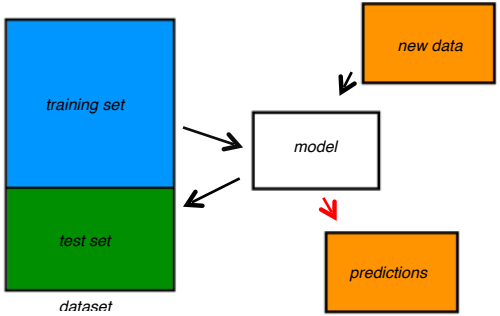


	Continuous	Categorical
Supervised	*regression* ✓	*classification* ✓
Unsupervised	*dimension reduction*	*clustering*

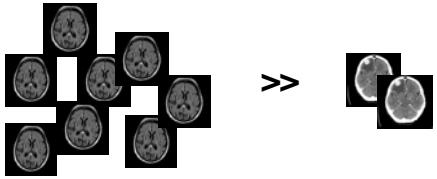
Regression & Regularization



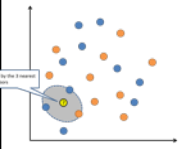
Train - Test Split



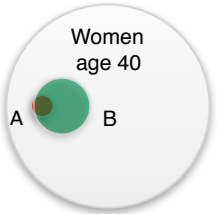
Imbalanced classes



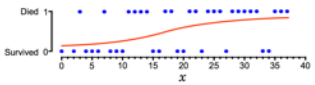
5 Classification methods



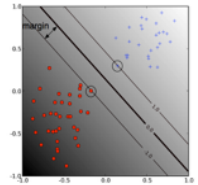
KNN



Naïve Bayes



Logistic Regression



SVMs

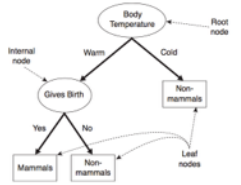
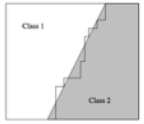


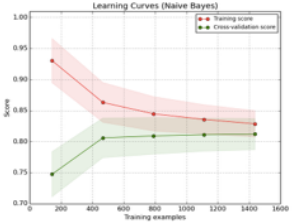
Figure 4.4. A decision tree for the mammal classification problem.

Decision Trees

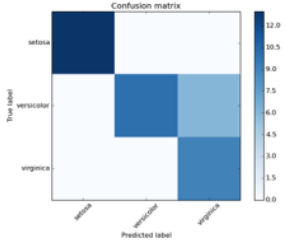
Ensembles



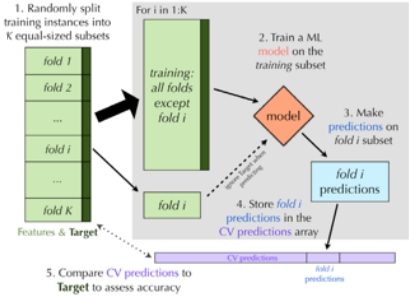
Learning curves

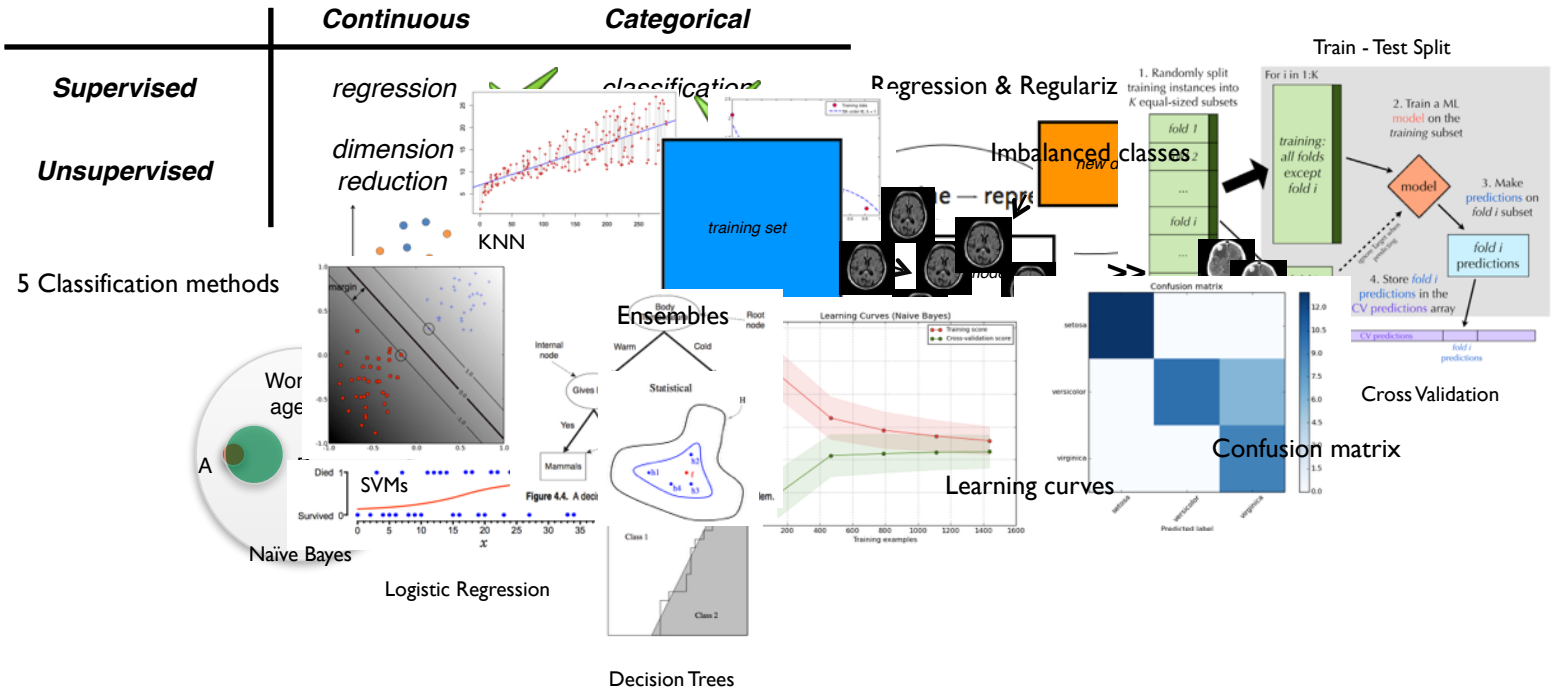


Confusion matrix



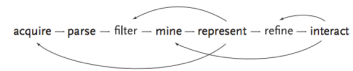
Cross Validation

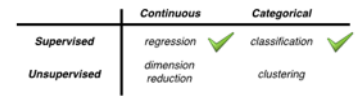


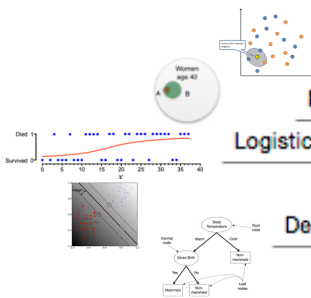

Continuous
Categorical
Supervised
Unsupervised
regression
dimension reduction
classification
Regression & Regulariz
Imbalanced classes
Train - Test Split
KNN
training set
5 Classification methods
Ensembles
Cross Validation
Confusion matrix
Learning curves
SVMs
Naïve Bayes
Logistic Regression
Decision Trees

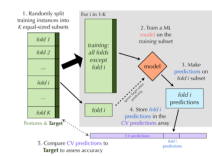
Theory

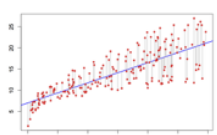
- Ensembles
- Data science process
- Imbalanced Classes
- Machine Learning Problems
- Classification
  - KNN
  - Naïve Bayes
  - Logistic Regression
  - SVMs
  - Decision Trees
- Train-Test split
- Cross Validation
- Regression
- Confusion Matrix

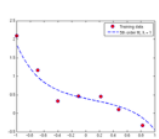


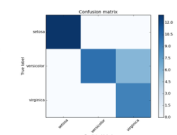


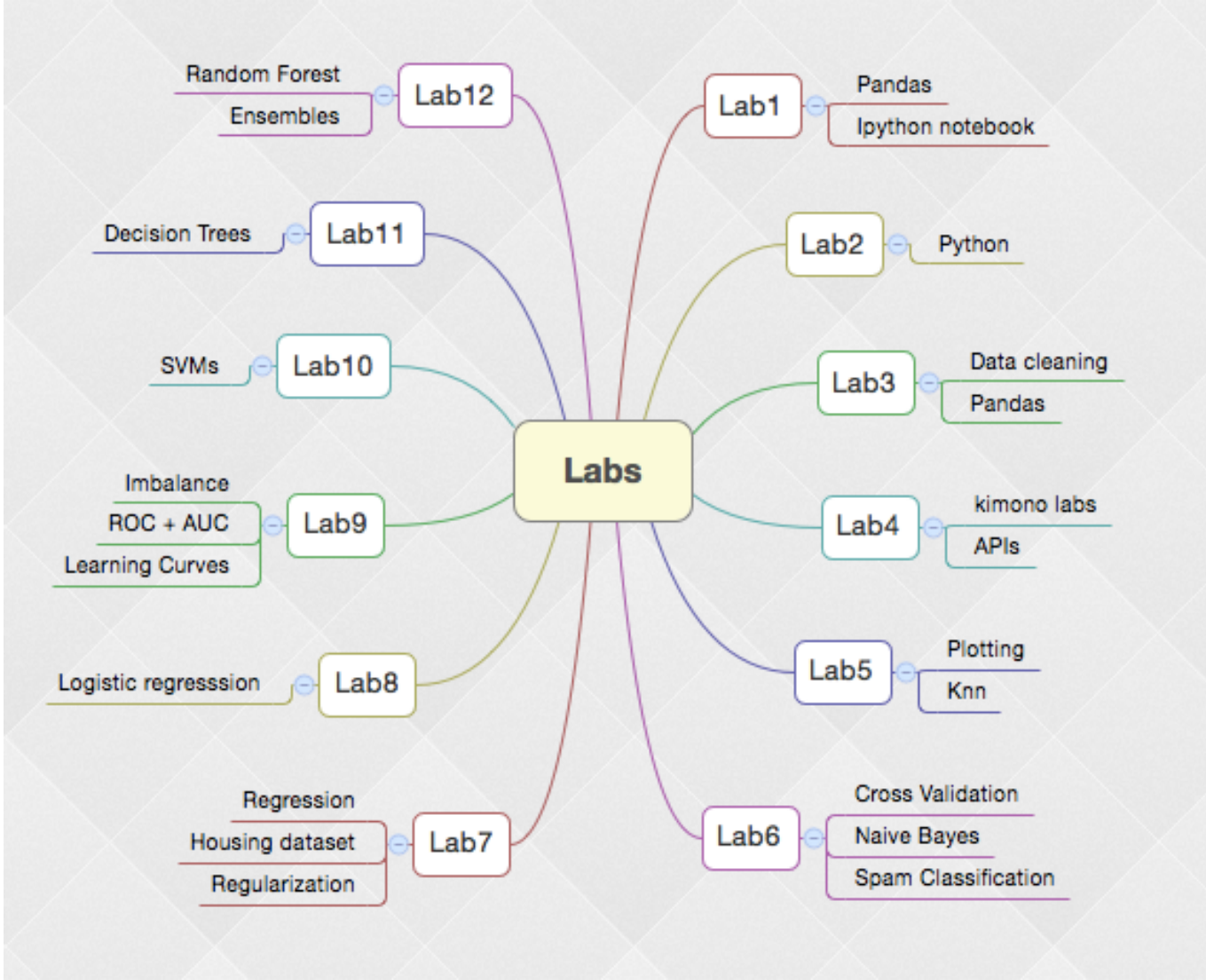




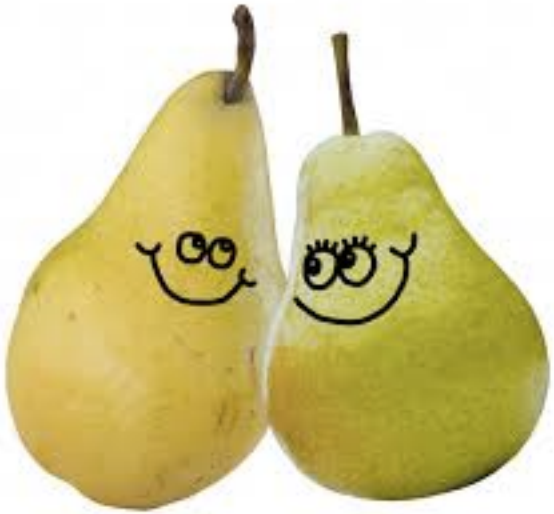







Labs
Lab1
Pandas
Ipython notebook
Lab2
Python
Lab3
Data cleaning
Pandas
Lab4
kimono labs
APIs
Lab5
Plotting
Knn
Lab6
Cross Validation
Naive Bayes
Spam Classification
Lab7
Regression
Housing dataset
Regularization
Lab8
Logistic regresssion
Lab9
Imbalance
ROC + AUC
Learning Curves
Lab10
SVMs
Lab11
Decision Trees
Lab12
Random Forest
Ensembles

IN PAIRS



- Choose a lab for your pair, communicate to class
- Open the lab code and the lecture presentation
- Go through the code and discuss
  - are there any parts that are cryptic/not clear?
- Find links between lab code and theory
- Highlight missing links
  - are there any aspects of the theory you would have wanted to cover in lab or vice versa?

PAIR PROGRAMMING

PAIR PROGRAMMING

- Choose Driver and Navigator
- Open Pair programming template
- Driver writes code, Navigator thinks about the problem at hand
- After 10 minutes switch roles

ELEVATOR PITCHES