

Hands on Weka: Part II

Petra Kralj Novak
 4.12.2012

Exercise 2: CAR dataset

- 1728 examples
- 6 attributes
 - 6 nominal
 - 0 numeric
- Nominal target variable
 - 4 classes: unacc, acc, good, v-good
 - Distribution of classes
 - unacc (70%), acc (22%), good (4%), v-good (4%)
- No missing values

Preparing the data for WEKA - 1

Data in a spreadsheet
 (e.g. MS Excel)

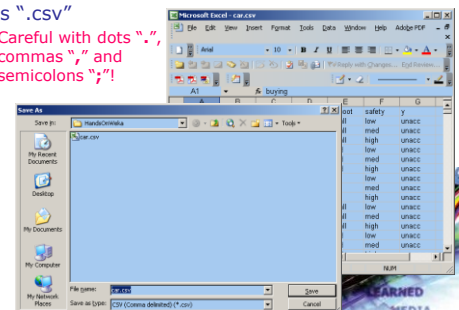
- Rows are examples
- Columns are attributes
- The last column is the target variable

	A	B	C	D	E	F	G
1	buying	unacc	doors	persons	lugboot	safety	?
2	unhigh	unhigh	2	2	small	low	unacc
3	unhigh	unhigh	2	2	small	med	unacc
4	unhigh	unhigh	2	2	small	high	unacc
5	unhigh	unhigh	2	2	med	low	unacc
6	unhigh	unhigh	2	2	med	med	unacc
7	unhigh	unhigh	2	2	med	high	unacc
8	unhigh	unhigh	2	2	big	low	unacc
9	unhigh	unhigh	2	2	big	med	unacc
10	unhigh	unhigh	2	2	big	high	unacc
11	unhigh	unhigh	2	4	small	low	unacc
12	unhigh	unhigh	2	4	small	med	unacc
13	unhigh	unhigh	2	4	small	high	unacc
14	unhigh	unhigh	2	4	med	low	unacc
15	unhigh	unhigh	2	4	med	med	unacc

Preparing the data for WEKA - 2

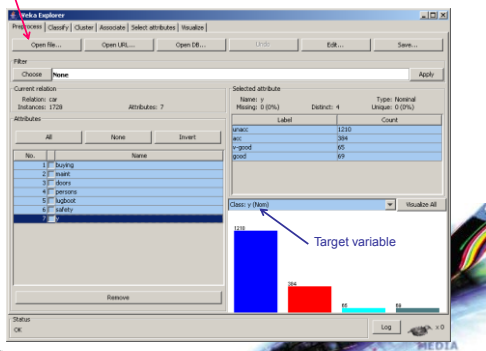
Save as ".csv"

- Careful with dots ".", commas "," and semicolons ";"!

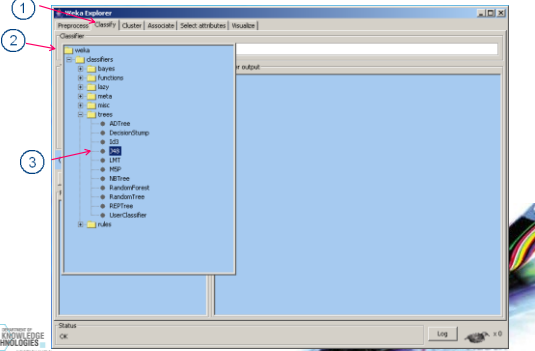


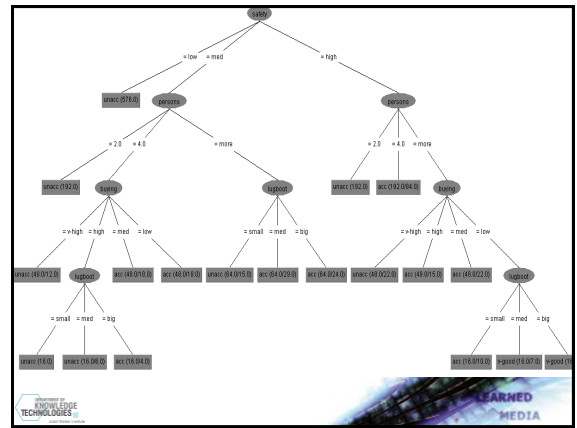
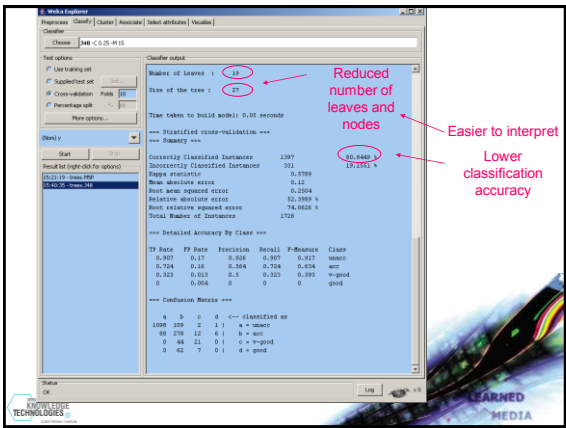
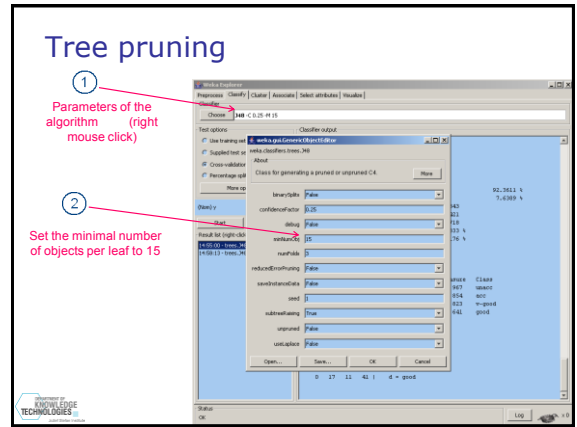
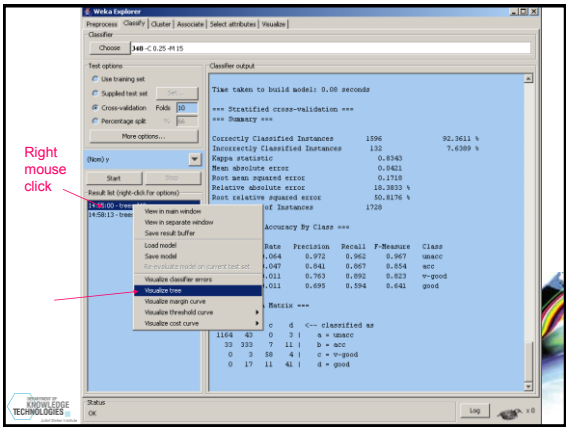
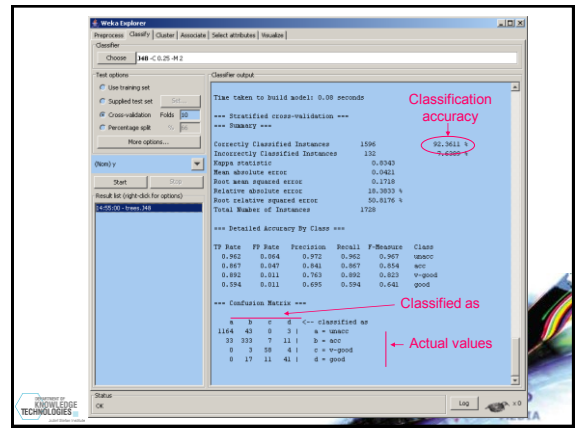
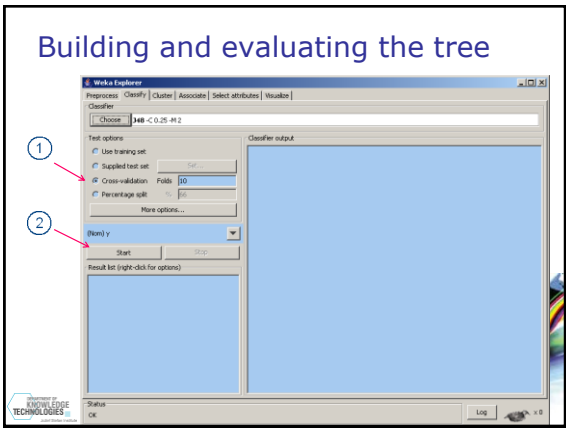
Load the data

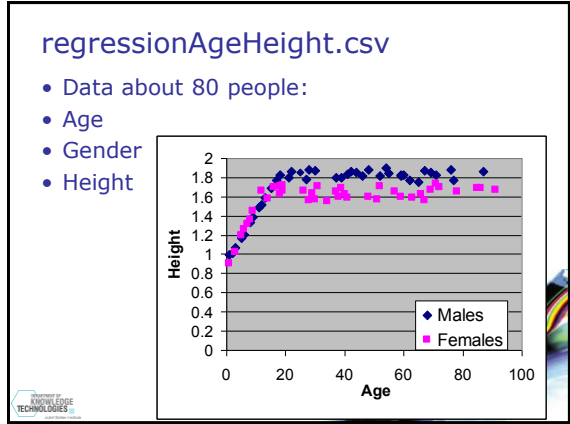
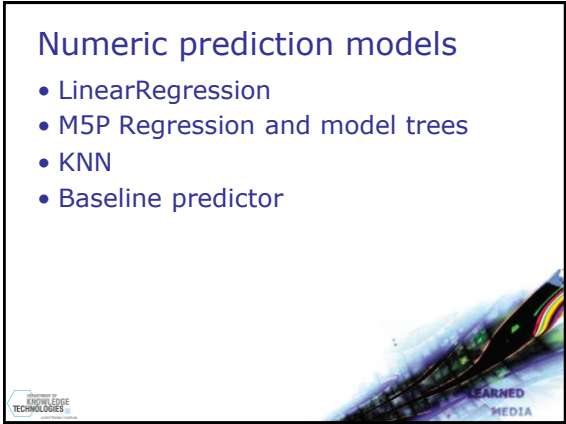
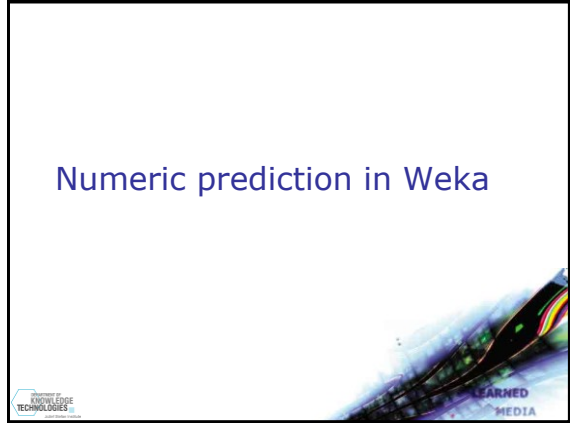
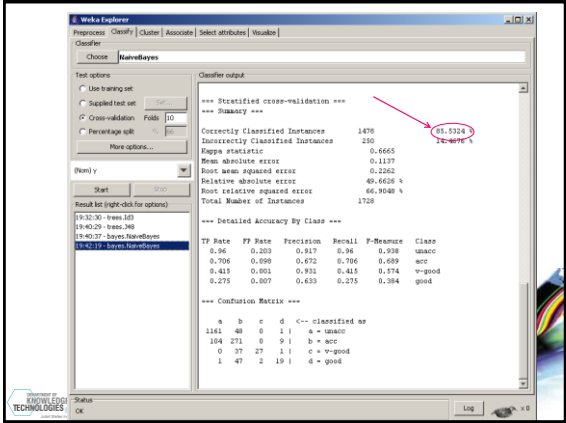
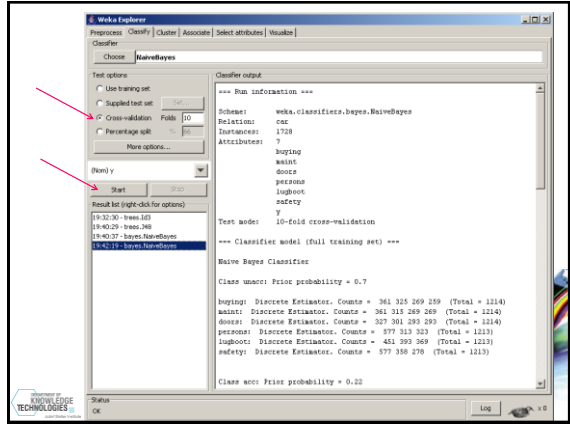
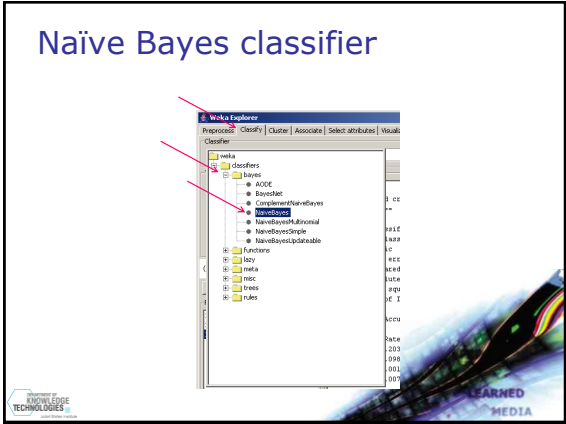
Car.csv

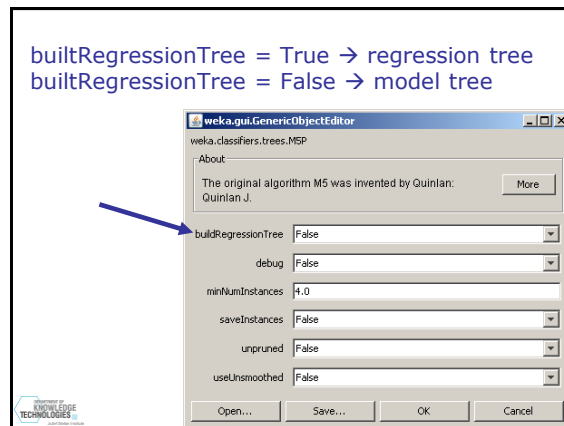
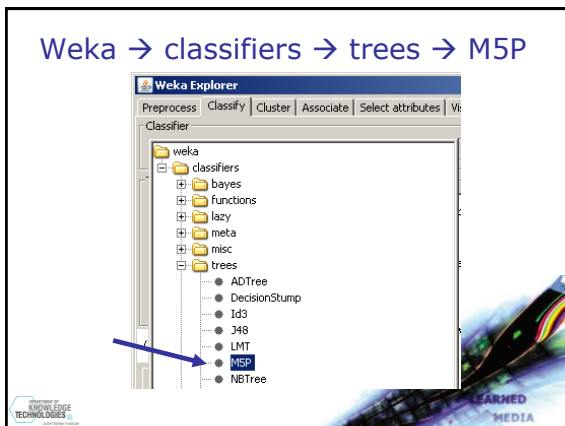
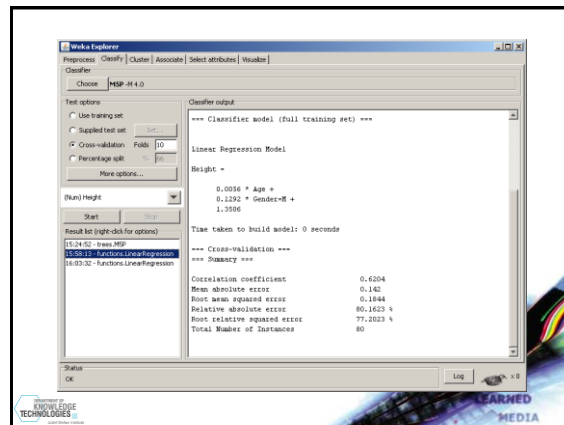
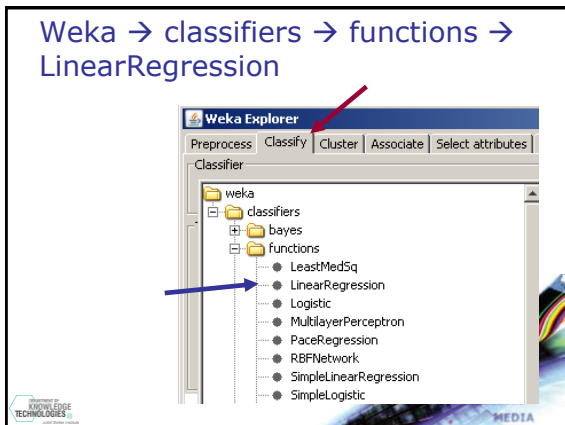
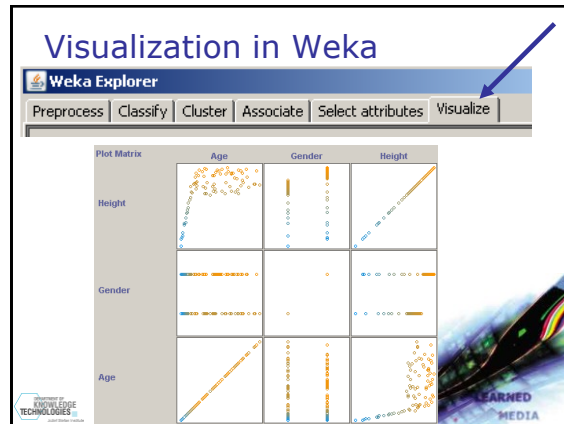
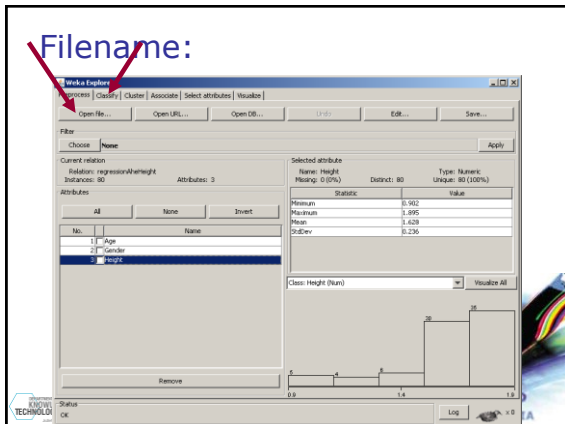


Choose algorithm J48

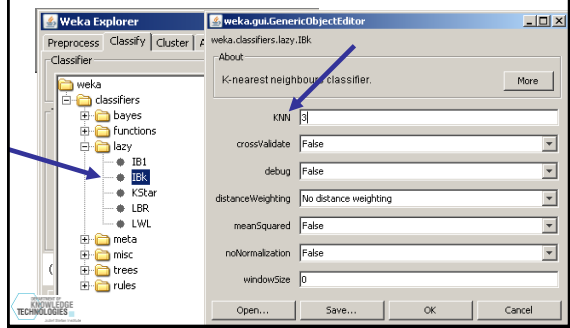








kNN: weka → classifiers → lazy
→IBk



Weka → classifiers → rules →
ZeroR

