University of Zurich[UZH]

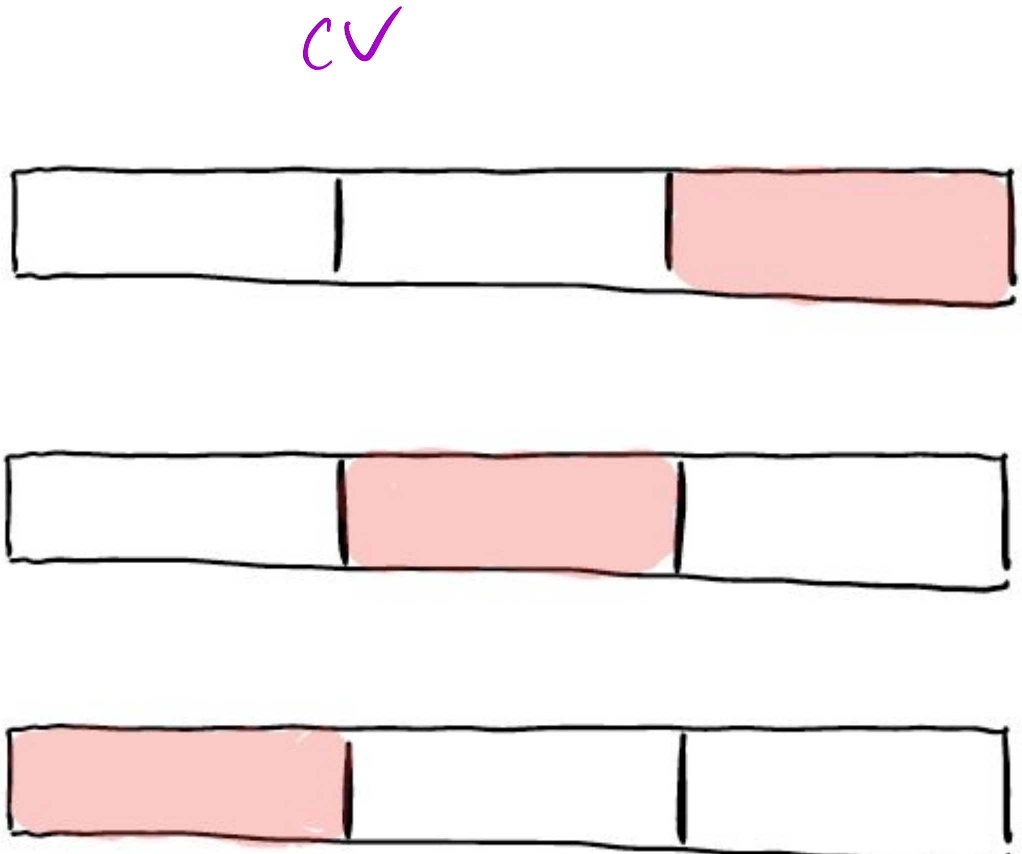
Institute of Computational Linguistics

Introduction to Machine Learning

Lesson 6

Mathias Müller, Phillip Ströbel

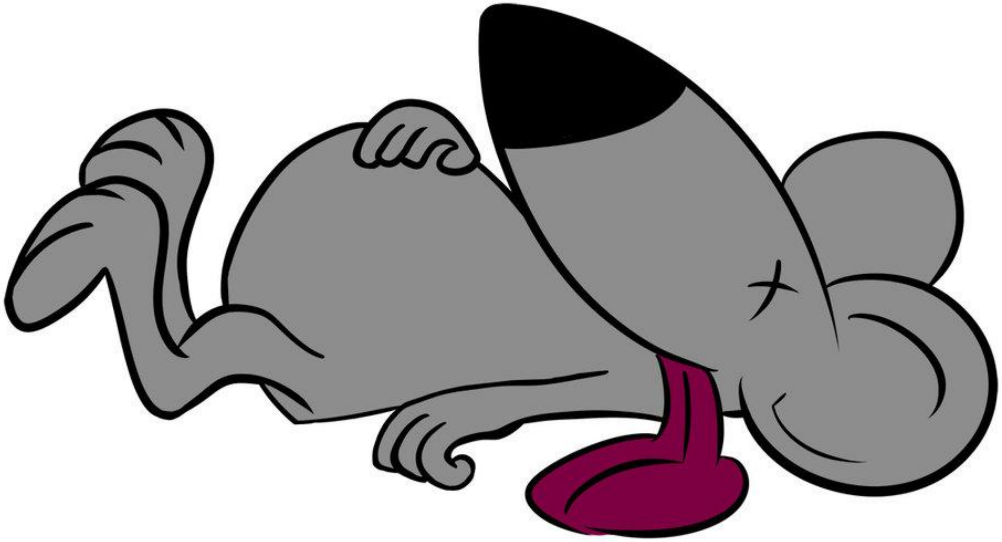
Wo waren wir



Today

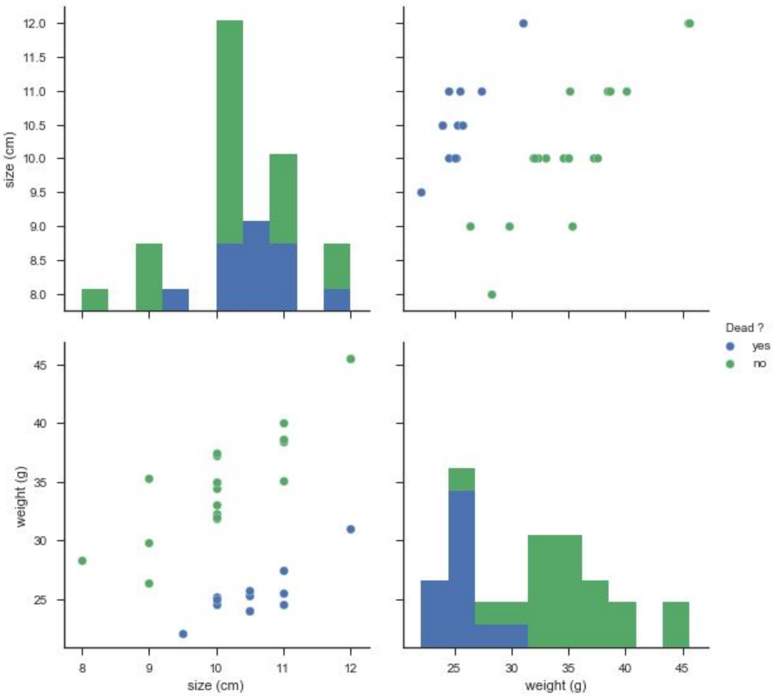
- **Overview of important classifiers**
  - (KNN)
  - Logistic regression
  - Decision Trees, Random Forests
  - Naïve Bayes

Dead Mice



Data Set

animal ID	size (cm)	weight (g)	Dead ?
1	11	25.5	yes
2	10.5	24	yes
3	9.5	22.1	yes
4	10	24.5	yes
5	10.5	24	yes
6	10.5	25.3	yes
7	8	28.3	no
8	11	38.4	no
9	11	38.6	no
10	10	25.2	yes
11	11	35.1	no
12	10	35	no
13	12	45.5	no



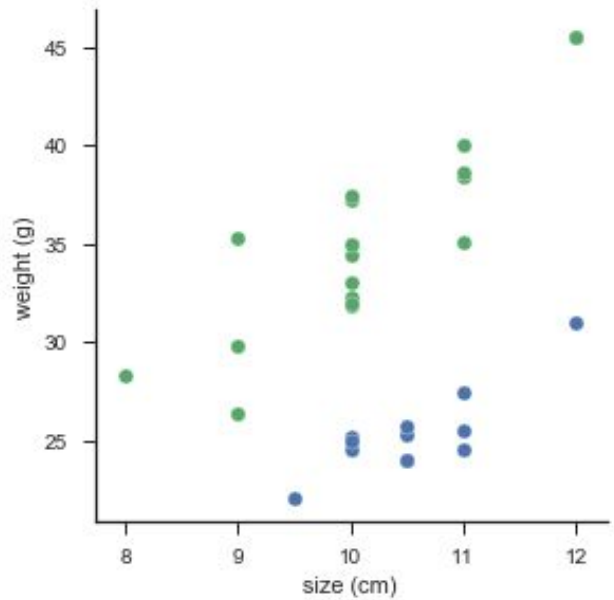
animal ID	size (cm)	weight (g)	Dead ?
1	10	26.2	

Notation

animal ID	size (cm)	weight (g)	Dead ?
1	11	25.5	yes
2	10.5	24	yes
3	9.5	22.1	yes
4	10	24.5	yes
5	10.5	24	yes
6	10.5	25.3	yes
7	8	28.3	no
8	11	38.4	no
9	11	38.6	no
10	10	25.2	yes
11	11	35.1	no
12	10	35	no
13	12	45.5	no

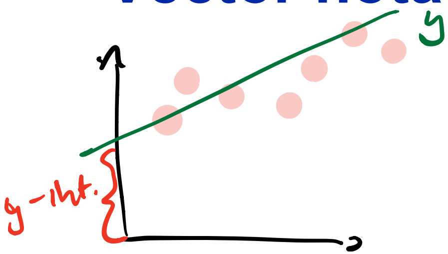
animal ID	size (cm)	weight (g)	Dead ?
1	10	26.2	

KNN



animal ID	size (cm)	weight (g)	Dead ?
1	10	26.2	
8	11.5	30	
9	11	41	

Vector notation for linear regression



$$y = c_1 x_1 + c_2 x_2 + b$$

coefficients (c_1, c_2); b: y-intercept, bias

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$c^T = [c_1 \;\; c_2]$$

$$y = c^T x + b$$

$$[c_1 \;\; c_2] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$0.7 + 0.2 = 0.9$$

$$c = \begin{bmatrix} c_1 \\ c_2 \\ b \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \qquad y = c^T x$$

animal ID	size (cm)	weight (g)	Dead ?
1	11	25.5	yes

How about linear regression for classification?

case → $y = w^T x$

logistic regression

17

−11

labels:

pos 1.0

pos 1.0

neg 0.0

$$\begin{cases} y >= 0.5 \rightarrow \text{pos} / 1.0 \\ y < 0.5 \rightarrow \text{neg} / 0.0 \end{cases}$$

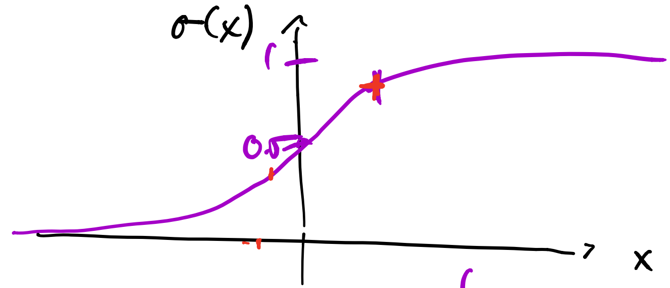
Logistic Regression

$y = \sigma(w^T x)$ → 17, −11

sigmoid function

17 → 0.9 → pos

−11 → 0.4 → neg



$$\sigma(x) = \frac{1}{1+e^{-x}}$$

y is probability of positive class

only 2 classes!

binary classification!

Dead Mice

Jupyter Notebook dead_mice.ipynb

2 ways to extend binary logistic regression

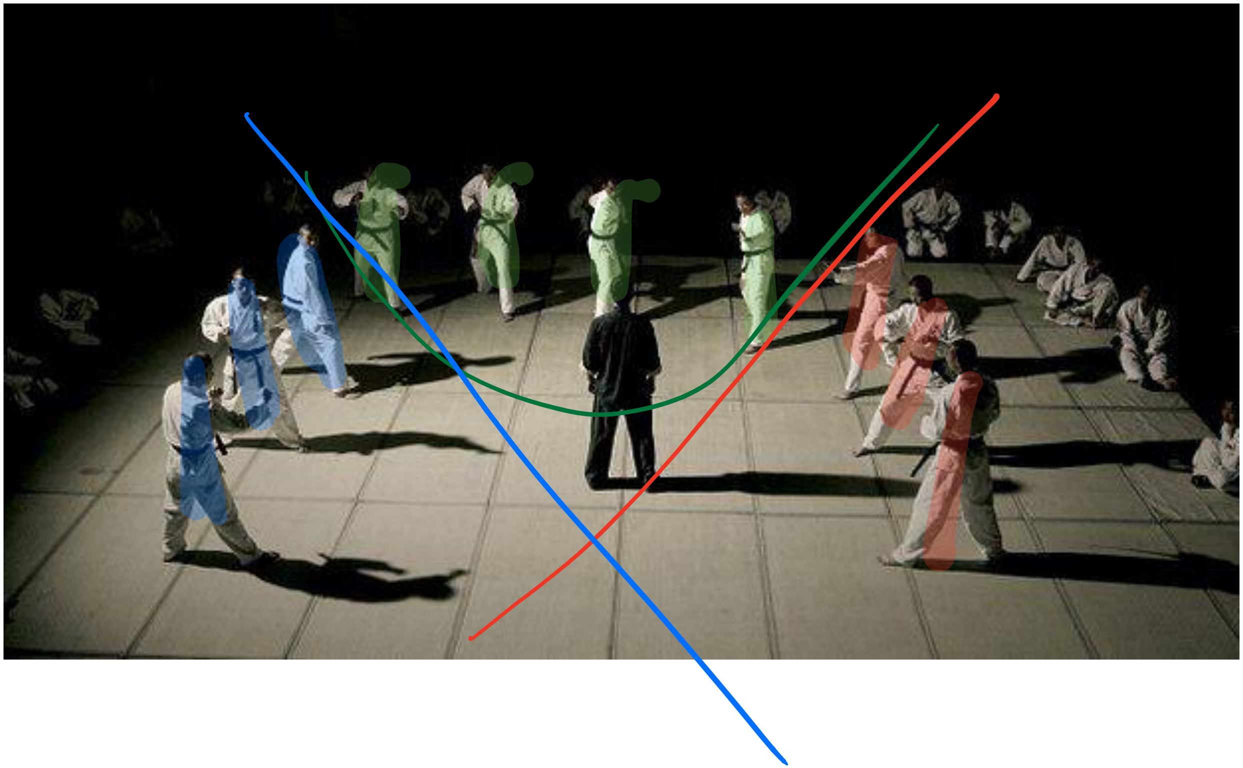
rest

- One-versus-all logistic regression
- Softmax regression (= Maximum Entropy)

Multinomial log. regression

One-versus-all multi-class logistic regression

$$y = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix} \quad \begin{matrix} 0.8 \\ 0.1 \\ 0.5 \end{matrix}$$



One-versus-all multi-class logistic regression

Softmax regression (= Maximum Entropy)

if $k=2$ → like logistic regression

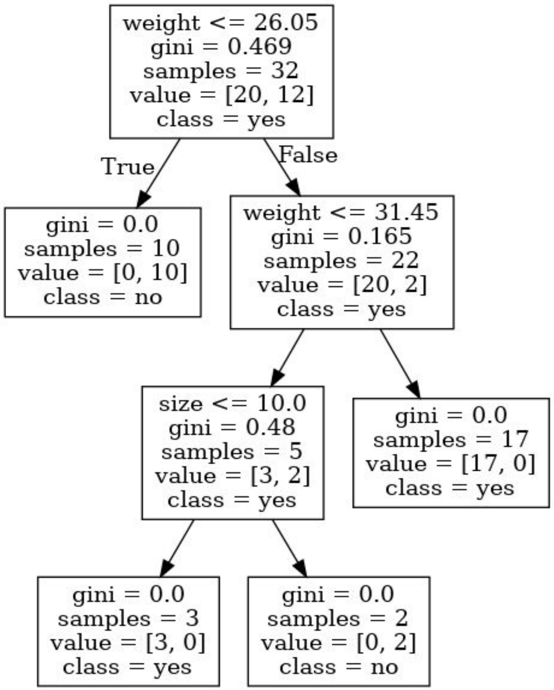
$$S = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

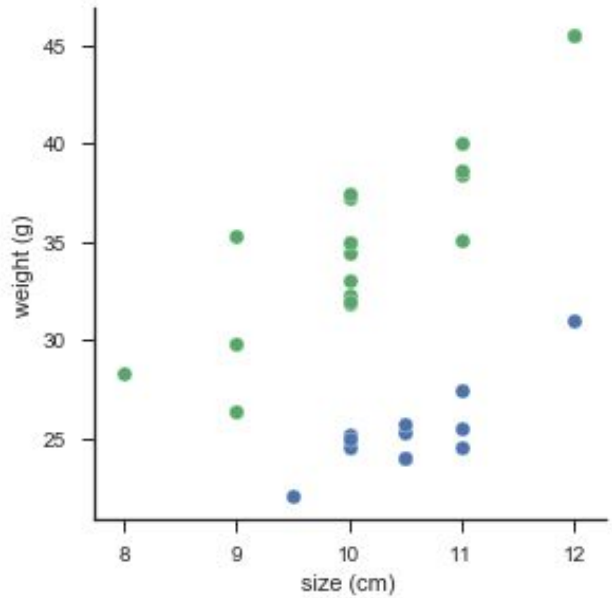
most probable class ↑

$$S\left(\underbrace{\begin{bmatrix} 0.1 & 0.5 & 1 \\ 1.1 & 2.3 & 1 \\ -1.1 & -2.3 & 1 \\ -1.5 & -2.5 & 1 \end{bmatrix}}_{4 \times 3}^{X} \underbrace{\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.1 & 0.2 & 0.3 \\ 0.01 & 0.1 & 0.1 \end{bmatrix}}_{3 \times 3}^{W} \right) = \begin{bmatrix} 0.1 & 0.2 & 0.9 \\ \vdots & & \\ \vdots & & \end{bmatrix}$$

y: 0.2 0.3 0.5

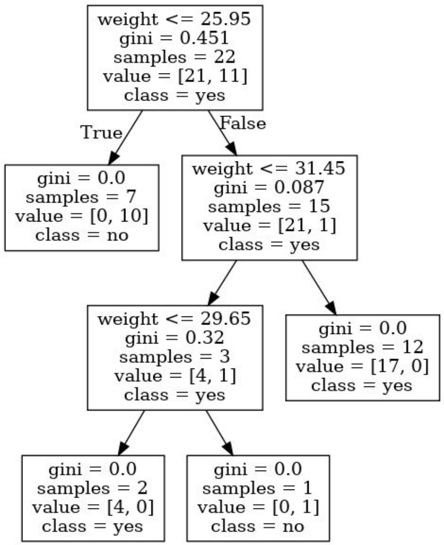
Decision Trees

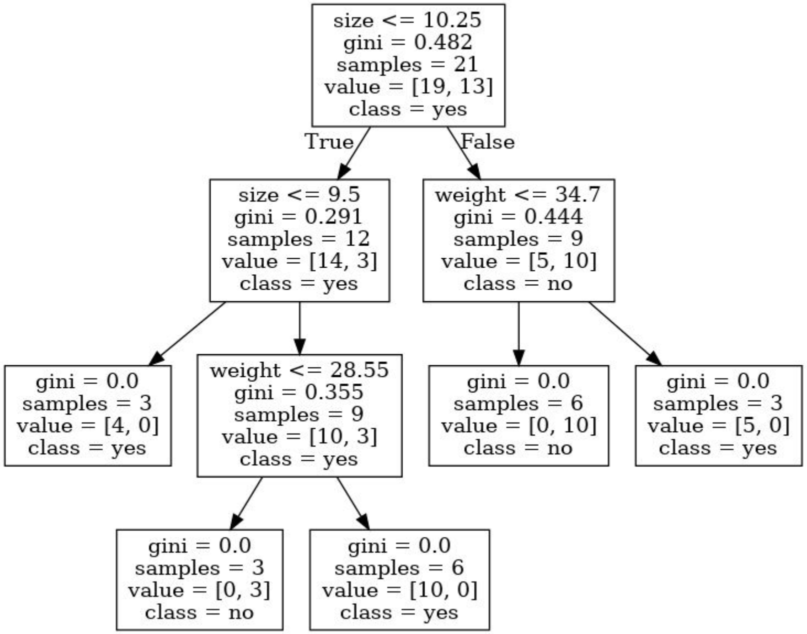




2 ways of using ensembles of decision trees

- Random Forests





Random Forests

- random subsets of trees

Naïve Bayes

Small excursion into probability theory

Naïve Bayes Theorem

Naïve Bayes Example

Naïve Bayes Example Data

play	outlook	temperature	humidity	windy
no	sunny	hot	high	false
no	sunny	hot	high	true
yes	overcast	hot	high	false
yes	rainy	mild	high	false
yes	rainy	cool	normal	false
no	rainy	cool	normal	true
yes	overcast	cool	normal	true
no	sunny	mild	high	false
yes	sunny	cool	normal	false
yes	rainy	mild	normal	false
yes	sunny	mild	normal	true
yes	overcast	mild	high	true
yes	overcast	hot	normal	false
no	rainy	mild	high	true

Naïve Bayes Example

~~Naïve Bayes Example~~

Summary

- vector notation for lin. reg.
- almost the same for log. regression
- logistic reg. only for binary problems
- extension for > 2 classes
  - one-vs-rest
    ↳ still log. regression
  - softmax classifier
    ↳ probabilities for each class

Exercise

Notebook 6.ipynb → Logistic Regression