Technische Universität Braunschweig

Hannover Medical School

PLRI PETER L. REICHERTZ INSTITUTE FOR MEDICAL INFORMATICS

Logistic Regression

Lisa-Marie Bente

Division Data Science in Biomedicine
Peter L. Reichertz Institute for Medical Informatics
of TU Braunschweig and Hannover Medical School
lisa-marie.bente@plri.de, www.plri.de

Exercise

- Write a script (in Python, of course)
- Train your model
- Evaluate your model with testing data
- Prepare .csv file with results of prediction of validation data
- Validation data can be found in Stud.IP

Save stuff with pickle

```
1  import pickle
2
3  with open('model.pkl', 'wb') as file:
4      pickle.dump(model, file, pickle.HIGHEST_PROTOCOL)
5
6  with open('model.pkl', 'rb') as file:
7      model = pickle.load(file)
```

Data

- Heart disease
- Classes: risk vs. no risk
- 16 attributes, 3811 samples
- .csv

Attributes

- male
- age
- education
- currentSmoker
- cigsPerDay
- BPMeds: nominal
- prevalentStroke: nominal
- prevalentHyp: nominal, hypertension
- diabetes: nominal
- totChol: continuous, cholesterol level
- sysBP: continuous, systolic blood pressure
- diaBP: continuous, diastolic blood pressure
- BMI: continuous, Body Mass Index
- heartRate
- glucose: continuous, glucose level
- **TenYearCHD**: binary, 10 year risk of coronary heart disease (1: yes, 0: no)

Things to keep in mind

- Explore your data before you start with the model
- Deal with missing data
- Split your data into training and testing set
- Have fun!