Applied Machine Learning and Predictive Modelling 1

Dr. Luisa Barbanti
luisa.barbanti@hslu.ch

Fall Semester 2025 (HSLU)

Outline

Section 1

Admin

Lecturers

- Luisa Barbanti (weeks 1, 2, 3 and 7)
- Daniel Meister (weeks 4, 5 and 7)
- Module responsible: Luisa Barbanti
- Module assistant: Ramón Christen

Admin

- 6 lessons (weekly!, not lecture on October 2nd 2025)
- 08:15 - 11:35
- 45-50 minutes lecture alternated to breaks
- Complementary exercise sheets (solutions available on the weekend)
- Based on **R**, available in python too
- Written exam and group work
- 3 ECTS

Office Hours

- Office hours are online on Wednesdays 5-6PM
- Same link as for the lecture :-)
- If nobody shows up in the first 15 minutes, Ramón will leave the zoom meeting
- First office hour: Wednesday, September 24th
- Last office hour: Wednesday, November 5th

Assumed prior knowledge

- Basic statistics
- Linear Models
- Generalised Linear Models (basics)?
- Basic **R**

Course Aims

- Students familiarise themselves with the main machine learning methods[1] currently used in Data Science
- Focus is on application of the statistical learning methods &
- Correct interpretation of the modelling results

[1] ML1 focuses on supervised learning.

Topics

1. Admin + Linear Models
2. Extending the Linear Model: Non-linearity
3. Extending the Linear Model: Generalised Linear Models
4. Model Validation
5. Support Vector Machines
6. Neural Networks
7. Big picture + Conclusions

Supporting Material

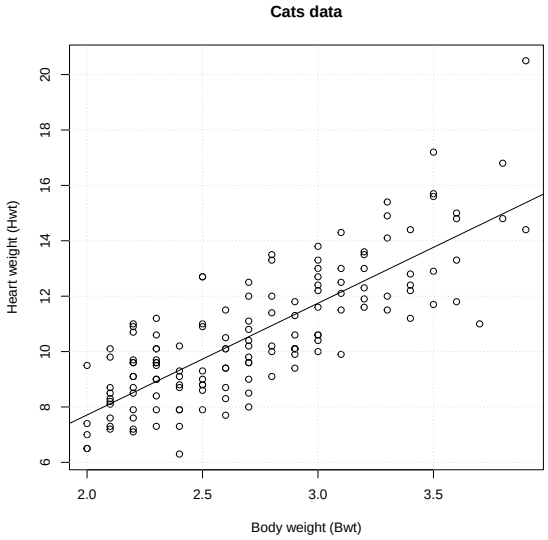
- Book: "An Introduction to Statistical Learning with Applications in R". Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 2023
- Free book download: `https://www.statlearning.com`
- Book: "An Introduction to Statistical Learning with Applications in Python". Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor, Springer, 2023
- Free book download: `https://www.statlearning.com`
- Video lectures by Hastie & Tibshirani: `https://www.r-bloggers.com/in-depth-introduction-to-machine-learning-in-15-hours-of-expert-videos/`
- NB: **the course does not follow the book very closely. However,** it is a very good idea to complement the course with a book or online material
- The latter comment is especially true in Data Science where very many fields flow together and therefore terminology and views can differ quite dramatically

About me

Section 2

Linear Models

Regression


Cats data
Heart weight (Hwt)
Body weight (Bwt)
6
8
10
12
14
16
18
20
2.0
2.5
3.0
3.5

Regression

$$y = \beta_0 + \beta_1 \cdot x_1 + \varepsilon$$

$$Hwt = \beta_0 + \beta_{Bwt} \cdot Bwt + \varepsilon$$

```
lm(Hwt ~ Bwt, data = d.cats)
```

$\Rightarrow$ **See R-code "Cats Lab"**

Regression: Coefficients

- β_0 and β_1 are the regression parameters
- β_0 is the intercept
- β_1 is the slope for the "body weight" predictor
- Regression coefficients are estimated from data
- Regression coefficients are estimated with the "Least Squares" method

Regression: Coefficients

- Estimated regression coefficients are denoted with a hat (e.g. $\hat{\beta}_1$)
- Fitted values (i.e. what the model predicts) are denoted with a hat as well (i.e. $\hat{y}$)
- $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1$
- Residuals are the difference between observed and predicted values
- $\text{res} = y - \hat{y}$

Regression: Coefficients

- If the errors are normally distributed, regression coefficients can be tested with t-tests[2]
- "Dichotomising" p-values into "significant"/"non-significant" is very bad practice!
- The intercept coefficient is not to be tested. Any linear model must contain an intercept
- **R** uses treatment contrasts to deal with categorical variables

[2] More about distributional assumptions will come later in this course.

Regression: Coefficient of determination

- The "goodness of fit" can be quantified with R^2
- $R^2 = \text{corr}(y, \hat{y})^2$
- $0 \leq R^2 \leq 1$
- R^2 must not be considered as an "absolute" measure, but rather as case-specific
- R^2 is not used to formally compare models (see F-test later)
- The adjusted R^2 takes into account the complexity of a model (i.e. the number of parameters)

P-values

- The p-value quantifies the probability of observing the value of the test statistic, or a more extreme value, under the null hypothesis
- Low p-values are coherent with a rejection of the null hypothesis stating that there is no effect
- For example, in the cats data the p-values for the predictor body weight is extremely small (close to zero) indicating that there is strong evidence that this variable plays a role
- Large p-values (close to 1) do not imply the we can accept the null hypothesis