Applied Machine Learning and Predictive Modelling 1

Dr. Luisa Barbanti and Dr. Matteo Tanadini
luisa.barbanti@hslu.ch

Fall Semester 2025 (HSLU)

Outline

Section 1

Modelling non-linearities

Linear Models

Given a Linear Model:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \varepsilon$$

This model is said to be linear because it is linear in its coefficients. Indeed, the following is a Linear Model.

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_1^2 + \beta_3 \cdot log(x_2) + \varepsilon$$

Linear Models

A model such as:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot {x_1}^2 + \beta_3 \cdot log(x_2) + \varepsilon$$

CAN model non-linear relationships. In **R** this model would be fitted as:

```
lm.non.lin.1 <- lm(y ~ x_1 + I(x_1^2) + log(x_2), data = someData)
```

Non-linear relationships

By including polynomials (e.g. $x_1 + {x_1}^2$) we can model non-linear relationships with a Linear Model.

This is also true for extensions of the Linear Model such as Generalised Linear Models.

Non-linear effects

- Linear models can model non-linear effects
- e.g. by adding quadratic terms
- $y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \varepsilon$

- **non-linear models are non-linear in their coefficients:**
- $y = \beta_0 + \beta_1 \cdot x^{\beta_2} + \varepsilon$

Generalise Additive Models

In alternative to polynomials we can use GAMs, which are an extension of the Linear Model.

Generalise Additive Models

GAMs come with advantages and disadvantages compared to e.g. polynomials. Here a very short selection of the most practice-relevant ones:

+ the degree of complexity must NOT be set by the user
+ the "estimated degrees of freedom" gives the user an indication of the complexity of a given smooth term
+ smooth terms can be visualised
− GAMs can run into computational issues (e.g. models that do not converge)
− the use of a quadratic term is simpler to explain than a GAM to a non-technical audience
− in order to fit and understand the results of a GAM some technical knowledge is required... and GAMs are a complex topic

Must know for a Data Scientist

A widespread wrong myth in data science is

"... we don't use Linear Models (or their extensions) because they cannot deal with non-linear relationships"

This is wrong!
Linear Models CAN MODEL NON-LINEAR RELATIONSHIPS.