UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

HPAI
HIGH PERFORMANCE
ARTIFICIAL INTELLIGENCE

XNDL

Lab1 - GLMs

Generalized Linear Models

❖ Gaussian: Continuous response var (mean & variance)

■ Regression problems

■ Normalizing helps w/ outliers & coefficients

Residual plot

❖ x axis: linear prediction

❖ y axis: error measure

❖ Identify outliers

❖ Data non-/linearity

❖ Heteroscedasticity (error variance changes with predicted value)

■ Linear regression assumption

[1,2,3]

Checking Normality

- ❖ Shapiro-Wilk

    - ■ Distribution shift at each point

    - ■ p-value: Prob. of sampling from a Gauss.

    - ■ Low p-value (threshold?) -> Not normal

- ❖ Kolmogorov-Smirnov

    - ■ Distance in cumulative distribution

    - ■ Compare against "Gaussian"

    - ■ Low value -> High normality

[4]

To Do

❖ Understand data

❖ Check coefficients

❖ Find which dataset is better fit for a linear regression.

❖ Add an outlier and try to detect it in the plots

❖ Improve code:

■ Print functions

■ Load dataset function & input param

■ stats functions

Generalized Linear Models

❖ Poisson: Positive integer response var (num. occurrences)

- Count problems

Incidence Rate Ratios / Confidence interval

❖ Importance of variables for target

❖ Precision of estimates

❖ To Do

  ■ Find most relevant variables

  ■ Improve code & structure

References

[1] https://statisticsbyjim.com/regression/check-residual-plots-regression-analysis/

[2] https://statisticsbyjim.com/regression/heteroscedasticity-regression/

[3] https://www.statisticshowto.com/residual-plot/

[4] https://www.spss-tutorials.com/spss-shapiro-wilk-test-for-normality/

Closure

The end