Text, table

Description automatically generatedText

Description automatically generatedText, letter

Description automatically generatedGraphical user interface, text, application

Description automatically generated

Linearity assumptions (for LS regression)

Mallows Cp statistic

Q-Q plot



Order data

Does indicator variable affect thorax length?

Reject H0 if Pr(>F) <= 0.05

Scatterplot



LOOCV in R

A screenshot of a computer

Description automatically generated with low confidence

Non-parametric regression in R



Nadaraya Watson (norm kernel)

Local polynomial

Text, letter

Description automatically generated

Smoothing splines



TODO: Hat matrix (serie 5 methods and non-parametric regression)

TODO: Equivalent ways of calculating LOOCV using hat matrix and manually computation

Bootstrap in R





Tukey Anscombe is residuals versus fitted values. Q-Q is empirical quantiles vs standard normal quantiles.

Note that the linear regression model does not assume a normal distribution for the predictors, but a skewed distribution and outliers often result in  
regression solutions that are largely determined by very few points.

Forward selection (in case p, the nr. of predictors vars. Is too large for exhaustive search):

1) start with smallest model

2) add predictor which reduces the MSE the most

3) repeat 2 until all predictors or large nr. of predictors selected (now a seq. models is produced)

4) choose the model in the seq. which has smallest Cp statistic.

Backward selection is obvious.

which is prop. to Cp(M):

The MSE can be estimated by:

Text

Description automatically generatedA picture containing text, clock, watch

Description automatically generatedA picture containing text, clock, watch

Description automatically generatedA picture containing text, watch

Description automatically generatedA picture containing text, watch, gauge

Description automatically generatedA picture containing text, clock

Description automatically generated

Bootstrapping LDA/QDA

Maybe add all-subsets regression in R (from HW2)?

Whereas Q-Q plot is given by

For lm, Tukey Anscombe is given by



Forward and backward selection in R



Decision boundary plot in R



Decision boundary between class 0 and 1:

QDA Classifier (more easily overfits if p is big)

if the covariance matrix ∑ is diagonal and const. then w is parallel to the line connecting µ0 and µ1



Prediction of LDA found by taking argmax: of:

LDA Classifier

Bayes Risk Bayes Classifier

Expected size of out-of-bootstrap sample [roughly 1/3 of points will be out of sample]:

“basic” = “Reversed quantile”

“norm” = “Normal”

“perc” = “Quantile”