Text, table

Description automatically generatedA picture containing text, clock, watch

Description automatically generatedText, letter

Description automatically generatedGraphical user interface, text, application

Description automatically generated

LHS is logit transform of π

(Linear) logistic regression

Linearity assumptions (for LS regression)

Scatterplot

Diagram

Description automatically generated with low confidence

For J>2

A picture containing text, screenshot

Description automatically generated

Likelihood (for Bernoulli x then for Binomial)







LOOCV in R





A screenshot of a computer

Description automatically generated with low confidence

Multinomial logistic regression

will give the probability of heart disease (N/m)

Non-parametric regression in R



Nadaraya Watson (norm kernel)



Does indicator variable affect thorax length?

Reject H0 if Pr(>F) <= 0.05

Text

Description automatically generated

where the probabilities of each class are stored in the columns (columns same order as levels (…)) but we can also achieve logistic regression by using dummy encoding and glm:

Order data

Local polynomial

Text, letter

Description automatically generated

Text

Description automatically generated

Q-Q plot

Smoothing splines



ROC and Misclassification graph in R



Text

Description automatically generated

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



which plots the TPR vs FPR. To plot the misclassification rate as a function of threshold prob. (i.e., after this it is considered positive)



K-Fold CV in R

Mallows Cp statistic

Text

Description automatically generated

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 generatedText

Description automatically generatedA screenshot of a computer

Description automatically generated with medium confidenceText

Description automatically generatedText

Description automatically generatedText

Description automatically generated with medium confidenceText

Description automatically generated with medium confidenceA picture containing text, clock

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 generated

To plot the same thing but for lm use:

Bootstrapping LDA/QDA

To plot the splines do:

Alternatively,

Generalized additive model in R

Poly. regression using a polynomial of degree d (no interaction term):

And then pred.cv <- prediction (all.y.pred, all.y.true) and proceed as normal (but add avg = “threshold” for ROC and avg = “vertical” for misclassification) use add = T to add them to same plot. One can specify the cost for FN and FP using cost.fp = 0.8, cost.fn = 1.2 as args to performance().

To plot the average ROC (across all folds) initialize a vector of lists and assign them as follows:

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”