]overfit

validation is test data set

record drop

box plot

categorial value

diffrence between R^2 and R^2 adj

**data processing:**

step1:Data prepocessing: outlier and missing value :rows

step 2 :modeling : variable importance :(select/drop)

step3 validation

(y prediction)

nltk

seq to seq

**logistic regression:**

sigmoid curve,s curve

cut off should increase 0/1

1 must be less

confusion matrix:pred/act

sensitivity

type1 error

type2

specificity

concordance

discordance

**hosmer-lemeshow test**