January 6th 2020

Week 1

Modeling: assuming something theoretical

For example,

a = bx0+ b1x1+b2x2….

…Assumes linearity

…Assumes finite variables

…Assumes quantifiable variables

Most examples or lab work will be in R

Survey class

Problem sets, usually dealt on Monday, due Sunday.

Mostly lecture, will be interactive. Read the material and engage.

Communicate on piazza. Github. Github invitation will be sent out through uchicago email.

One single rendered pdf for handing in the homework.

Closed book, pen and paper exam for the finals.

3/16 exam (and a review session for the finals before)

r markdown crash course will be posted on piazza

January 8, 2020

Measurement for coming up with a new theory

What you observe

Events division in time series

Controlled events different outcome

Counterfactual

New theory that challenges/ reconcile new, unexplainable data

Theoretically, exogenous, endogenous

System created in the function, within: endogenous

Theory: coming up with a new world

Case study: linking theory and model. Eg. rational utility

Formal theory, choice theory

January 13th, 2020

To find f(x):

Prediction: maximize accuracy, might neglect the mechanism and the ability to generalize (blackbox)

making inference : explanation over prediction, mechanism, meaningful change

1.3

make assumption:

OLS linearity and additivity(independent variables)

Fit, difference between prediction and actual value

Minimize the complexity of the model (not the space)

Making assumption <-> over-fitting

Low s smoother

Span (how rough or smooth)

1.7

supervised: measured x and y

unsupervised: no target you are predicting, recover latent structure in the data

1.8 statistical learning and machine learning

goals: statistical (for inference, to generalize, phehomenon, the data generating process),

machine learning (maximize accuracy, best learner to predict, to program itself )

classification and regression

regression (quantitative features)/ classification(qualitative features)

1.10

to find the optimal values

em estimation

markup chain monte carlo

1.11

MLE,

OLS being one type of MLE

Y

g(mu)

g-1(E(Y))

2.

error: comes from suboptimal estimation

better technique, less reducible error

perfect model, where error comes from epsilon (irreducible), x just not predicts y

prediction value/ observed value

Loss function L(Y, f-hat(x))

(absolute error, aquared error, mean squred error(MSE))

pick the model that minimize the MSE

training error not so important, because its only for this dataset (can’t generalize)

more importantly, the test error

bias = E[f-hat(x)- f(x)]

var = E[f-hat(x)-E[f(x)]]2

bias- variance trade-off

3.4 classification

cross validation/ bootstrap, for limited dataset

(aic bic)