11-06-2020

Meeting with annie

Those doctors I met: super good in stats

Choosing models: dimensionality vs number of people; how difficult the question is; complexity

Explain basic stuff about knowledge and training and testing set; etc…

Include how much data points = test results

Bins: Doctors tend to work in bins, its more informative and easier to interpret

ROC curve + sensitivity + specificity: don’t we need just the ROC curve

ODDS ratio vs RISK ratio!! There are different in case there is low occurrences.

Likelihood ratio? Basically how much do I trust my test and go from prior Bayesian to post Bayesian. LR is not affected by prevalence, so can be used to adjust PPV, NPV! But is independent from LR. CHECK THIS.

Positive predictive value and negative PV depend on prevalence; they can be derived using Bayes.

Importance plots: 1) bivariate analysis: take dataset, theses are labels, these are features: test each feature and its relation to the label (loooong table) using odds ratio to quantify, with each value with a p-value. This is a selection procedure. Then check out oh this one should be related (from knowledge) but it’s not…. Maybe there are confounding factors, other associations. Adjusting = sort of adding features to the bivariate analysis to see how the feature behaves with another feature; then if there is, then you should correct / normalize / adjust to it! Before it was 6, now the value is 3 – adjusted value.

So maybe just add bivariate importance; then just do actual effect bivariate.

Add a CI for what we just talked. Basically want a CI for anything, but especially for the unadjusted values of the feature importance

ML 433 epfl ; Think about what work I can give CHECK

Bradford-hill causality criteria CHECK; very contextual: strength, consistency,

LIME: input not enough maybe include range with FEATURE + VALUE