



Ideas NOT related to G.C. Gen:

- ① Finding keystroke metrics that are good predictors of cognitive effort (e.g. t/len, revision no., jumps, deletes)

How: use keystroke metrics to predict variables that are already established predictors of ~~comp~~ cognitive effort
e.g. certain complexity measures / eye tracking.

Results: Hist Boosting model (lightBM) w/ L2 normalization

ks measure	R ² avg over 32 measures
t/len	0.5 %
fleng/lteng	1.4 %
per.info	0.4 %
jumps/deletes	3.2 %
Length	32.0 %
length + ks	44.0 %
none	-0.1 %

Extra: LSTM - NN
that automatically
generates best
keystroke metric
summaries from
raw log

- ② Finding groups of linguistic measures that are similar in nature / good predictors based on ks data

How: 1-by-1 analysis on difference in ks values over different values of linguistic complexity

Results: single dashed surfaces => ...

volume per complexity => ...

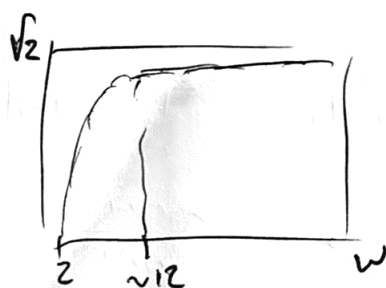
Also, R² of model for small groups for quantitative data

Ideas related to CoCoGen

key idea CoCoGen only makes sense if cplx metrics for close-by sentences influence one another / are related

- ① Basic statistical model to show correlation of values on \geq window size of w

Results



suggest best window size ≈ 12

- ② Use CoCoGen info as additional feature for model that predicts k_s = cognitive effort as a function of complexity

Research Questions

- ①: How can we compress information on the k_s log in metrics
- ②: How do specific linguistic metrics (# of clauses, ...) relate to specific k_s metrics \Rightarrow explain smth about writing process
- ③: Argue complexity is a LOCAL property, and CoCoGen as a tool makes sense