prediction

regression / classification

Y number   Y class / category

goal: get good predictions

evaluate predictions:
want good predictions on the population
training data = sample from population
 ↳ use this to estimate well for the population
    data splitting / CV

prediction error ⟶ combination of bias + variance

bias / variance tradeoff

linear models:
- linear regression
- variable selection in linear regression
- LASSO / Ridge / PLS
- spline / smoothing / GAM
- logistic regression
- LDA / QDA

non-linear methods:
- K-nearest neighbours
- neural networks
- trees + tree ensembles

unsupervised methods
- PCA
- K-means cluster

for a given method,
① what problem can you use it for?
② what good/bad properties does it have?
③ bias/variance tradeoff
④ overfitting
⑤ computational
⑥ types of predictors
⑦ interpretable
⑧ compare methods
⑨ work for small/large data