Today

- Ensemble methods
  - bagging (bootstrap aggregation)

Ensemble methods

- Train many models
- Average the models to predict
- Averaging reduces variance

e.g. $$\mathrm{Var}(\bar{y}) = \frac{\sigma_y^2}{n}$$

Bagging

- Bootstrap
  - form new datasets by resampling from the data
- Aggregate
  - average over bootstrapped model fits

Bagging algorithm

for many repetitions
    resample the data with replacement
    train the base model
    record prediction
final prediction = mean of predictions

Base model: can be any type of model

Out of bag error estimate

- Each bootstrap we can use the other samples to gauge prediction error
- Approx = LOOCV

Variable importance

- Average RSS decrease over splits for each variable