

Machine Learning

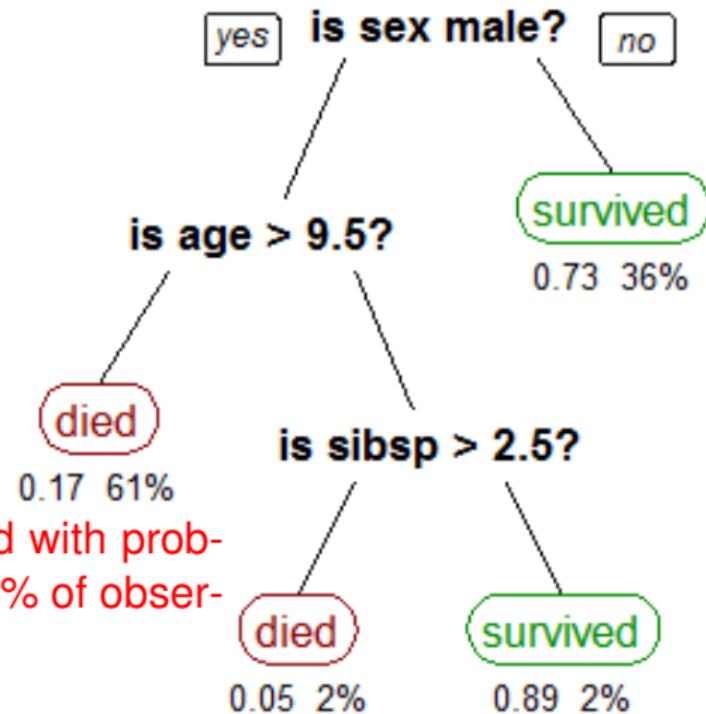
Random Forests

Jeff Abrahamson

9, 16 décembre 2016 13 janvier, 9 mars 2017



Decision Trees



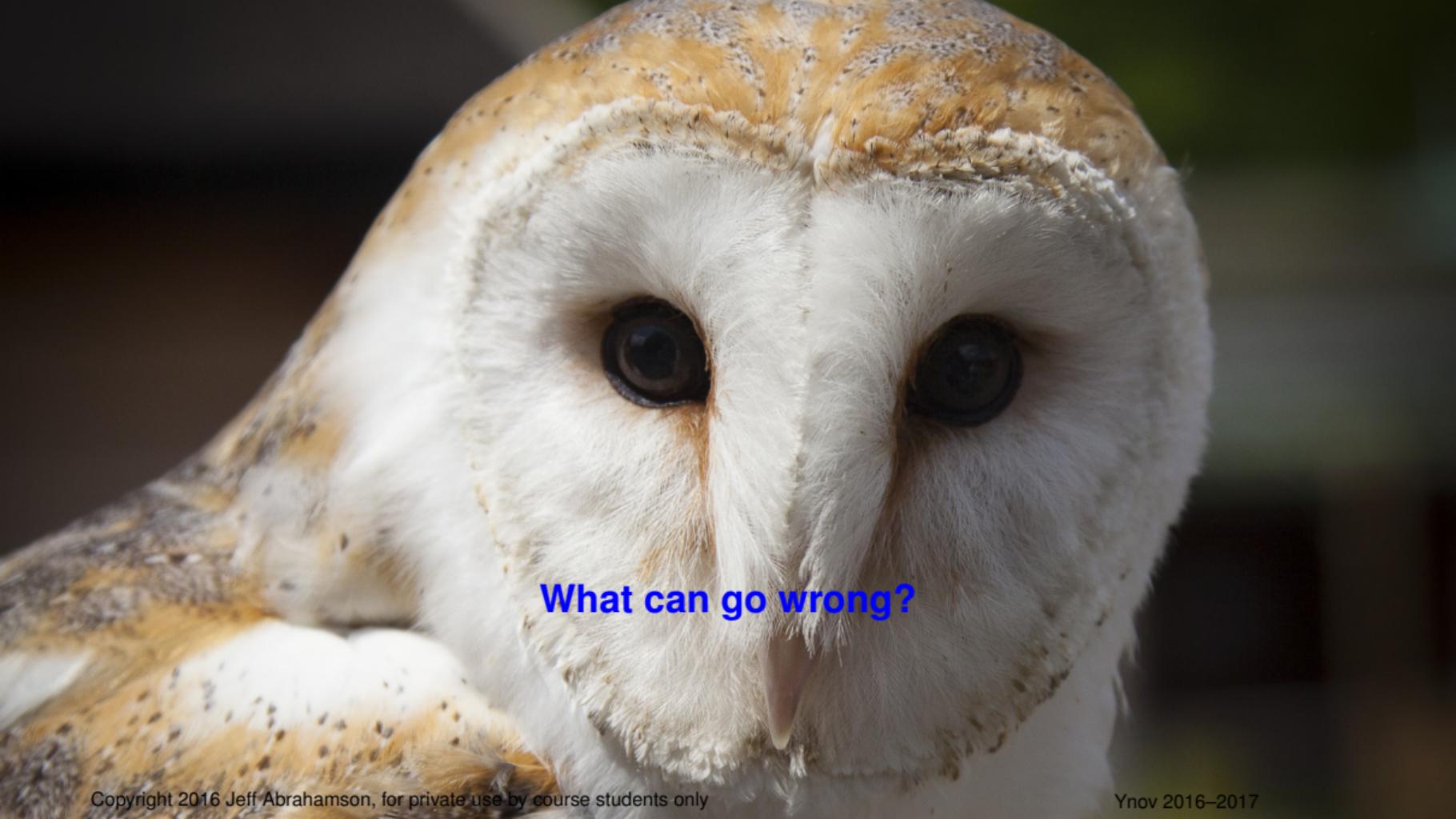
E.g., passengers died with probability .17 which is 61% of observations

Stephen Milborrow

Decision Trees

Variations

- Classification tree
- Regression tree

A close-up photograph of a Barn Owl's face. The owl has large, dark, almond-shaped eyes and a white, feathered facial disc with a distinct dark 'M' shape. Its plumage is a mix of light brown and white with dark spots. The background is blurred.

What can go wrong?

Decision Trees

Ensemble methods

- Bagging
- Random forest
- Boosted trees (*gradient boosted trees*)
- Rotation forest

Bootstrap aggregating = bagging

Bootstrap

A family of statistical methods using sampling with replacement.

Bootstrap aggregating = bagging

- Increase stability
- Increase accuracy
- Reduce variance
- Avoid overfitting

A type of model averaging.

Bootstrap aggregating = bagging

- Training set D of size n
- Sample D *with replacement* to create D_1, \dots, D_k of size n'
- If $n = n'$, expect $1 - 1/e \approx 63.2\%$ repeats

Bootstrap aggregating = bagging

- Training set D of size n
 - Sample D *with replacement* to create D_1, \dots, D_k of size n'
 - If $n = n'$, expect $1 - 1/e \approx 63.2\%$ repeats
-
- Train k models
 - Average (regression) or vote (classification)

Bootstrap aggregating = bagging

Do not confuse with

- Boosting (and AdaBoost)
- Bootstrap (statistics)
- Cross validation

Random subspace method

attribute bagging = feature bagging

Random subspace method

Bagging (bootstrap aggregation) = resampling to create more data sets, train models on different samples

Attribute bagging = project to create more data sets, train models on different samples

Random forests

Combine bagging with random subspace method

A photograph of a dense forest in a thick fog. The trees are tall and thin, their trunks silhouetted against the light. Bare branches reach out from the left side of the frame. In the center, the word "questions?" is written in a bold, blue, sans-serif font.

questions?