Bagging

Chenyue Cai, Oscar Xu, Yintang Yang

Ensemble method

- Multiple learners on one single machine learning problem
- Use the aggregated prediction based on these multiple learners

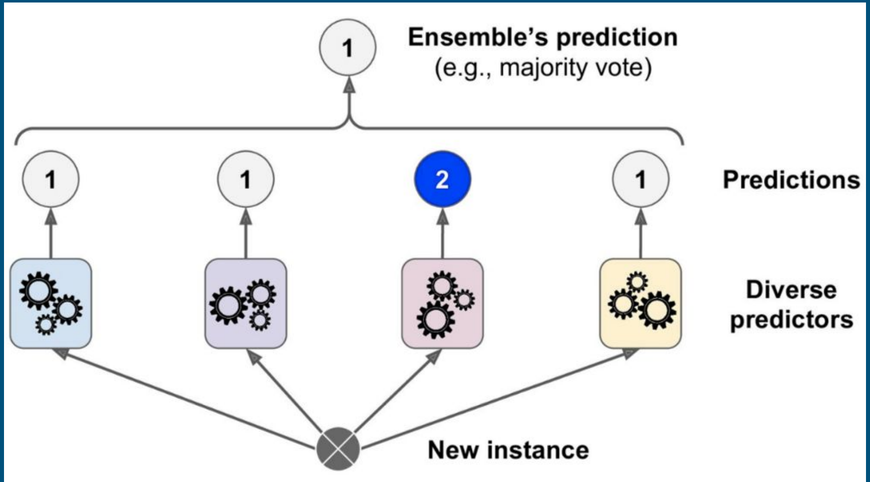


Figure 7-2. Hard voting classifier predictions

One issue that we always encounter :(

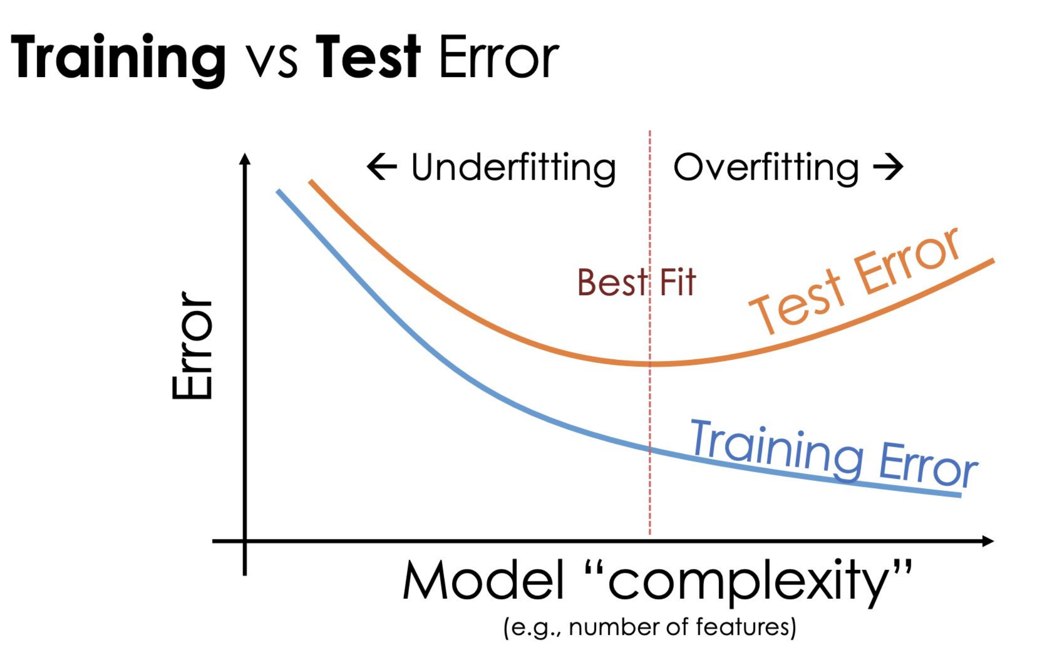


Image From: http://www.ds100.org/fa19/resources/assets/lectures/lec17/17-FEandRegularization.pdf

Overfitting a quick recap

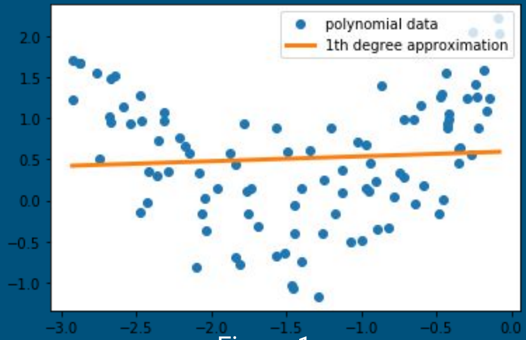


Figure 1

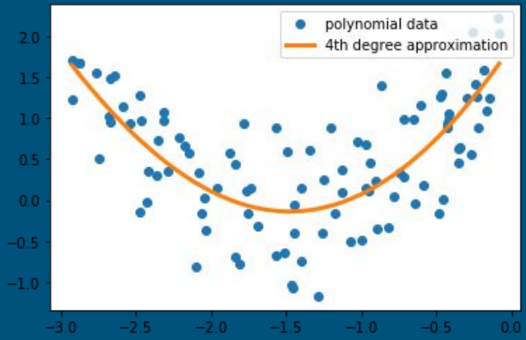


Figure 2

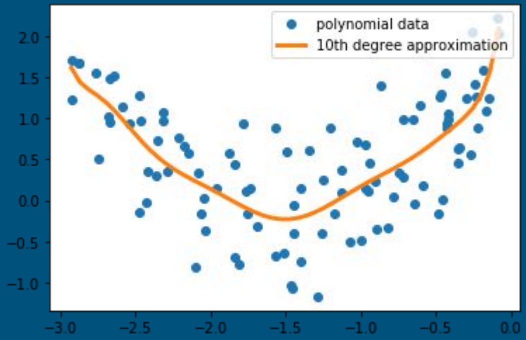


Figure 3

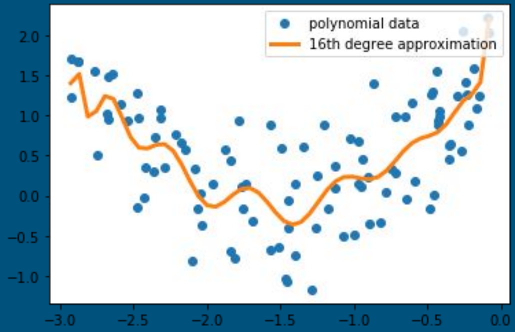
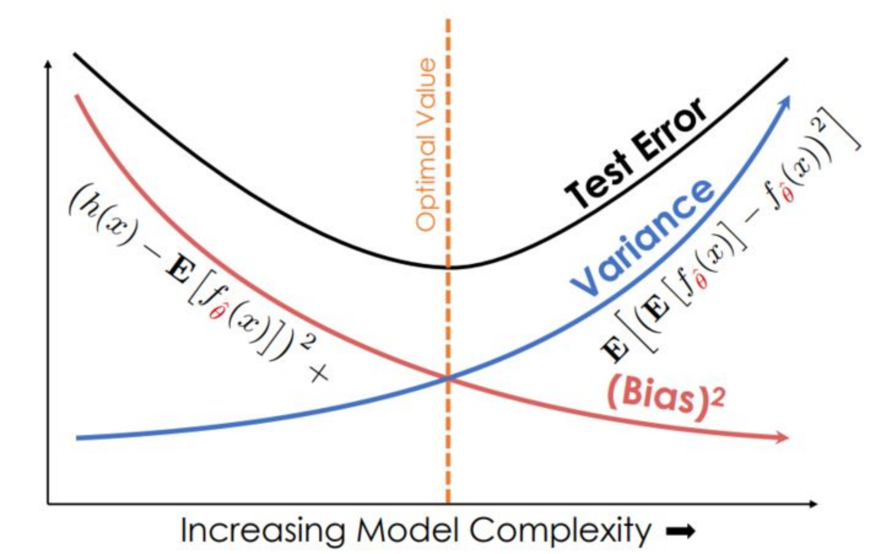


Figure 4

- Here is the best fit for different degrees
- Which one is underfitting?
- Which one is overfitting?

How to solve overfitting?

Recap: Interpretable metric of error
Bias and Variance



- Bias measures how well the model approximates underlying true function
- Variance measures how robust the model is towards perturbation
- large bias/small variance means few features, highly regularized, such as highly pruned decision trees, large-k kNN etc;
- While small bias/high variance means many features, less regularization, small-k k-NN etc.

Image from:
https://www.textbook.ds100.org/ch/15/bias_modeling.html

Derivation

$$
\begin{aligned}
\mathrm{E}\big[(y-\hat{f})^2\big] &= \mathrm{E}[y^2+\hat{f}^2-2y\hat{f}] \\
&= \mathrm{E}[y^2]+\mathrm{E}[\hat{f}^2]-\mathrm{E}[2y\hat{f}] \\
&= \mathrm{Var}[y]+\mathrm{E}[y]^2+\mathrm{Var}[\hat{f}]+\mathrm{E}[\hat{f}]^2-2f\mathrm{E}[\hat{f}] \\
&= \mathrm{Var}[y]+\mathrm{Var}[\hat{f}]+(f-\mathrm{E}[\hat{f}])^2 \\
&= \mathrm{Var}[y]+\mathrm{Var}[\hat{f}]+\mathrm{E}[f-\hat{f}]^2 \\
&= \sigma^2+\mathrm{Var}[\hat{f}]+\mathrm{Bias}[\hat{f}]^2
\end{aligned}
$$

Image from :
https://en.wikipedia-on-ipfs.org/wiki/Bias%E2%80%93variance_tradeoff.html

- Derivation in statistical terms:
  - Random variable
  - Variance and Expectation calculation

It seems that solving overfitting will always incur a bias variance trade off :(

Bagging for low bias high variance learners

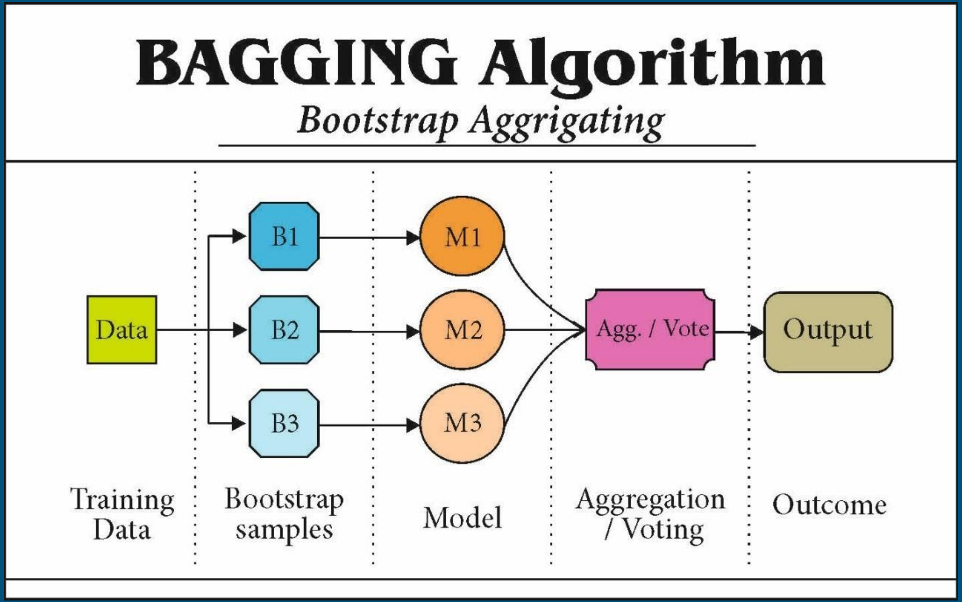
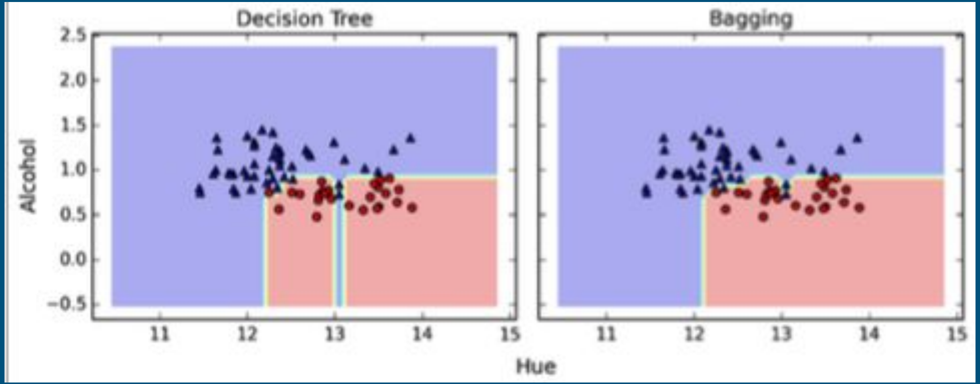


Image from:
https://analyticsindiamag.com/guide-to-ensemble-methods-bagging-vs-boosting/?utm_source=rss&utm_medium=rss&utm_campaign=guide-to-ensemble-methods-bagging-vs-boosting

Bagging Theory

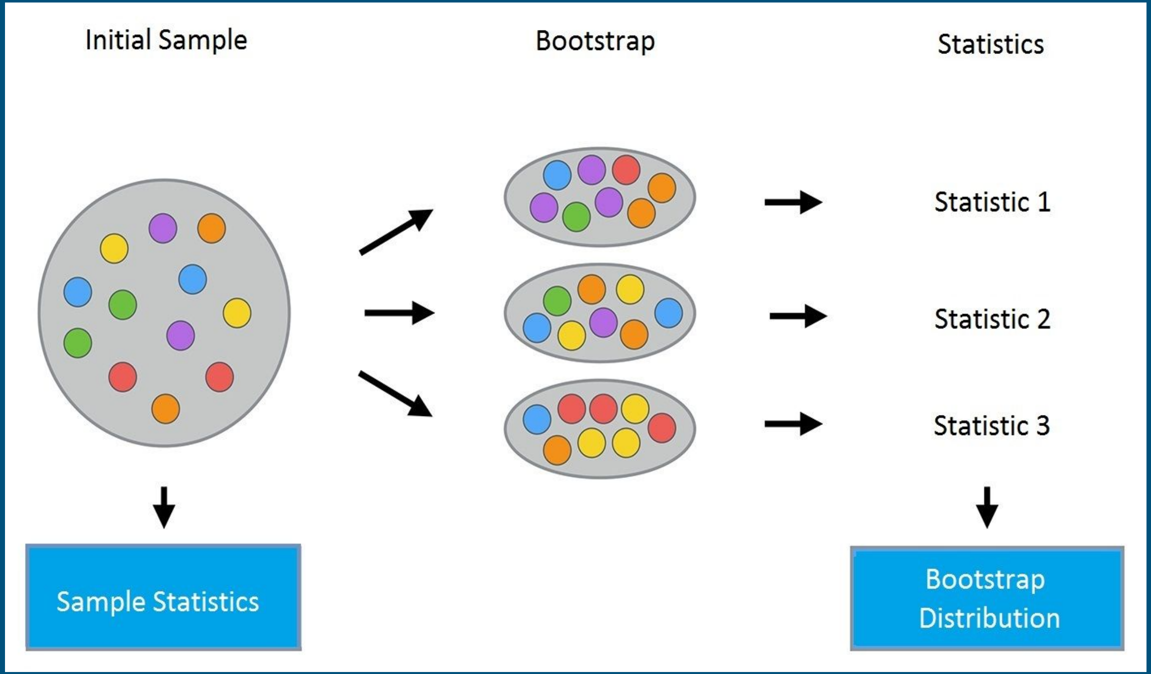
- Maintain bias reduce variance
- Suitable for low bias high variance learner



Bootstrap

- Train-Test Split
- Training set : used to fit model
- Test set: used to check generalization ability
- Validation set: used to evaluate the model we trained on the training set

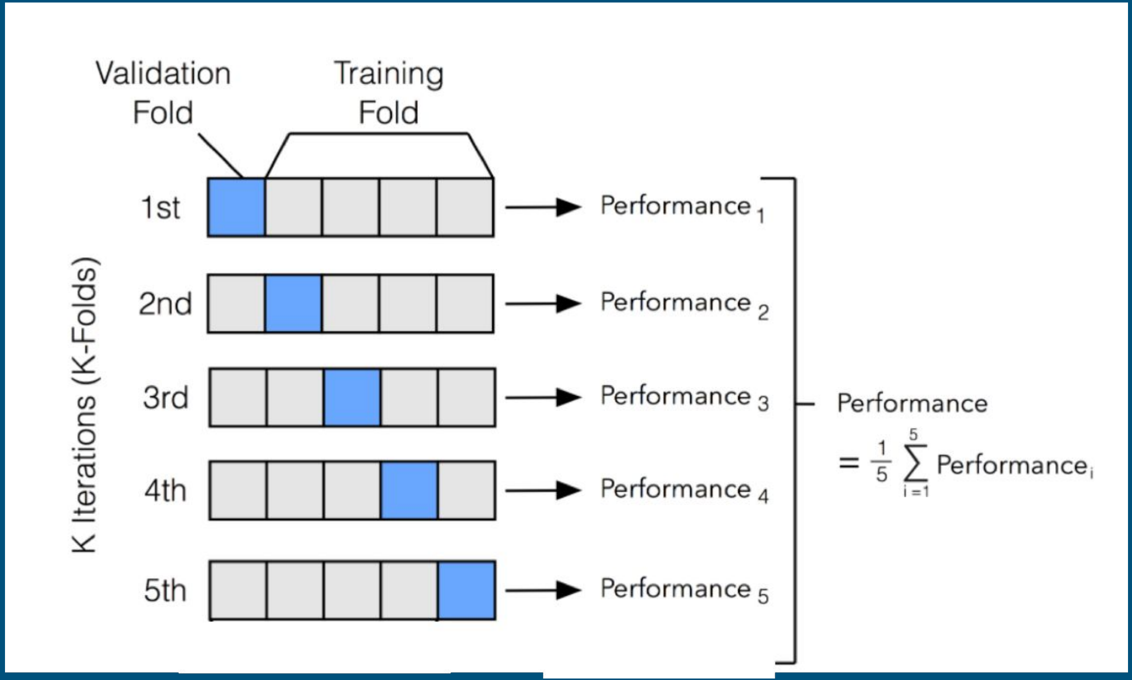
Bootstrap



- Treat sample as population
- Randomly select
- With replacement
- Avoid sample reduction

Image from: https://www.kaggle.com/kashnitsky/topic-5-ensembles-part-1-bagging

K-Fold Cross-Validation



- Train the model for Training Fold_1
- Use Validation Fold_1 to find Performance_1
- Repeat for 2...K folds
- Overall Performance is the average of each Performance_i

Image from :http://ethen8181.github.io/machine-learning/model_selection/model_selection.html

Decision Tree

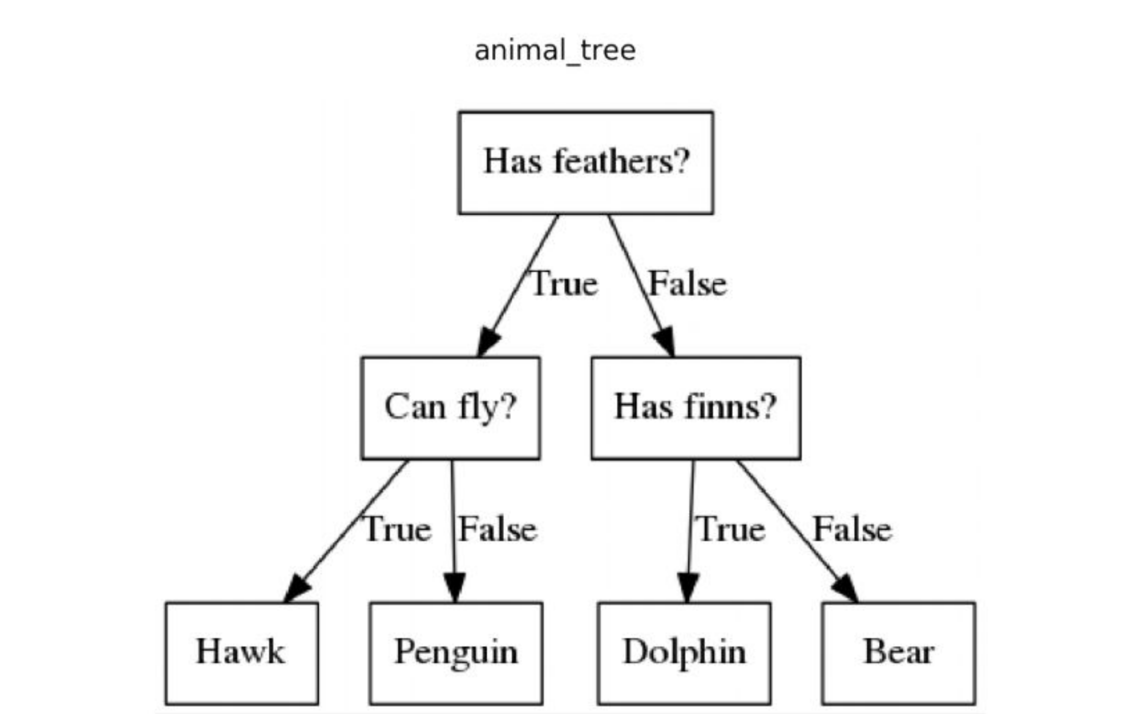
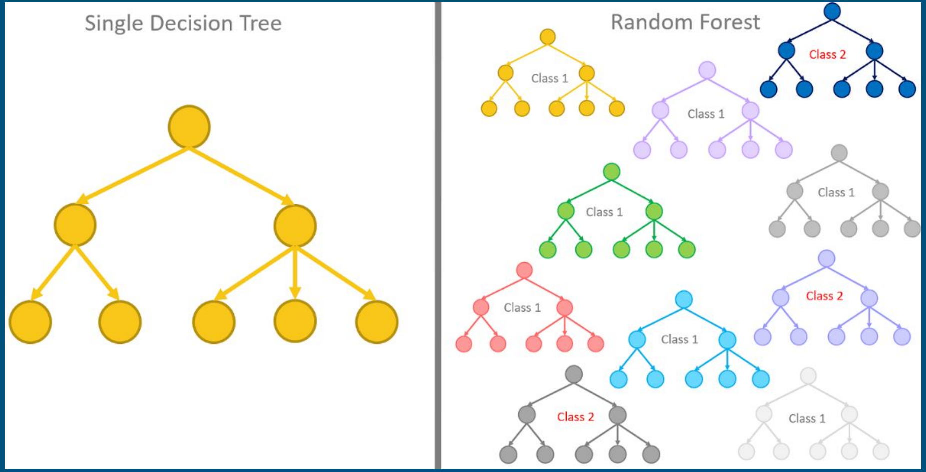


Image From:https://towardsdatascience.com/decision-tree-and-random-forest-explained-8d20ddabc9dd

- used for classification and regression problems
- answer sequential questions
- “If A, then B”

Random Forest

- An ensemble of many decision trees
- Each decision tree is used as parallel estimators
- It takes the mean value of the results from decision trees



End of Bagging

Thank you!