Bagging

Chenyue Cai, Oscar Xu, Yintang Yang

Ensemble methods:

- If you searched on google, the definition of ensemble is as following:
  - A group of musicians, actors, or dancers who perform together.
  - A group of items viewed as a whole rather than individuality.
- If you searched bagging on google scholar, the paper you find will say
  - Aggregated prediction based on multiple learners

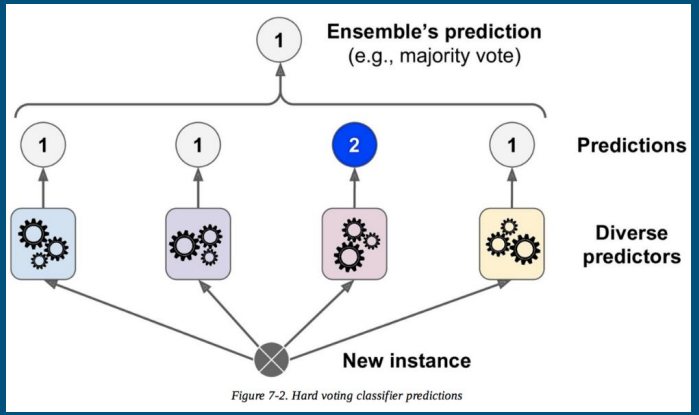


Figure 7-2. Hard voting classifier predictions

Image Source: https://images.app.goo.gl/Byx66JtKok4VQFoV8

Before all the math, and new concept, Let's identify something you know !

One issue that ML ppl always encounter :(

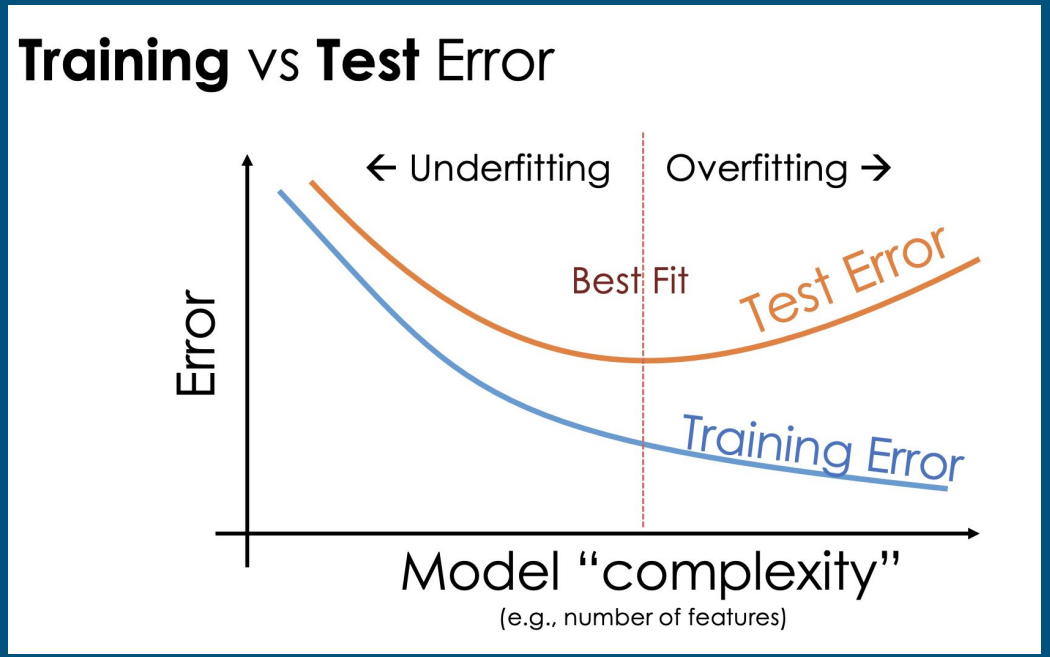


Image Source: http://www.ds100.org/fa19/resources/assets/lectures/lec17/17-FEandRegularization.pdf

Overfitting: A quick recap

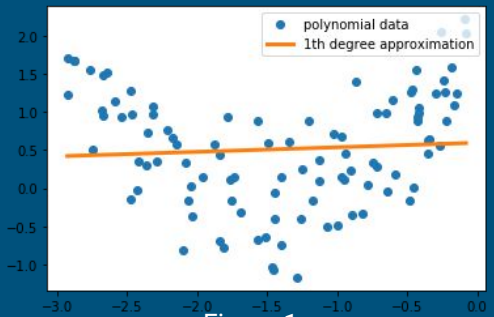


Figure 1

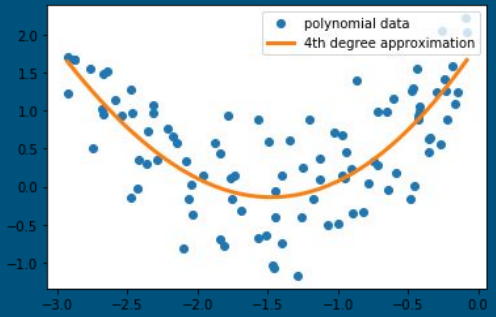


Figure 2

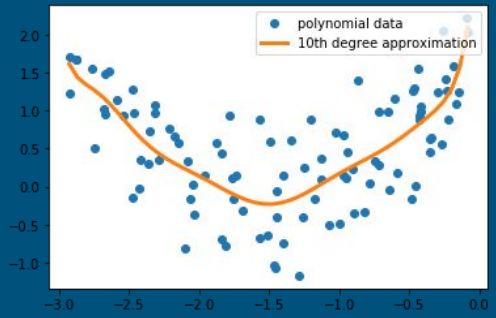


Figure 3

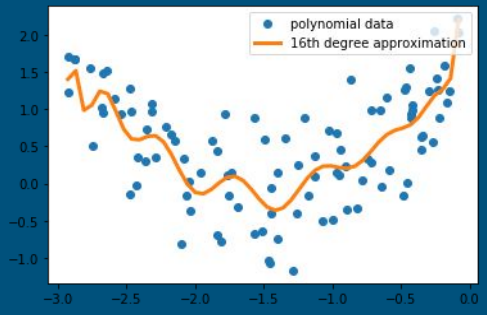


Figure 4

- Questions: given these line of best fit for different degree of polynomials identify the following:
  - Which one has the largest degree?
  - Which one has the smallest degree?
  - Which one is overfitting or underfitting?

How to solve Overfitting?

Image source : https://images.app.goo.gl/gcPhwoyDiERQygtbA

Recap: Interpretable metric of error Bias and Variance

- Bias Variance Definition
  - 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;
- small bias/high variance means many features, less regularization, small-k k-NN etc.

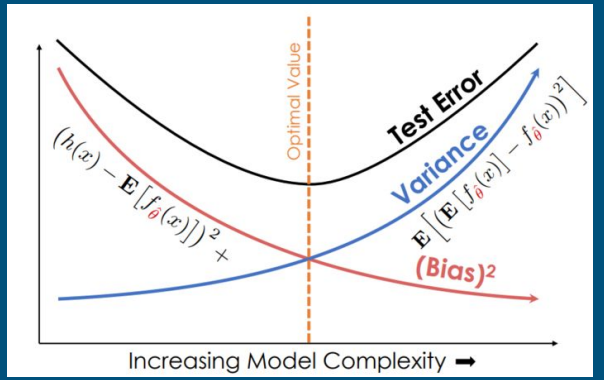


Image from: https://www.textbook.ds100.org/ch/15/bias_modeling.html

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

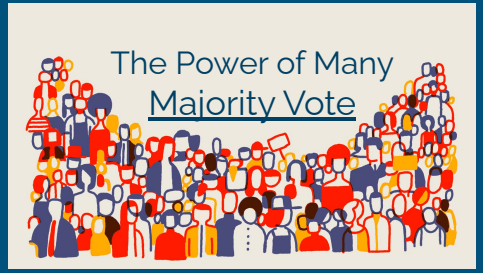
Bagging

"Our secret weapon against the inevitable bias variance trade-off"

- Single model: we are stuck with bias variance trade-off, reducing variance WILL HURT our model accuracy.
- Ensemble model: we can utilize the nature of combining multiple models to achieve our goal: keeping bias low while improving variance.
- In the following slides, we will answer the following questions:
  - How do we implement bagging?
  - Why does bagging work?
  - What does bagging improve?

Bagging: Implementation

- Comparing Bagging with Democracy
  - Dictatorship: Single Model Prediction
  - Democracy: Aggregate multiple learners result by a "Majority vote"
- "Democracy" Elements
  - "Voters": Different models that learns from a subsample of the whole training data
  - "Vote": Simply averages the prediction (in regression case), decide the results based on the most popular vote (in classification case)



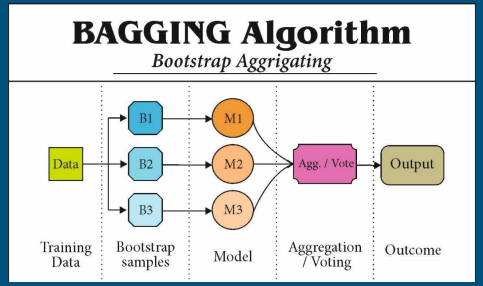


Image Source:
https://analyticsindiamag.com/guide-to-ensemble-methods-bagging-vs-boosting/?utm_source=rss&utm_medium=rss&utm_campaign=guide-to-ensemble-methods-bagging-vs-boosting

Bagging Implementation

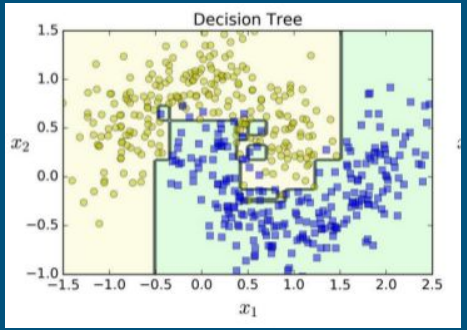
- Training:
  - Select a base learner model: M
  - Bootstrap from training data N samples
  - for i in range (N):
    - Train with i th sample and get a trained model M_i
- Predicting:
  - for i in range(N):
    - Get predicted result y_pred_i
  - Average out y_preds and attain the final result

Bagging Theory:

- Consider each model as a random variable Y, then we have Y1, Y2 ….. Yn. Each Y is independent of each other and have the same mean mu and variance sigma. We will explore the result of “majority vote”.
  - Problem set up: iid. Yi ~ (μ, σ^2)
  - Evaluate: the mean and variance of $1/n * \Sigma Y_i$
- Since it’s iid random variables of identical mean and variance, the mean is just μ, while the variance by linearity of variance is $1/n * \sigma^2$
- Thus, as n increases, the variance drops linearly while mean stays the same.

Bagging improvement:

- Maintain bias, Reduce variance
  - Thus suitable for low bias high variance learners
  - Can you name some of these learners on your own? (High degree polynomial)
- Smoothen out the decision boundary
  - We also see smoothen decision boundary when we do other stuff (e.g. ridge regression, Kernel RBF) It all deems to tackle overfitting!
- Increase test time accuracy



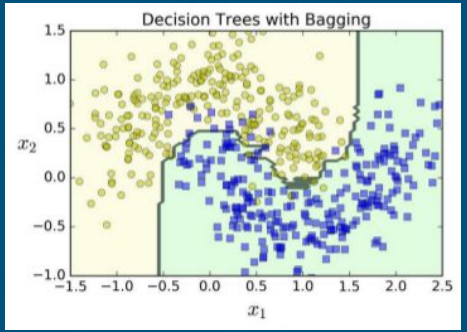


Image Source:
https://www.google.com/search?q=decision+boundary+bagging&tbm=isch&ved=2ahUKEwiZ4e6i2sTtAhWQhZ4KHbolCAYQ2-cCegQIABAA&oq=decision+boundary+bagging&gs_lcp=CgNpbWcQAzoECAAQEzoGCAAQCBAeOgQIABAYUPt9WKqFAWDChgFoAHAAeACAAV2IAZUFkgEBOJgBAKABAaoBC2d3cy13aXotaW1nwAEB&sclient=img&ei=t8HSX9mpBJCL-gS6y6Aw&hl=zh-CN#imgrc=zU22DM-AQvX3NM

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
- Common ratio: 70% train, 15% val, 15% test, or 80% train, 10% val, 10% test.
- Check sklearn.model_selection.train_test_split( ) function for coding details

Bootstrap

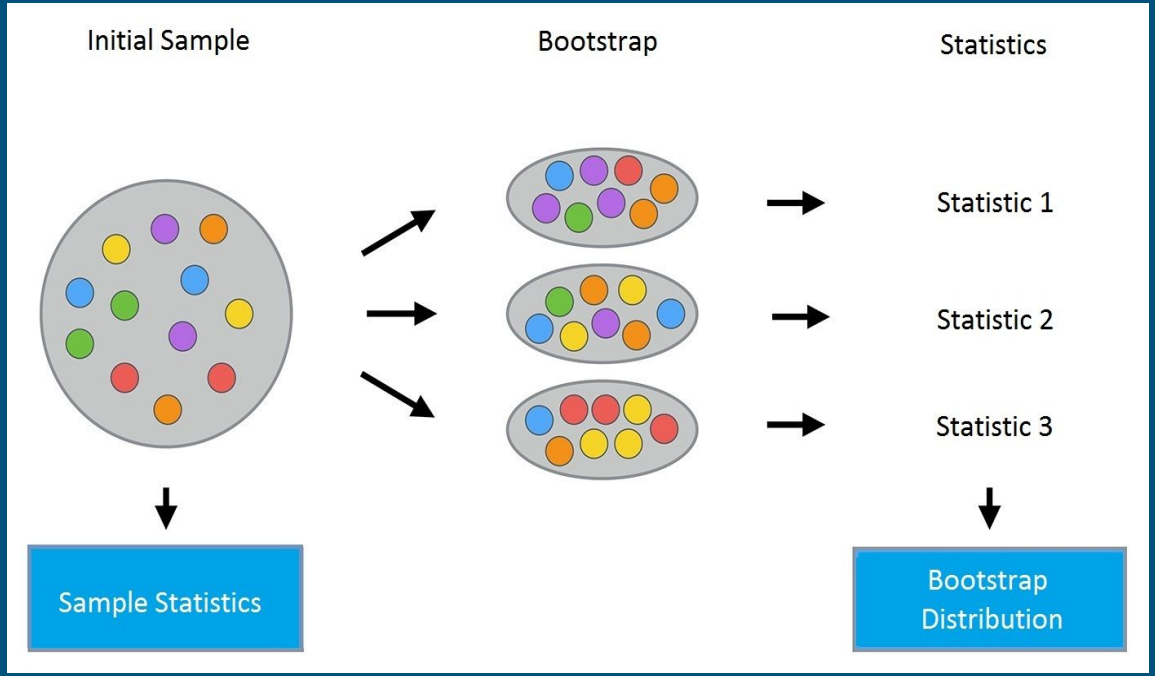
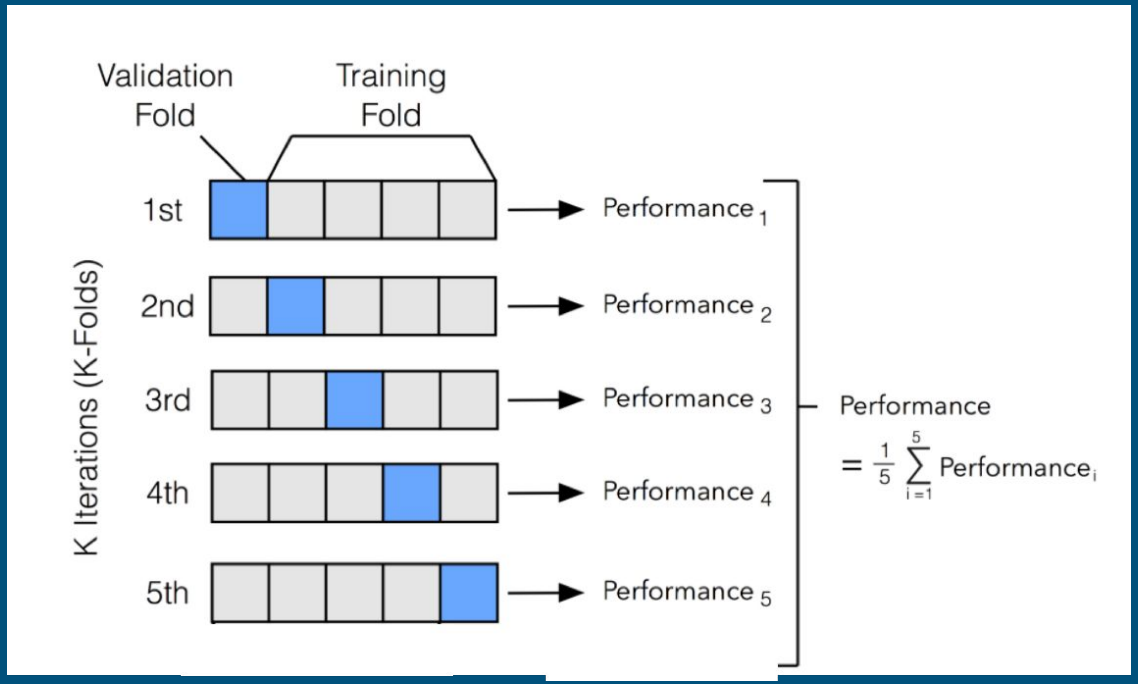


Image source: https://www.kaggle.com/kashnitsky/topic-5-ensembles-part-1-bagging

- Treat initial sample as population
- Randomly select
- With replacement
- Average statistics to get bootstrap distribution
- Avoid sample reduction
- Reduce variance
- Avoid overfitting

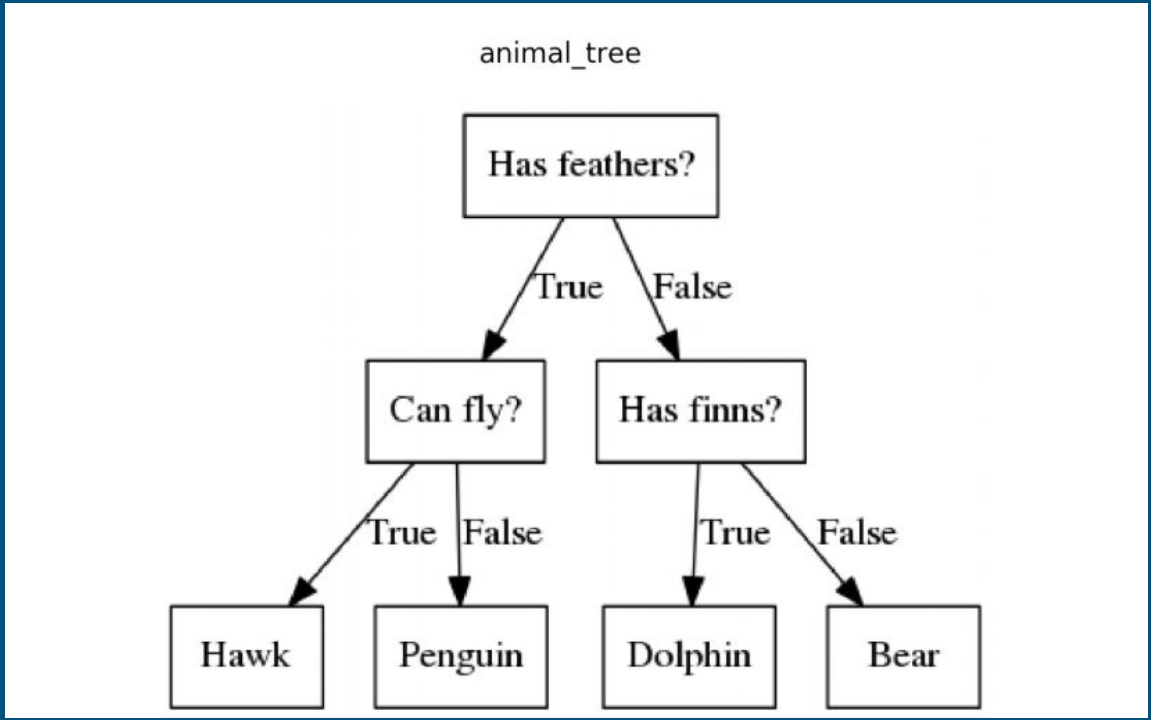
K-Fold Cross-Validation



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

Decision Tree



- Used for classification and regression problems
- Answer sequential questions
- “If A, then B”

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

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

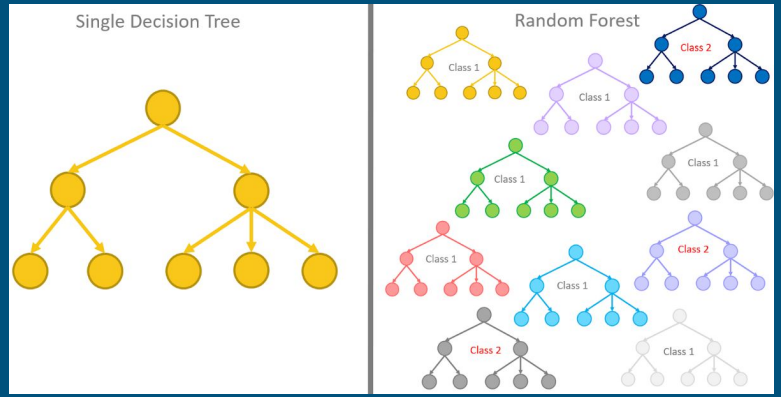


Image Source :
https://images.app.goo.gl/E51b5zT4WaJ2PGd18

Exercise

True or False: One example of bagging is random forest, it is used for classification. Thus bagging is a method specifically for classification problem but not regression problem.

Solution

False.

Bagging can be used for both classification and regression.

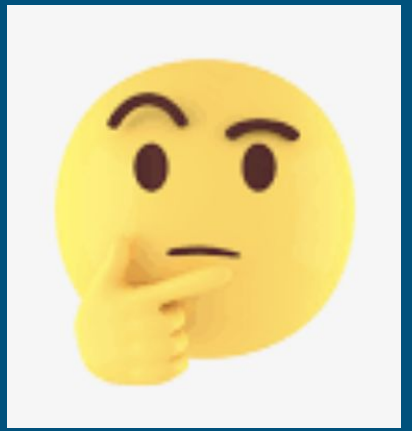


Image source:https://images.app.goo.gl/uDHKvcGYqP5pDV7E9

The End.
Thank you!