DATA ANALYTICS

Unit 4: Recommender Systems

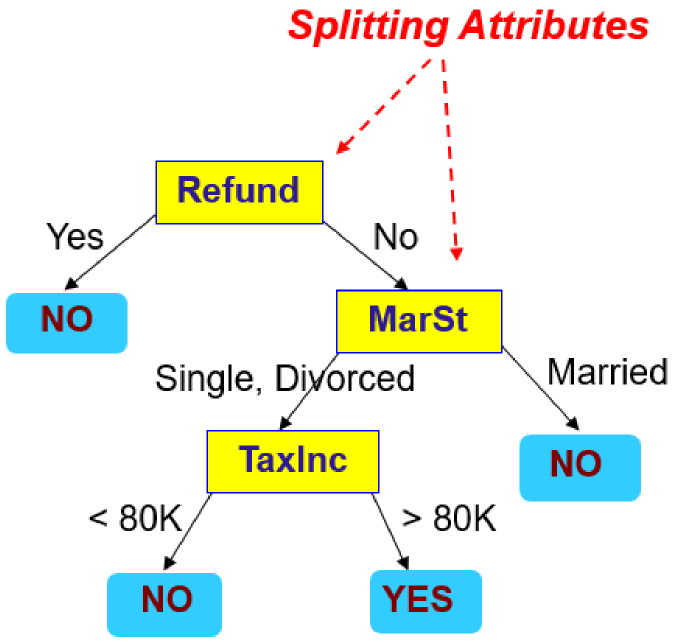
Jyothi R.
Assistant Professor
Department of Computer Science
and Engineering

Example of a Decision Tree

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data



Model: Decision Tree

Using the following condition

85k < Taxable Income <= 95k?

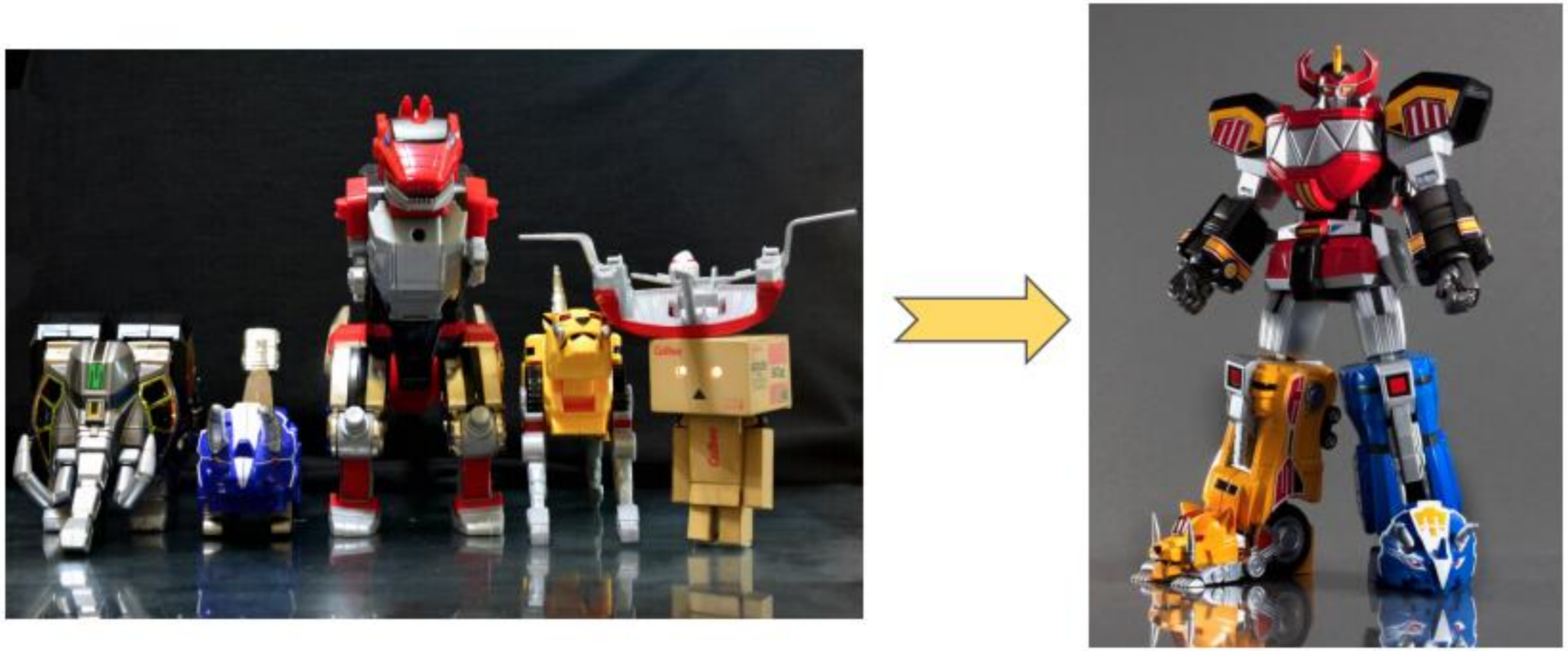
causes a multiway split.

We have three branches:

(1) Income < 85k
(2) Income between 85k-95k
(3) Income > 95k

Ensemble Methods

- General Idea Combine multiple classifiers usually with different strengths to build a bigger, better classifier



Ensemble Methods and Boosting

- The ensemble method is a machine-learning-algorithm that generates several classifiers using different sampling strategies such as bootstrap aggregating.

- A majority-voting-approach may be used for classifying a new observation using the multiple classifiers that are part of the ensemble method.

- In the ensemble method, several techniques such as logistic regression, CHAID, CART etc., are used.
- For a new observation, its class is identified using all the classifiers that are part of the ensemble method
- Different classifiers are likely to classify a new observation into different categories
- The final class of a new observation is decided based on a majority vote in which each classifier is given equal weightage

- Adaboosting: Boosting algorithms assign weights to each classifier based on their accuracy

Ensemble methods

Typical application: classification

Ensemble of classifiers is a set of classifiers whose individual decisions combined in some way to classify new examples.

1. Generate multiple classifiers
2. Each votes on test instance
3. Take majority as classification

Classifiers different due to different sampling of training data, or randomized parameters within the classification algorithm

Aim: Take a simple, mediocre algorithm and transform it into a super classifier without requiring any new or fancy algorithm.

Ensemble methods: Summary

Differ in training strategy, and combination method

1. Parallel training with different training sets: bagging

2. Sequential training, iteratively re-weighting training examples so current classifier focuses on hard examples: boosting

3. Parallel training with objective encouraging division of labor: mixture of experts

Notes:

- Also known as meta-learning
- Typically applied to weak models, such as decision stumps (single-node decision trees), or linear classifiers

Variance-bias tradeoff?

Minimize two sets of errors:

1. Variance: Error from sensitivity to small fluctuations in the training set

2. Bias: Erroneous assumptions in the model

Variance-bias decomposition: is a way of analyzing the generalization error as a sum of 3 terms: variance, bias and irreducible error (resulting from the problem itself)

Why do ensemble methods work?

Based on one of two basic observations:

1. Variance reduction: if the training sets are completely independent, it will always helps to average an ensemble because this will reduce variance without affecting bias (e.g., bagging) -- reduce sensitivity to individual data points.

2. Bias reduction: for simple models, average of models has much greater capacity than single model (e.g., hyperplane classifiers, Gaussian densities).

   Averaging models can reduce bias substantially by increasing capacity, and control variance by fitting one component at a time (e.g., boosting).

Ensemble methods: Justification

Ensemble methods more accurate than any individual members:

- Accurate (better than guessing)
- Diverse (different errors on new examples)

Independent errors: prob k of N classifiers

Ensemble methods: Netflix

- Clear demonstration of the power of ensemble methods
- Original prize winner (BellKor) used an ensemble of 107 models!

- “Our experience is that most efforts should be concentrated in deriving substantially different approaches, rather than refining a simple technique.”

- “We strongly believe that the success of an ensemble approach depends on the ability of its various predictors to expose different complementing aspects of the data. Experience shows that this is very different than optimizing the accuracy of each individual predictor.”

Bootstrap estimation

- Repeatedly draw n samples from D
- For each set of samples, estimate a statistic
- The bootstrap estimate is the mean of the individual estimates
- Used to estimate a statistic parameter and its variance
- Bagging: bootstrap aggregation

Bagging

- Simple idea: generate M bootstrap samples from your original training set. Train on each one to get y , and average them
- For regression: average predictions
- For classification: average class probabilities

  (or take the majority vote if only hard outputs available)
- Bagging approximates the Bayesian posterior mean. The more bootstraps the better, so use as many as you have time for Each bootstrap sample is drawn with replacement, so each one contains some duplicates of certain training points and leaves out other training points completely

Cross-validated committees

- Bagging works well for unstable algorithms

  can change dramatically with small changes in training data

- But can hurt a stable algorithm: a Bayes optimal algorithm may

  leave out some training examples in every bootstrap
- Alternative method based on different training examples:

  cross-validated committees:
- Here k disjoint subsets of data are left out of training sets
- Again uses majority for combination

Boosting

- Also works by manipulating training set, but classifiers trained sequentially
- Each classifier trained given knowledge of the performance of previously trained classifiers: focus on hard examples
- Final classifier: weighted sum of component classifiers

References

Text Book:

“Business Analytics, The Science of Data-Driven Making”, U. Dinesh Kumar, Wiley 2017 (Chapter 14)

“Recommender Systems, The text book, Charu C. Aggarwal, Springer 2016 Section 1 and Section 2.

Image Courtesy

http://webcache.googleusercontent.com/search?q=cache:vW5NL8dqVQkJ:www.cs.kent.edu/~jin/DM07/ClassificationDecisionTree.ppt+&cd=5&hl=en&ct=clnk&gl=in

https://cmci.colorado.edu/classes/INFO-4604/fa17/files/slides-16_ensemble.pdf

THANK YOU

Jyothi R.
Assistant Professor,
Department of Computer Science
jyothir@pes.edu