Fundamental Methods of Data Science

Class 6

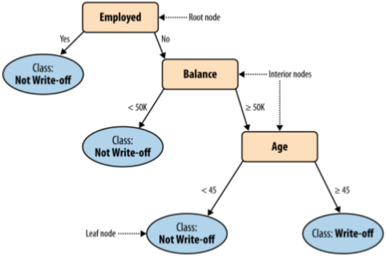
Tomer Libal

Tree-Structured Models

- In the previous class, you learned how to choose the most informative attributes
- Is it enough for creating a good model?

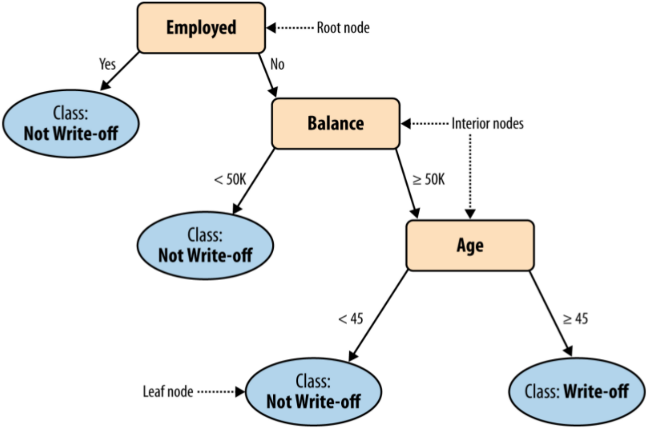
Tree-Structured Models

- In the previous class, you learned how to choose the most informative attributes
- Is it enough for creating a good model?



Tree-Structured Models

- Classify ‘John Doe’
  - Balance=115K, Employed=No, and Age=40



Tree-Structured Models: “Rules”

- No two parents share descendants
- There are no cycles
- The branches always “point downwards”
- Every example always ends up at a leaf node with some specific class determination
  - Probability estimation trees, regression trees (to be continued . . .)

Tree Induction

- How do we create a classification tree from data?

Tree Induction

- How do we create a classification tree from data?
  - **divide-and-conquer** approach
  - take each data subset and **recursively** apply attribute selection to find the best attribute to partition it

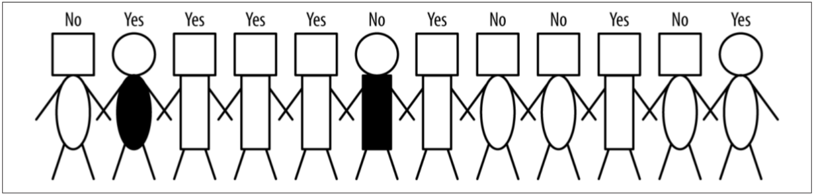
Tree Induction

- How do we create a classification tree from data?
  - **divide-and-conquer** approach
  - take each data subset and **recursively** apply attribute selection to find the best attribute to partition it
- When do we stop?

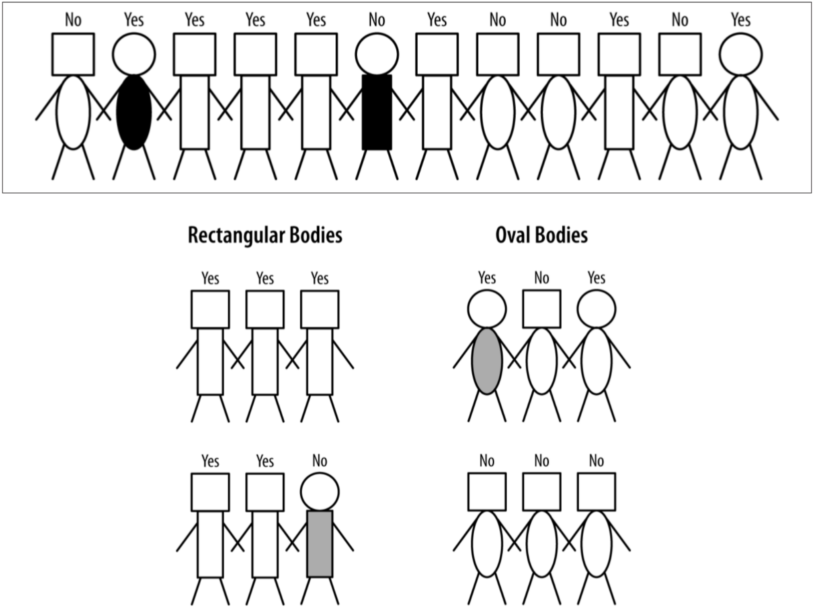
Tree Induction

- How do we create a classification tree from data?
  - **divide-and-conquer** approach
  - take each data subset and **recursively** apply attribute selection to find the best attribute to partition it
- When do we stop?
  - The nodes are pure, or
  - there are no more variables, or
  - even earlier (over-fitting, to be continued ...)

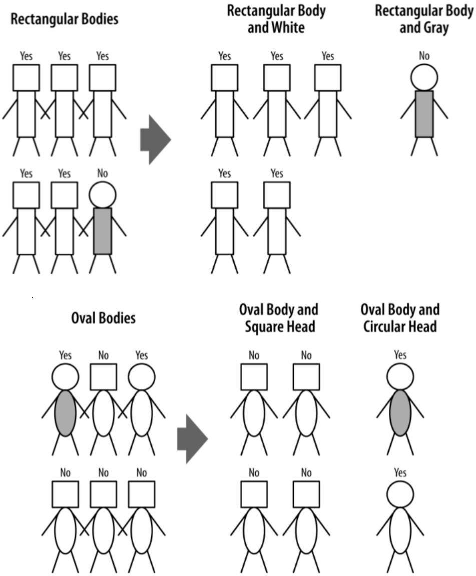
Example - Tree Induction



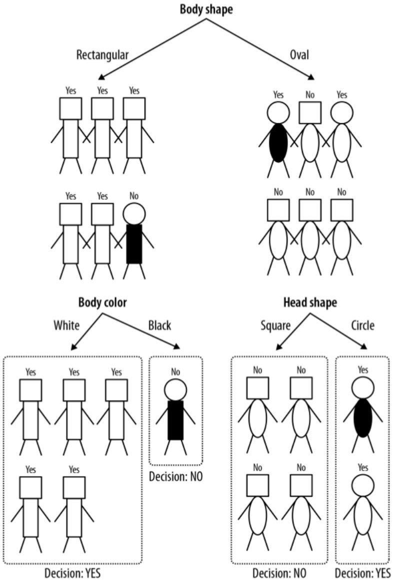
Example - Tree Induction



Example - Tree Induction



Example - Tree Induction



Why Trees?

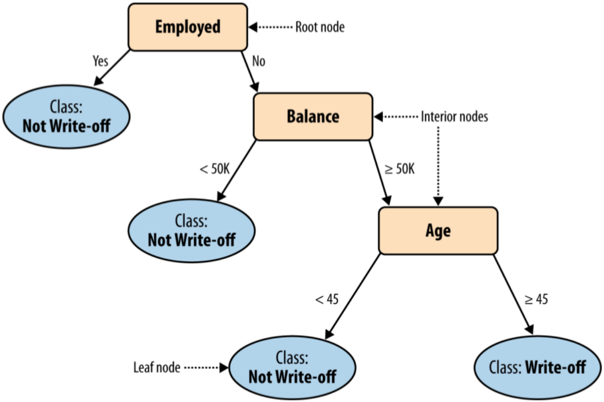
- Decision trees (DTs), or classification trees, are one of the most popular data mining tools
- They are:
  - Easy to understand
  - Easy to implement
  - Easy to use
  - Computationally cheap
- Almost all data mining packages include DTs
- They have advantages for model comprehensibility, which is important for:
  - model evaluation
  - communication to non-DM-savvy stakeholders

Trees as Sets of Rules

- The classification tree is equivalent to this rule set
  - Each rule consists of the attribute tests along the path connected with AND

Trees as Sets of Rules

- The classification tree is equivalent to this rule set
  - Each rule consists of the attribute tests along the path connected with AND



MegaTelCo: Predicting Customer Churn

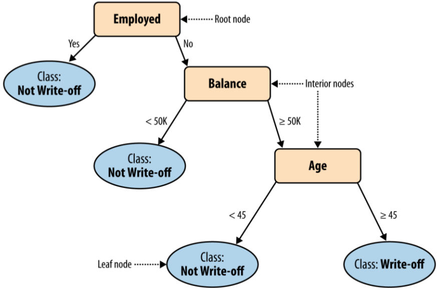
- Would a yes/no answer be enough? How can you improve over that?

MegaTelCo: Predicting Customer Churn

- Would a yes/no answer be enough? How can you improve over that?
  - MegaTelCo might want to rank customers according to their probability of leaving

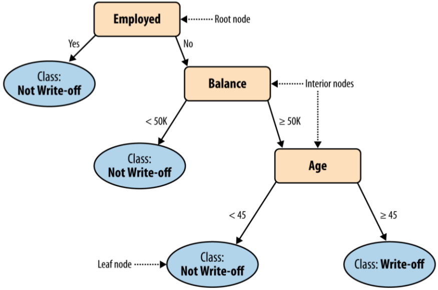
MegaTelCo: Predicting Customer Churn

- Would a yes/no answer be enough? How can you improve over that?
  - MegaTelCo might want to rank customers according to their probability of leaving



MegaTelCo: Predicting Customer Churn

- Would a yes/no answer be enough? How can you improve over that?
  - MegaTelCo might want to rank customers according to their probability of leaving



- How can you improve your answer?

From Classification Trees to Probability Estimation Trees

- Frequency-based estimate
  - Basic assumption: Each member of a segment corresponding to a tree leaf has the same probability to belong in the corresponding class
  - If a leaf contains `n` positive instances and `m` negative instances (binary classification), the probability of any new instance being positive may be estimated as $\frac{n}{n+m}$

From Classification Trees to Probability Estimation Trees

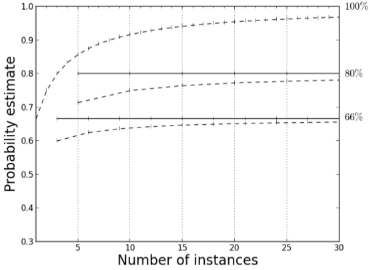
- Frequency-based estimate
  - Basic assumption: Each member of a segment corresponding to a tree leaf has the same probability to belong in the corresponding class
  - If a leaf contains `n` positive instances and `m` negative instances (binary classification), the probability of any new instance being positive may be estimated as $\frac{n}{n+m}$
- Prone to over-fitting, why?

Laplace Correction

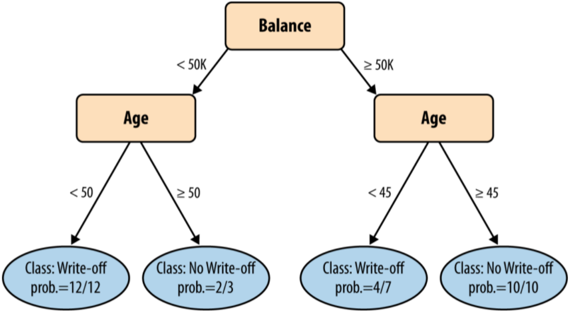
- $p(c) = \frac{n+1}{n+m+2}$
  - where `n` is the number of occurrences of `c` in the set and `m` is the number of all other occurrences

Laplace Correction

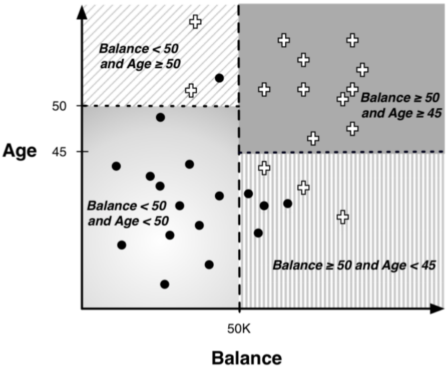
- $p(c) = \frac{n+1}{n+m+2}$
  - where `n` is the number of occurrences of `c` in the set and `m` is the number of all other occurrences



Visualizing Segmentations



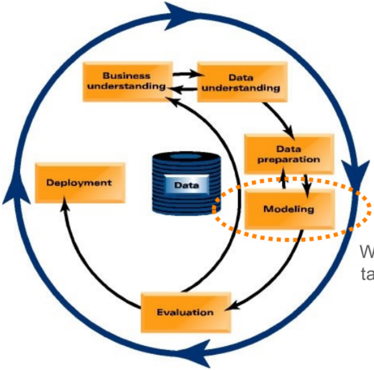
Visualizing Segmentations



The many faces of classification: Classification / Probability Estimation / Ranking

- Classification Problem
  - Most general case: The target takes on discrete values that are NOT ordered
  - Most common: binary classification where the target is either 0 or 1
- 2 Different Solutions to Classification
  - Classifier model: Model predicts the same set of discrete value as the data had
  - Probability estimation: Model predicts a score between 0 and 1 that is meant to be the probability of being in that class

Let's focus back in on actually mining the data..



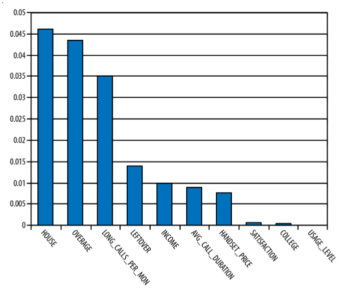
Which customers should TelCo target with a special offer, prior to contract expiration?

MegaTelCo: Predicting Churn with Tree Induction

Variable	Explanation
COLLEGE	Is the customer college educated?
INCOME	Annual income
OVERAGE	Average overcharges per month
LEFTOVER	Average number of leftover minutes per month
HOUSE	Estimated value of dwelling (from census tract)
HANDSET_PRICE	Cost of phone
LONG_CALLS_PER_MONTH	Average number of long calls (15 mins or over) per month
AVERAGE_CALL_DURATION	Average duration of a call
REPORTED_SATISFACTION	Reported level of satisfaction
REPORTED_USAGE_LEVEL	Self-reported usage level
LEAVE *(Target variable)*	Did the customer stay or leave (churn)?

- What is the first step?

MegaTelCo: Predicting Churn with Tree Induction



Rank	Info. gain	Attribute name
1	0.0461	HOUSE
2	0.0436	OVERAGE
3	0.0350	LONG_CALLS_PER_MON
4	0.0136	LEFTOVER
5	0.0101	INCOME
6	0.0089	AVG_CALL_DURATION
7	0.0076	HANDSET_PRICE
8	0.0003	SATISFACTION
9	0.000	COLLEGE
10	0.000	USAGE_LEVEL

- How will you build the tree?

MegaTelCo: Predicting Churn with Tree Induction