ABtree: An Algorithm For Subgroup-Based Treatment Assignment

Xiaofei (Susan) Wang

Lecturer/Research Scholar
Department of Statistics
Yale University

September 20, 2016

Joint work with Derek Feng.

Obama Campaign 2008

https://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/

Obama Campaign 2008

Existing button:

https://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/

Obama Campaign 2008

Existing button:

Other buttons to consider:

https://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/

Obama Campaign 2008

https://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/

Traditional A/B Testing

Consider two options A and B (e.g. buttons, lines of text, pictures)

Traditional A/B Testing

Consider two options A and B (e.g. buttons, lines of text, pictures)

Q: Which option A or B, when applied universally, maximizes some desired outcome (e.g. clickthrough rate, sign-up rate, purchase amount)?

Traditional A/B Testing

Consider two options A and B (e.g. buttons, lines of text, pictures)

Q: Which option A or B, when applied universally, maximizes some desired outcome (e.g. clickthrough rate, sign-up rate, purchase amount)?

- Run an experiment.

Traditional A/B Testing

Consider two options A and B (e.g. buttons, lines of text, pictures)

Q: Which option A or B, when applied universally, maximizes some desired outcome (e.g. clickthrough rate, sign-up rate, purchase amount)?

- Run an experiment.
- 2-sample t-test? 2-sample z-test?

Drawbacks of A/B Testing

1. Ignores other information that we have on individual users.

Drawbacks of A/B Testing

1. Ignores other information that we have on individual users.

VisitId	OS	Date/Time	Referrer	Checkout
340	Windows XP	9/18 5:03pm	Google	1
341	Windows 8.1	9/18 5:04pm	Google	0
342	Windows XP	9/18 5:04pm	(Direct)	0
343	OS X Yosemite	9/18 5:06pm	Google	1
344	OS X Yosemite	9/18 5:06pm	Yahoo	0
...				

Drawbacks of A/B Testing

1. Ignores other information that we have on individual users.

Checkout
1
0
0
1
0
...

Drawbacks of A/B Testing

1. Ignores other information that we have on individual users.

Checkout
1
0
0
1
0
...

2. Misses potential *subgroup effects*, in which subgroups of the population benefit from one version whereas others benefit from another.

Solution?

A (poor) solution: Divide up the population into different *market segments*.

Solution?

A (poor) solution: Divide up the population into different *market segments*.

Windows Segment:

VisitId	OS	Checkout
340	Windows XP	1
341	Windows 8.1	0
342	Windows XP	0
...		

Mac Segment:

VisitId	OS	Checkout
343	OS X Yosemite	1
344	OS X Yosemite	0
...		

Solution?

A (poor) solution: Divide up the population into different *market segments*.

Windows Segment:

Checkout
1
0
0
...

Mac Segment:

Checkout
1
0
...

Solution?

A (poor) solution: Divide up the population into different *market segments*.

Windows Segment:

Checkout
1
0
0
...

Mac Segment:

Checkout
1
0
...

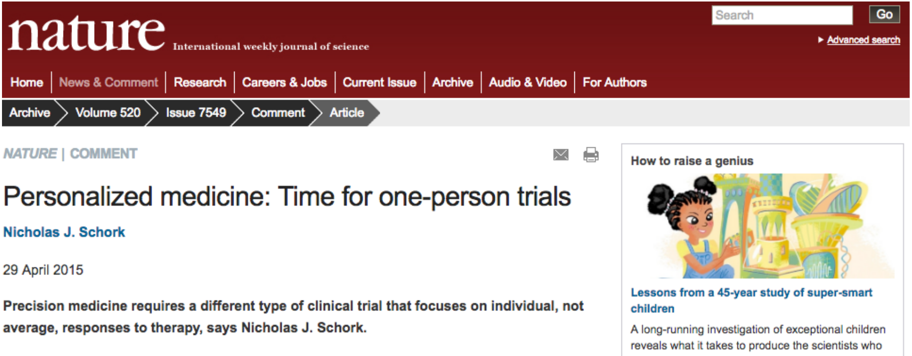
Better: **Automatically** detect *market segments* exhibiting differences in response to treatment.

Formal Question

- n individuals $i = 1, \ldots, n$
- k **treatments**
- individual i with covariates X_i gets randomly assigned to treatment T_i
- observe $Y_i | T_i$ (quantitative or binary)

Goal: For $i = 1, \ldots, n$, determine τ_i corresponding to the treatment which maximizes $E(Y_i | T_i = \tau_i)$.

Relationship to Precision Medicine



nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 520 > Issue 7549 > Comment > Article

NATURE | COMMENT

Personalized medicine: Time for one-person trials

Nicholas J. Schork

29 April 2015

Precision medicine requires a different type of clinical trial that focuses on individual, not average, responses to therapy, says Nicholas J. Schork.

How to raise a genius

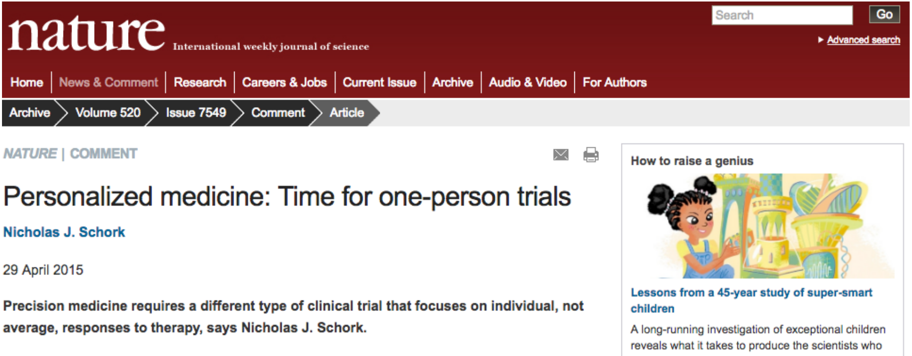
Lessons from a 45-year study of super-smart children

A long-running investigation of exceptional children reveals what it takes to produce the scientists who

---

http://www.nature.com/news/personalized-medicine-time-for-one-person-trials-1.17411

Relationship to Precision Medicine



"For instance, the drug Gleevec (imatinib) was found to double survival rates of leukaemia patients with a chromosomal abnormality in their tumours called the Philadelphia translocation."

---

http://www.nature.com/news/personalized-medicine-time-for-one-person-trials-1.17411

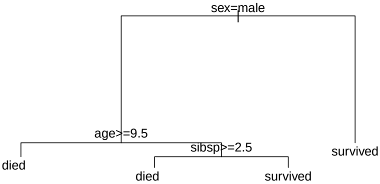
Trees

Classification and Regression Trees (CART)

[1] `ptitanic` dataset in `library(rpart.plot)`

Classification and Regression Trees (CART)

Titanic dataset[1]:



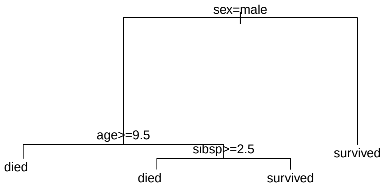
[1] ptitanic dataset in `library(rpart.plot)`

Classification and Regression Trees (CART)

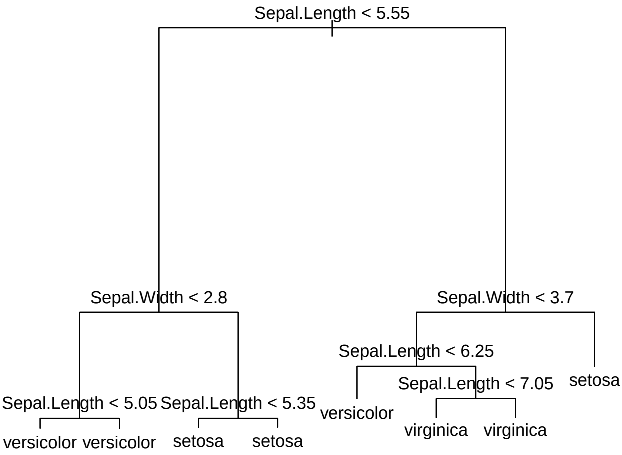
Breiman et al. 1984:

- Used to model a categorical or quantitative Y using predictors X where the relationships are possibly *non-linear*
- Easy to interpret
- Fast computation
- Extends to random forests

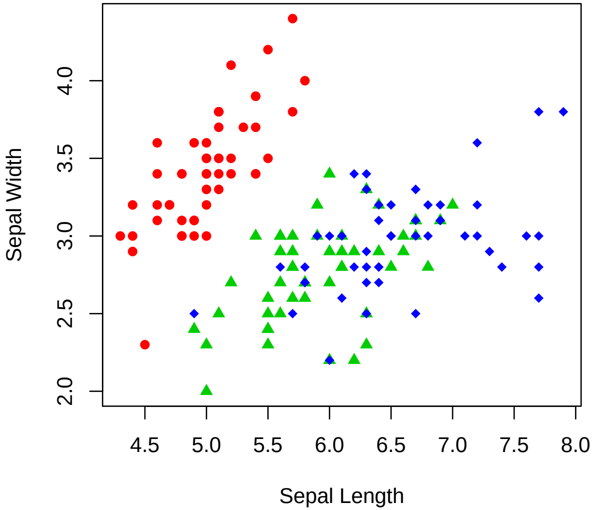
Titanic dataset:



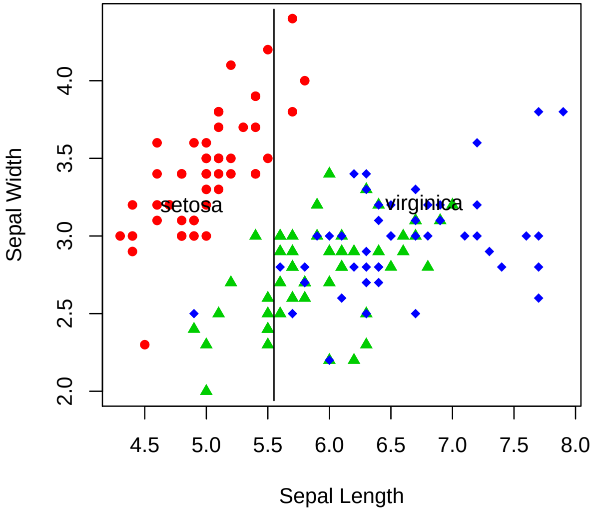
Classification and Regression Trees (CART)



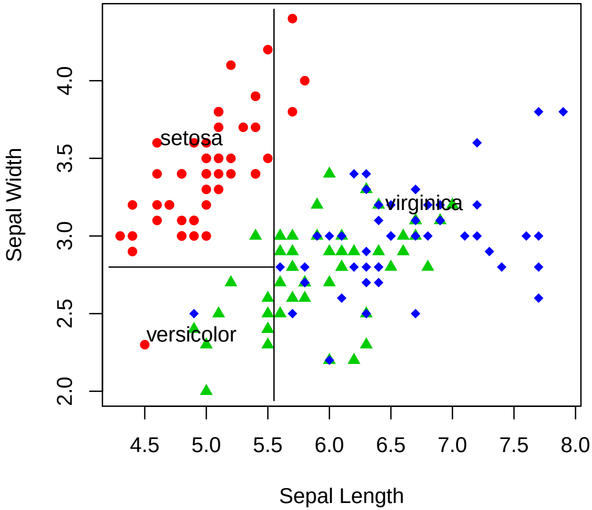
Classification and Regression Trees (CART)



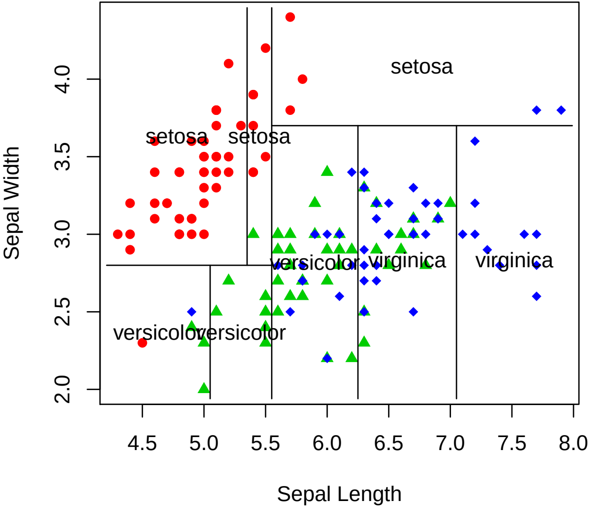
Classification and Regression Trees (CART)



Classification and Regression Trees (CART)



Classification and Regression Trees (CART)



Related Work (Using Trees)

- Interaction Trees (Su et al. 2009) - continuous Y
  - Maximize significance of interaction between **treatment** and **splits**.

Related Work (Using Trees)

- Interaction Trees (Su et al. 2009) - continuous Y
  - Maximize significance of interaction between **treatment** and **splits**.
- Virtual Twins (Foster et al. 2011) - binary Y
  - Use a random forest to estimate treatment effect Z:

$$Z = P(Y = 1 | T = 1) - P(Y = 1 | T = 0)$$

  - Use CART to estimate Z for all individuals.
  - Use arbitrary cutoff c such that leafs with $Z > c$ are flagged as subgroups.

Related Work (ct'd)

- SIDES (Lipkovich et al. 2011)
  - Multiple trees
  - Splitting criterion is a function of Z, test statistic for H_0 : treatment effect $= 0$ in subgroup, e.g.

$$|Z_L - Z_R|$$

  - Each split results in **good** and **bad** subgroup.
  - Only continue splitting **good** nodes.
  - Controls for overall Type I error rate.

Related Work (ct'd)

- SIDES (Lipkovich et al. 2011)
  - Multiple trees
  - Splitting criterion is a function of Z, test statistic for H_0 : treatment effect $= 0$ in subgroup, e.g.

$$|Z_L - Z_R|$$

  - Each split results in **good** and **bad** subgroup.
  - Only continue splitting **good** nodes.
  - Controls for overall Type I error rate.
  - Implemented in `SIDES` R package.
  - Numerous tuning parameters; slow.

Related Work (ct'd)

- QUINT (Dusseldorp & Van Mechelen 2013)
  - Splitting criterion is a weighted function of:
    - effect size
    - subgroup size
  - Requires the presence of subgroups where A beats B and subgroups where B beats A.

Related Work (ct'd)

- QUINT (Dusseldorp & Van Mechelen 2013)
  - Splitting criterion is a weighted function of:
    - effect size
    - subgroup size
  - Requires the presence of subgroups where A beats B and subgroups where B beats A.
  - Implemented in `quint` R package. Buggy.

ABtree: Splitting

Tree growth procedure uses binary recursive partitioning into subgroups $S_1, S_2, \ldots$ to **maximize**:

$$\sum_{S_j} Q(S_j) \tag{1}$$

for some measure Q.

What should we use for Q?

ABtree: Idea #1

- Goal is **profit maximization**

ABtree: Idea #1

- Goal is **profit maximization**
- Q should be a function of the *expected profit derived from the i-th individual conditional on receiving treatment t*

ABtree: Idea #1

An example:

$\bar{Y}_A$: 68
(n_A : 41)
$\bar{Y}_B$: 45
(n_B : 20)

- Total profit $= 68(41) + 45(20) = 3,688$.
- A is more profitable in this subgroup.
- Expect to gain $(68 - 45) \times 20 = 460$ in profit if everyone in this subgroup was assigned to A.

ABtree: Idea #1

An example:

A
$\bar{Y}_A$: 68
(n_A : 41)
$\bar{Y}_B$: 45
(n_B : 20)

- Total profit $= 68(41) + 45(20) = 3,688$.
- A is more profitable in this subgroup. Assign treatment A.
- Expect to gain $(68 - 45) \times 20 = 460$ in profit if everyone in this subgroup was assigned to A.

Implied splitting criterion:

$$Q_{max}(S_j) := |S_j| \max_t \bar{y}_{j|t}$$

where S_j is the j-th subgroup and $\bar{y}_{j|t}$ is the average profit of individuals in this subgroup receiving treatment t.

ABtree: Idea #2

L_2 Maximization:

- Choose splits that maximize total squared distance between best average profit $\max_t \bar{y}_{j|t}$ and average profit of other treatments $\bar{y}_{j|t}$:

$$Q_{L_2} := |S_j| \sum_{t'} (\max_t \bar{y}_{j|t} - \bar{y}_{j|t'})^2.$$

ABtree: Idea #2

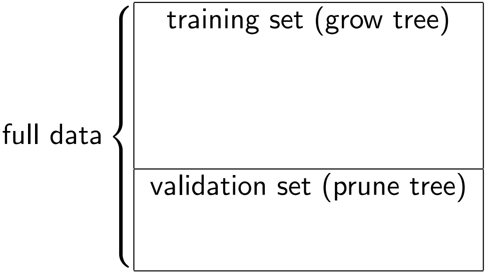
L_2 Maximization:

- Choose splits that maximize total squared distance between best average profit $\max_t \bar{y}_{j|t}$ and average profit of other treatments $\bar{y}_{j|t}$:

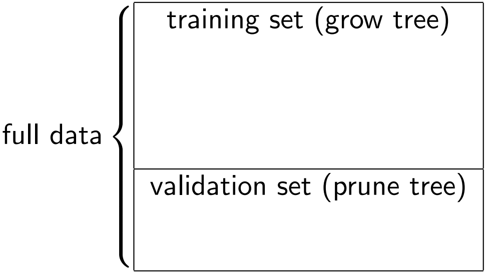
$$Q_{L_2} := |S_j| \sum_{t'} (\max_t \bar{y}_{j|t} - \bar{y}_{j|t'})^2.$$

This measure outperforms Q_{max} in simulation settings.

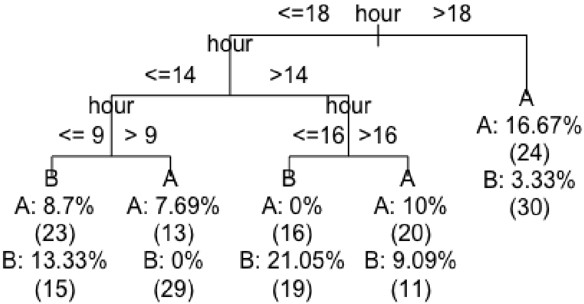
ABtree Pruning



- Pick subtree that does well

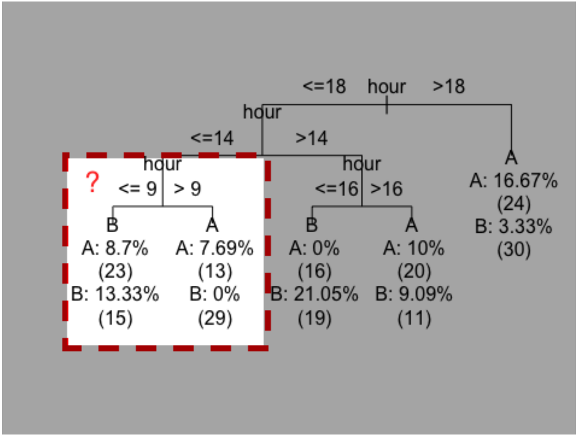


- Pick subtree that does well
- e.g. maximizes $\sum_{S_j} Q(S_j)$ in *validation set*



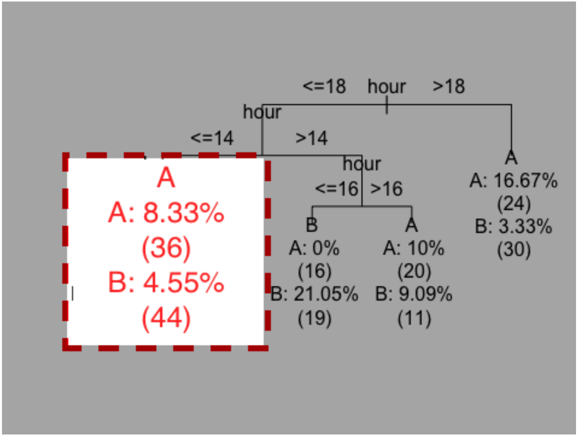
<=18 hour >18
hour
<=14
>14
hour
<= 9
> 9
hour
<=16
>16
B
A: 8.7%
(23)
B: 13.33%
(15)
A
A: 7.69%
(13)
B: 0%
(29)
B
A: 0%
(16)
B: 21.05%
(19)
A
A: 10%
(20)
B: 9.09%
(11)
A
A: 16.67%
(24)
B: 3.33%
(30)

ABtree Pruning



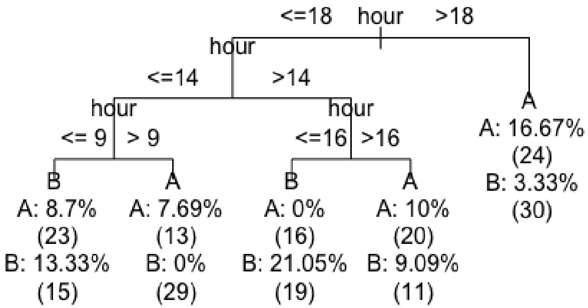
$$\sum Q_{L_2}(S_j) = (0.13 - 0.08)^2(23 + 15) + \cdots + (0.17 - 0.03)^2(24 + 30)$$
$$= 2.84$$

ABtree Pruning



$$\sum Q_{L_2}(S_j) = (0.08 - 0.05)^2(36 + 44) + \cdots + (0.17 - 0.03)^2(24 + 30)$$
$$= 2.63$$

ABtree Pruning



Since $2.84 > 2.63$, we do not prune this branch!

Example: National Supported Work Study (LaLonde, 1986)

- 1970's large-scale national and private program designed to provide work experience for $n = 722$ disadvantaged workers
- Treatments: A (control) and B (receiving benefits to improve employability)
- Sample sizes: $n_A = 425$ and $n_B = 297$
- Y (binary): whether individual earnings increased after the completion of experiment

Example: National Supported Work Study (LaLonde, 1986)

Covariates X:

Name	Description	Type
age	age (yrs)	quantitative
educ	education (yrs)	quantitative
race	(black, hispanic, white)	categorical
marr	married flag	categorical
nodegr	no degree flag	categorical
log.income75	log income in 1975	quantitative
u75	unemployment flag in 1975	categorical

Example: National Supported Work Study (LaLonde, 1986)

Assessment:

training set $n = 500$ (grow tree)
validation set $n = 150$ (prune tree)
test set $n = 72$ (assessment)

Example: National Supported Work Study (LaLonde, 1986)

Assessment:

training set $n = 500$ (grow tree)
validation set $n = 150$ (prune tree)
test set $n = 72$ (assessment)

2 groups in **test set**:

- received recommended treatment (**match**)
- did not receive recommended treatment (**no match**)

Example: National Supported Work Study (LaLonde, 1986)

Assessment:

training set $n = 500$ (grow tree)
validation set $n = 150$ (prune tree)
test set $n = 72$ (assessment)

2 groups in **test set**:

- received recommended treatment (**match**)
- did not receive recommended treatment (**no match**)

	match	no match
$y = 1$	d	e
$y = 0$	f	g

Example: National Supported Work Study (LaLonde, 1986)

Assessment:

training set $n = 500$ (grow tree)
validation set $n = 150$ (prune tree)
test set $n = 72$ (assessment)

2 groups in **test set**:

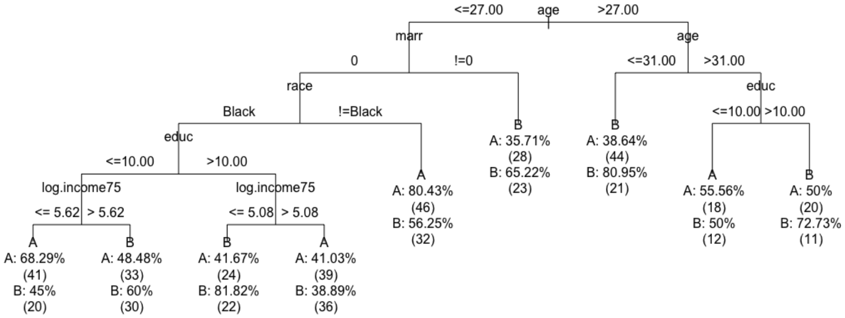
- received recommended treatment (**match**)
- did not receive recommended treatment (**no match**)

	match	no match
$y = 1$	d	e
$y = 0$	f	g

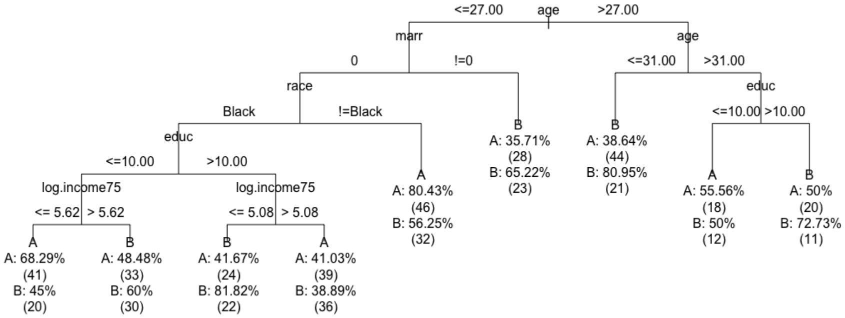
Tree works well if:

$$\frac{d}{d+f} > \frac{e}{e+g}$$

Example: National Supported Work Study (LaLonde, 1986)



Example: National Supported Work Study (LaLonde, 1986)



	match	no match
$y = 1$	28	21
$y = 0$	10	13
$\%success$	73.7%	61.8%

Summary

- ABtree supports analysis on **2 or more treatments**.

Summary

- ABtree supports analysis on **2 or more treatments**.
- The response Y can be binary or quantitative.

Summary

- ABtree supports analysis on **2 or more treatments**.
- The response Y can be binary or quantitative.
- A/B testing uses large sample sizes (in contrast to clinical trials).

Summary

- ABtree supports analysis on **2 or more treatments**.
- The response Y can be binary or quantitative.
- A/B testing uses large sample sizes (in contrast to clinical trials).
- As with most tree-based algorithms, splits obtained are sensitive to the underlying data.

Summary

- ABtree supports analysis on **2 or more treatments**.
- The response Y can be binary or quantitative.
- A/B testing uses large sample sizes (in contrast to clinical trials).
- As with most tree-based algorithms, splits obtained are sensitive to the underlying data.
- If only interested *treatment assignment* (rather than a sequence of *assignment rules*, we can use **random forests** to improve accuracy.

Summary

- ABtree supports analysis on **2 or more treatments**.
- The response Y can be binary or quantitative.
- A/B testing uses large sample sizes (in contrast to clinical trials).
- As with most tree-based algorithms, splits obtained are sensitive to the underlying data.
- If only interested *treatment assignment* (rather than a sequence of *assignment rules*, we can use **random forests** to improve accuracy.

Summary

- ABtree supports analysis on **2 or more treatments**.
- The response Y can be binary or quantitative.
- A/B testing uses large sample sizes (in contrast to clinical trials).
- As with most tree-based algorithms, splits obtained are sensitive to the underlying data.
- If only interested *treatment assignment* (rather than a sequence of *assignment rules*, we can use **random forests** to improve accuracy.

Thank You!

References

- Su, Xiaogang, et al. "Subgroup analysis via recursive partitioning." *Journal of Machine Learning Research* 10. Feb (2009): 141-158.
- Breiman, Leo, et al. *Classification and regression trees*. CRC press, 1984.
- LaLonde, R.J. 1986. "Evaluating the econometric evaulations of training programs with experimental data." *American Economic Review*, Vol.76, No.4, pp. 604-620.
- Lipkovich, Ilya, et al. "Subgroup identification based on differential effect searcha recursive partitioning method for establishing response to treatment in patient subpopulations." *Statistics in medicine* 30.21 (2011): 2601-2621.
- Foster, Jared C., Jeremy MG Taylor, and Stephen J. Ruberg. "Subgroup identification from randomized clinical trial data." *Statistics in medicine* 30.24 (2011): 2867-2880.
- Dusseldorp, Elise, and Iven Van Mechelen. "Qualitative interaction trees: a tool to identify qualitative treatment-subgroup interactions." *Statistics in medicine* 33.2 (2014): 219-237.