Primer on Machine Learning

attilagk

Lazy8

Machine Learning is Everywhere

- pattern detection and recognition (iphone touch ID, face ID, word autocomplete, speech to text)
- history based recommendation (youtube, facebook, google search, amazon,...)
  - products for customers
  - customers for providers
- email filtering and classification (gmail)

Machine Learning and Artificial Intelligence

1950s	*[getting] machines to exhibit behavior, which if done by humans, would be assumed to involve the use of intelligence*[1]
now	*computational methods to automatically learn and to improve with experience*[2]

ML[3]	statistical ("statistical learning")
AI	analytical (knowledge, logic)

[1] Arthur Samuel, 1983

[2] http://www.mlplatform.nl/what-is-machine-learning/

[3] may mean Maximum Likelihood: abbrev. not widely used

Machine Learning Now

- big data
  - data science, data mining, ...
  - myth: machine learning needs big data[4]
- fast computers
- emerging new methods
  - deep learning, reinfocement learning, ...

[4] see NY/SF homes toy data

General Outline

1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

General Outline

1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

skill set	
EN	expert knowledge
ST	statistics
PR	programming
DSS	domain spec. software
HPC	high perf. comp.
LIB	ML libraries[a]
DOC	LaTeX, Web

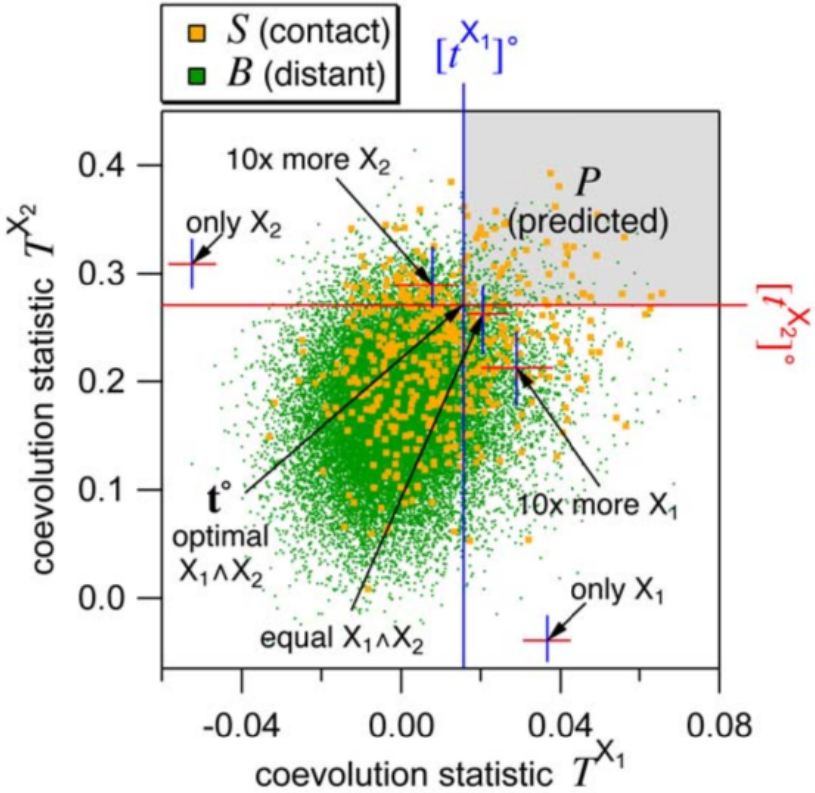
[a] Python, R, Java, Julia, Scala, Mat

My Story with Machine Learning[5]

Myth: it's like cooking

skill	2006	2017
expert knowledge	?	?
statistics	-	+
programming	-	+
domain spec. softw.	-	+
high perform. comp.	-	?
ML libraries	-	?
LaTeX, Web.	-	+

PLoS One. 2012;7(5):e36546.



[5] http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036546

General Outline

1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

The "Home" Data

Useless except for demonstration

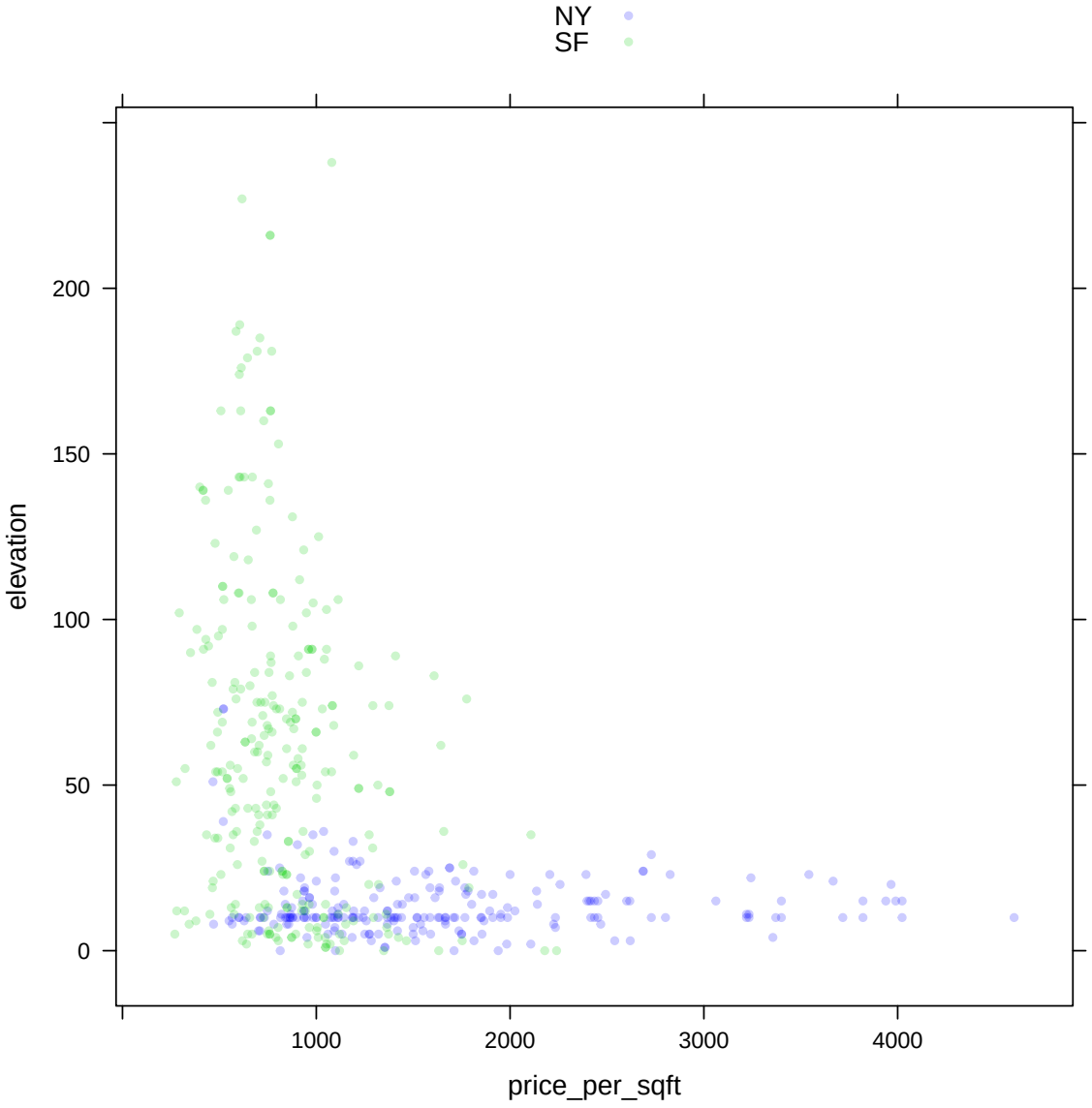
observation	input features / variables				output
i	x_{i1}	x_{i2}	...	x_{ip}	y_i
home	price/sqft	elevation	...	beds	city
training data					
1	999	10	...	2	NY
2	1939	0	...	2	NY
⋮	⋮	⋮	⋮	⋮	⋮
491	764	163	...	1	SF
492	762	216	...	3	SF
test data					
493	1196	40	...	2	?
⋮	⋮	⋮	⋮	⋮	⋮

General Outline

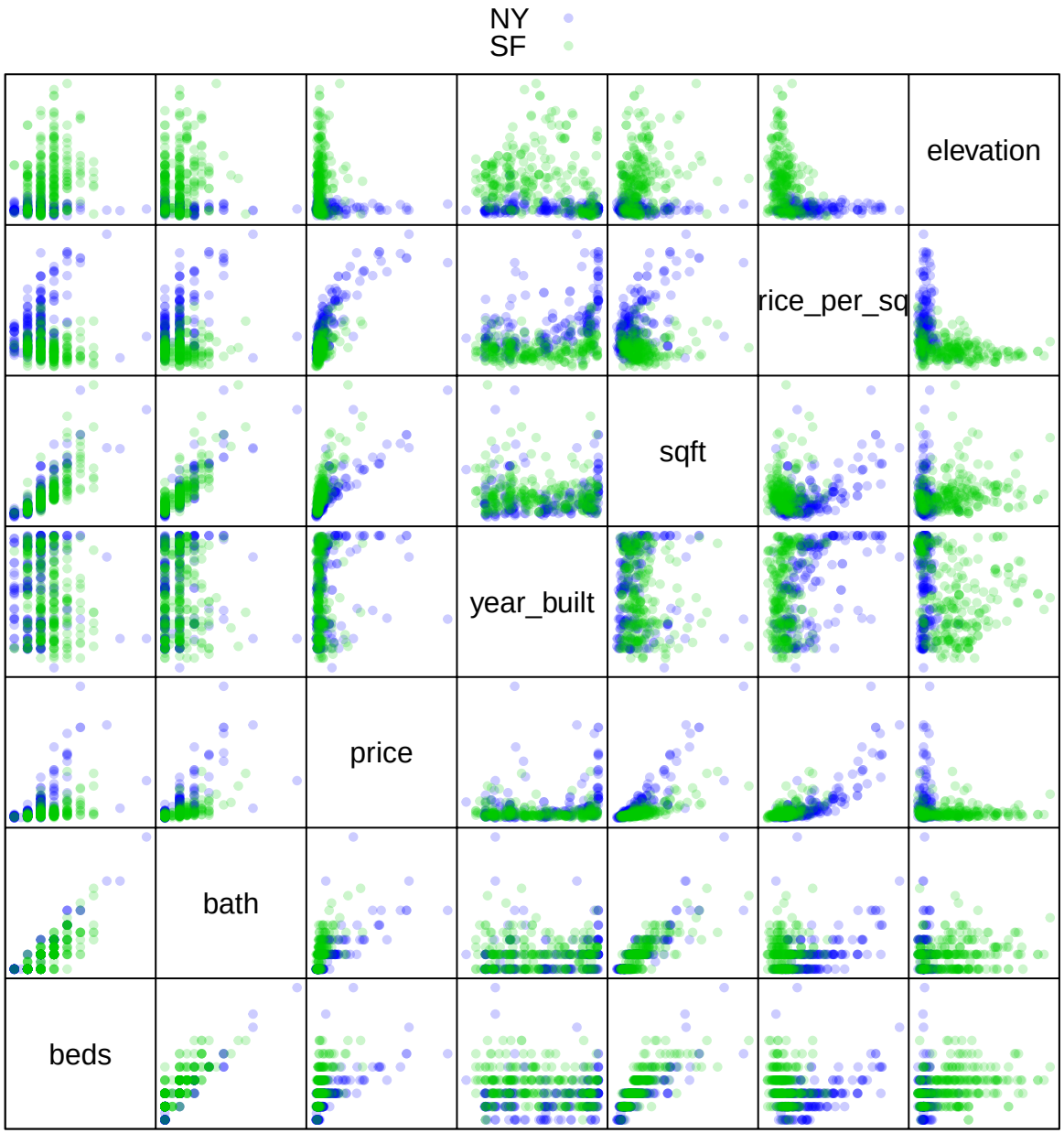
1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

Inspecting Dependencies

2 input features: 2D plots



All Inputs



Scatter Plot Matrix

General Outline

1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

The "Home" Data

Useless except for demonstration

observation	input features / variables				output
i	x_{i1}	x_{i2}	...	x_{ip}	y_i
home	price/sqft	elevation	...	beds	city
training data					
1	999	10	...	2	NY
2	1939	0	...	2	NY
⋮	⋮	⋮	⋮	⋮	⋮
491	764	163	...	1	SF
492	762	216	...	3	SF
test data					
493	1196	40	...	2	?
⋮	⋮	⋮	⋮	⋮	⋮

Tasks

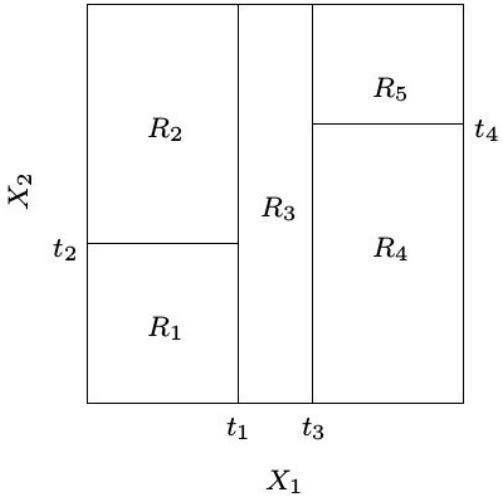
1. supervised learning: training *and* test data
   - prediction, classification
     - pattern recognition
     - business, medical, ... predictions & decisions
2. unsupervised learning: *only* training data
   - structure discovery
     - social, biol., tech. networks, associations,...
     - probabilistic expert systems
   - hypothesis testing, feature subset selection
     - research, marketing
   - matrix completion (imputation)
     - recommendation systems

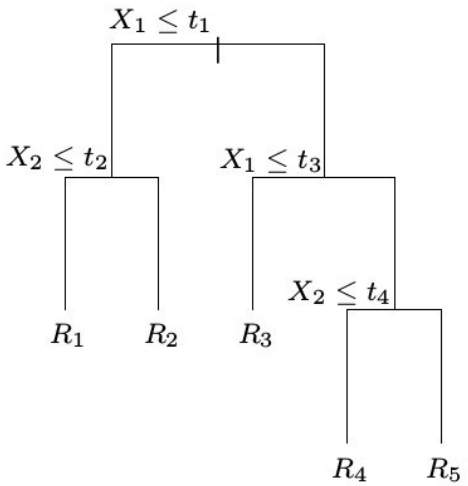
General Outline

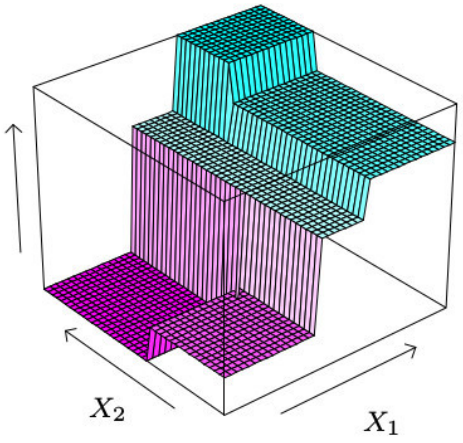
1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

Decision Tree is a Simple Model for Classification

A.k.a. CART: Classification And Regression Tree[6]



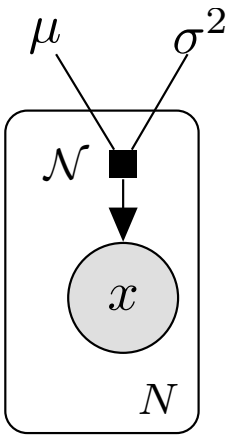


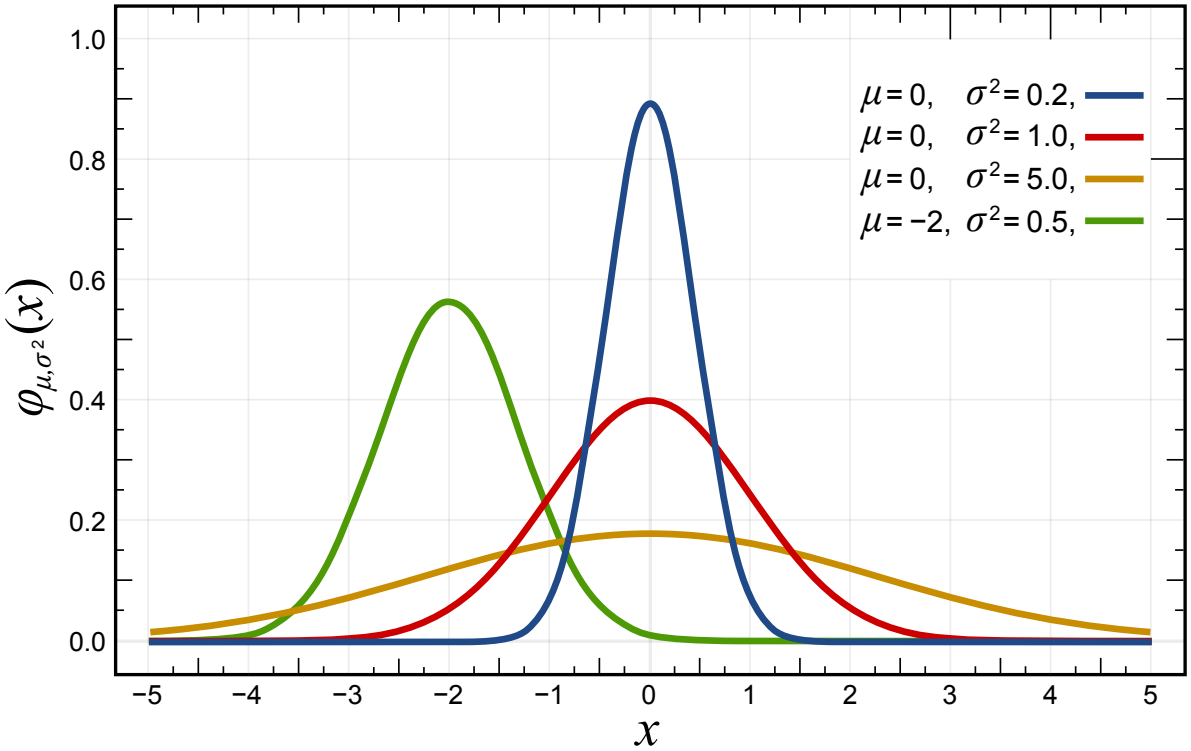


[6] https://web.stanford.edu/ hastie/Papers/ESLII.pdf

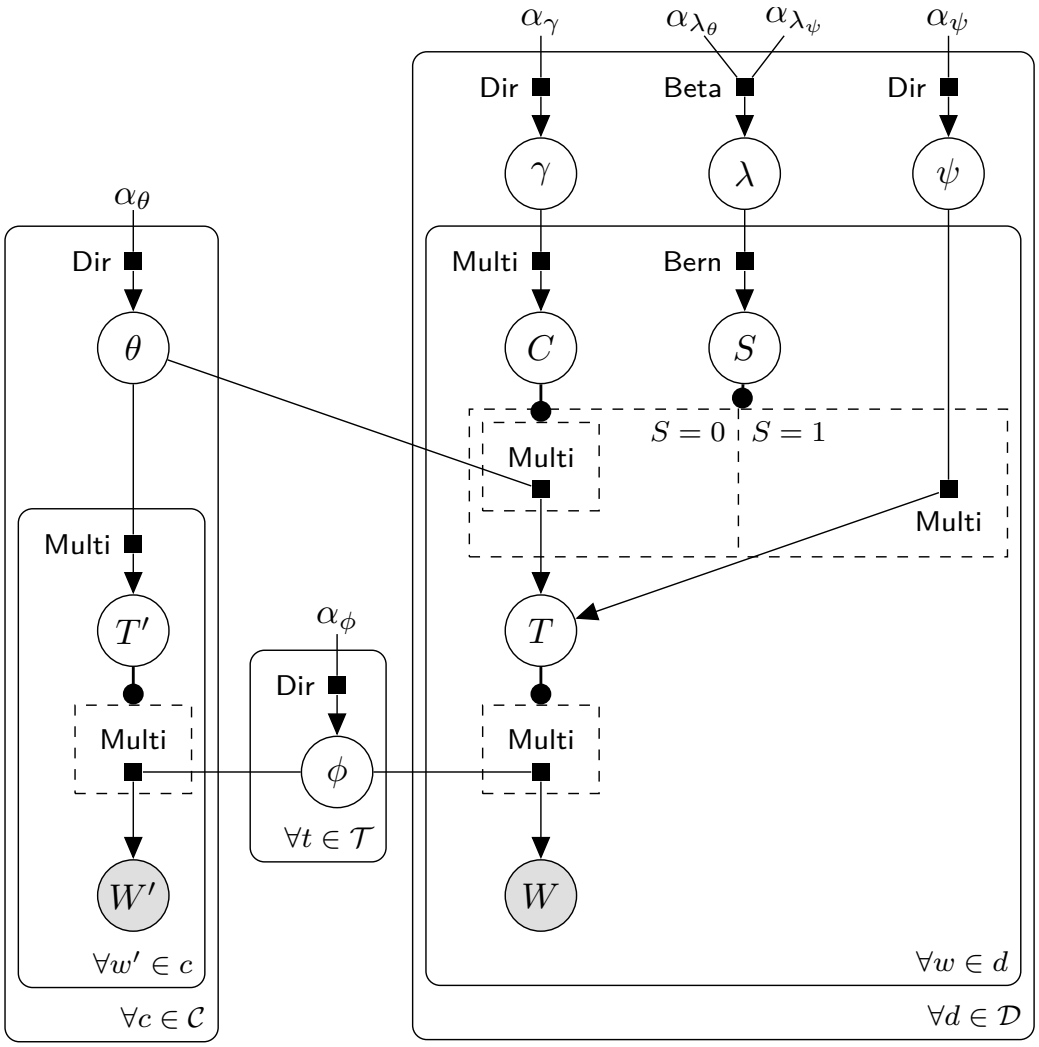
Normal Model of Data $x_1, \ldots, x_N$ for Prediction/Inference

Normal distribution $\mathcal{N}$ with parameters μ, σ^2





Model for Unsupervised Prediction of Citation Influences[7]



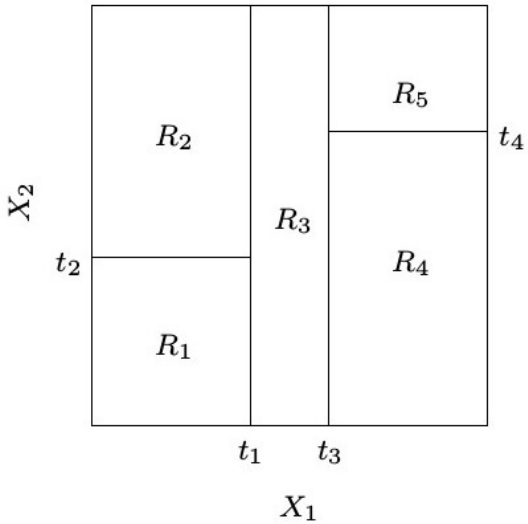
[7] http://www.machinelearning.org/proceedings/icml2007/papers/257.pdf

General Outline

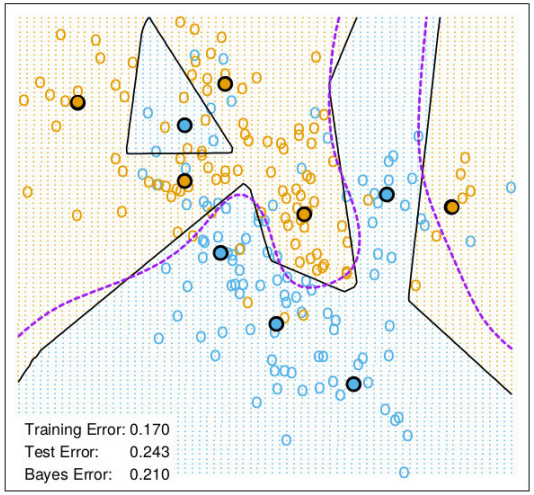
1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

Various Fitted Models Partitioning Input Space[8]

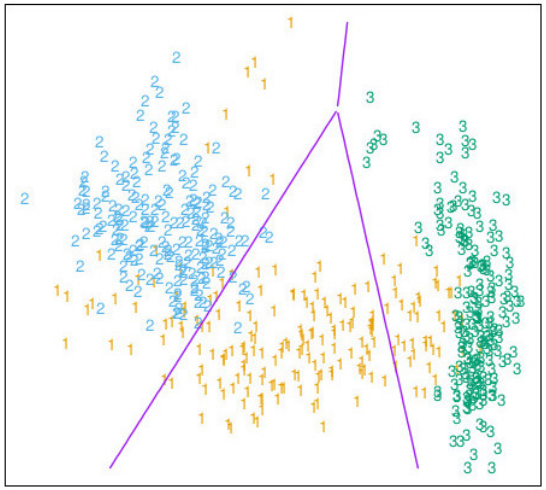
decision tree



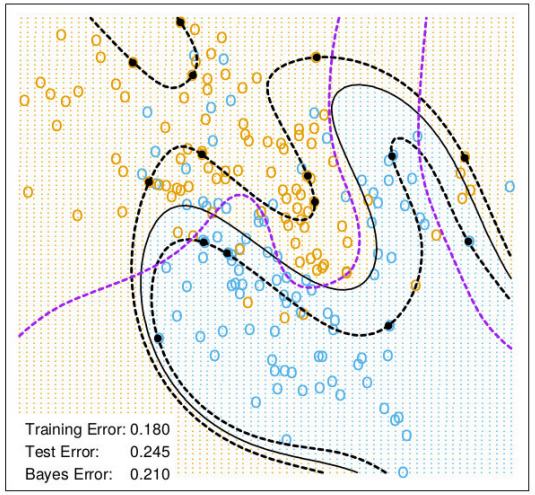
K-means classifier



generalized linear regression



support vector machine



[8] https://web.stanford.edu/ hastie/Papers/ESLII.pdf

Fitting Decision Trees with R and `rpart`

Why R?

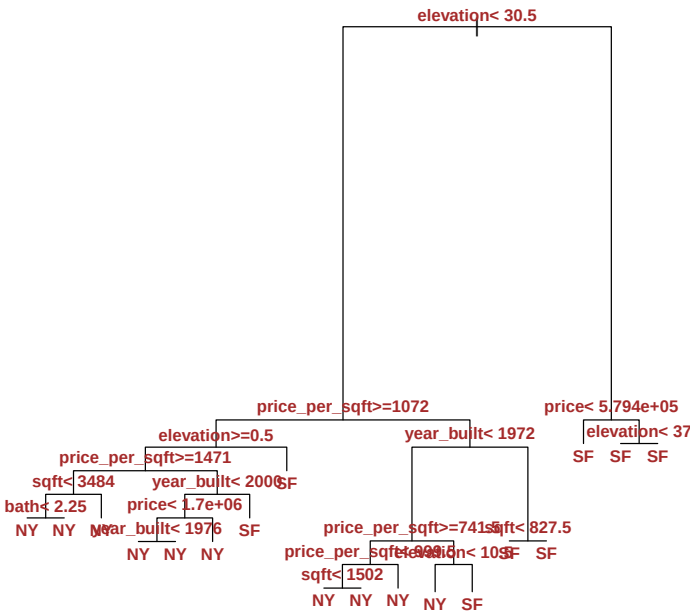
- created by and for biostatisticians
- functional language (like JavaScript)
- open source
- mature
- lots of machine learning packages
- R2D3[9]

[9] http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Fitted Decision Tree(s)[10]

Several related trees may be fitted.

This one is rather complex.



[10] https://attilagk.github.io/R-you-experienced/2017-10-16-fixed-and-mixed-models.html

Demo with "Visual Intro"[11]

Observe progressive growth of tree!

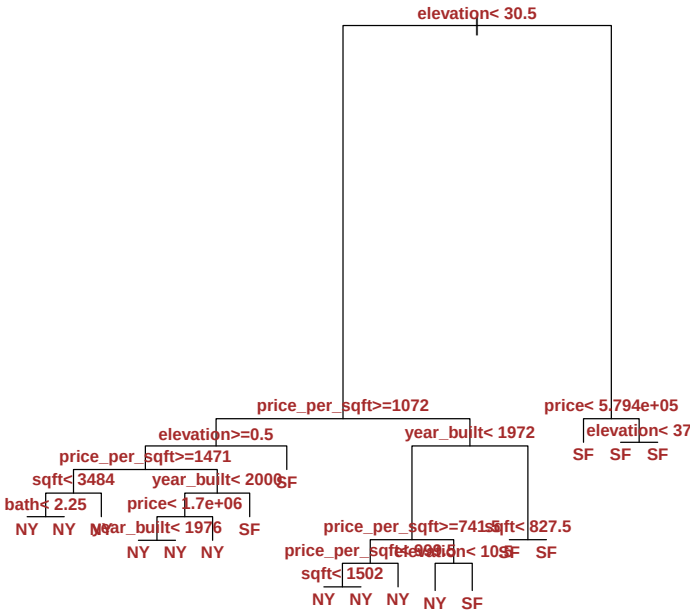
[11] http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

General Outline

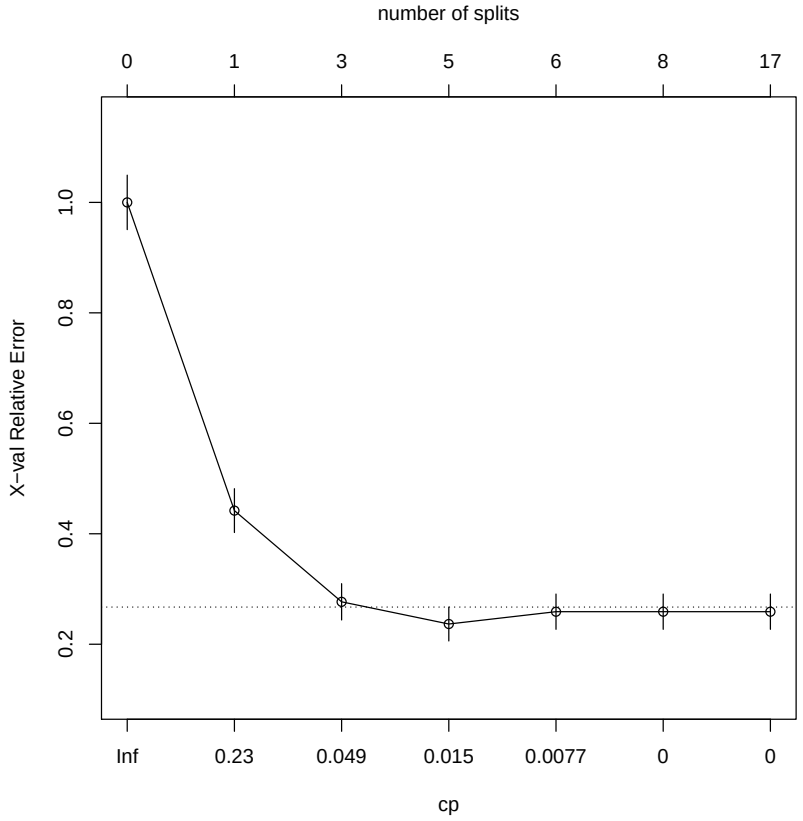
1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

Fitted Decision Tree(s)[10]

Several related trees may be fitted.
This one is rather complex.



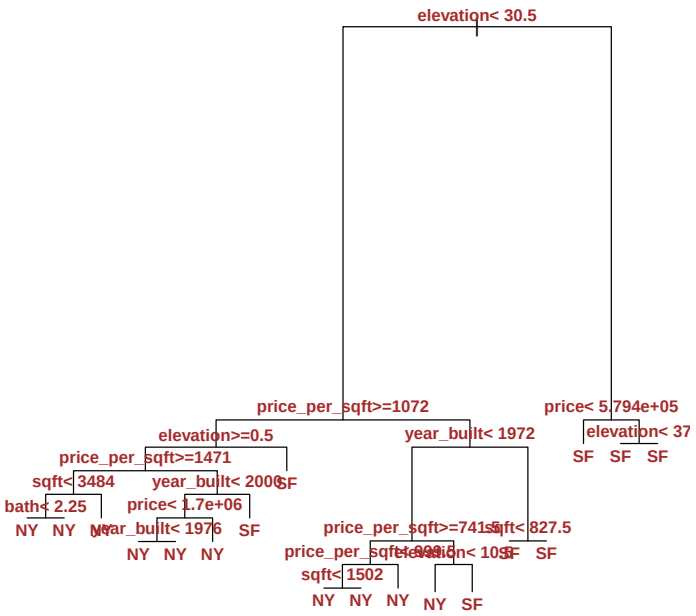
Tree selection based on fit (error) and complexity (cp)



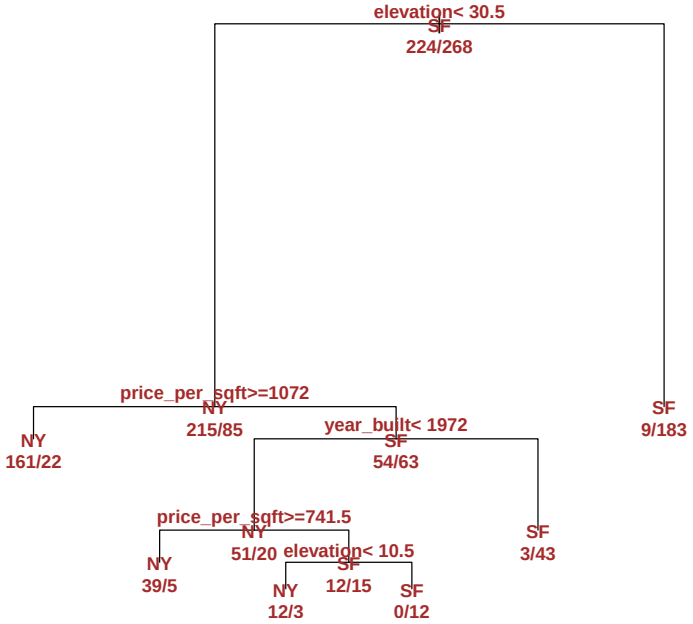
[10] https://attilagk.github.io/R-you-experienced/2017-10-16-fixed-and-mixed-models.html

Fitted Decision Tree(s)[10]

Several related trees may be fitted.
This one is rather complex.



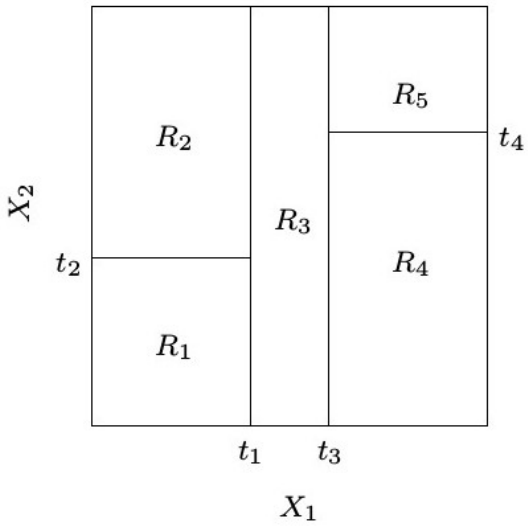
The optimal tree



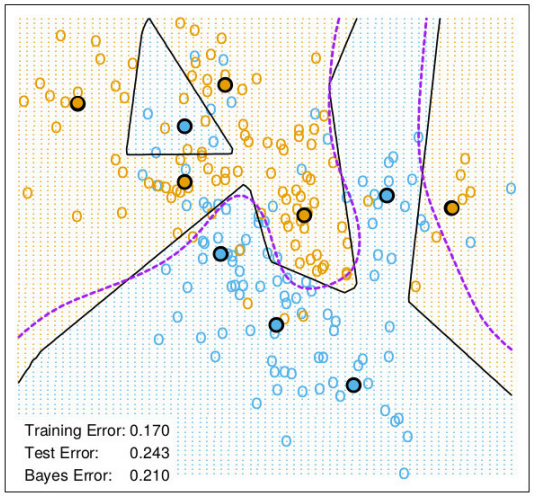
[10] https://attilagk.github.io/R-you-experienced/2017-10-16-fixed-and-mixed-models.html

Various Fitted Models Partitioning Input Space[8]

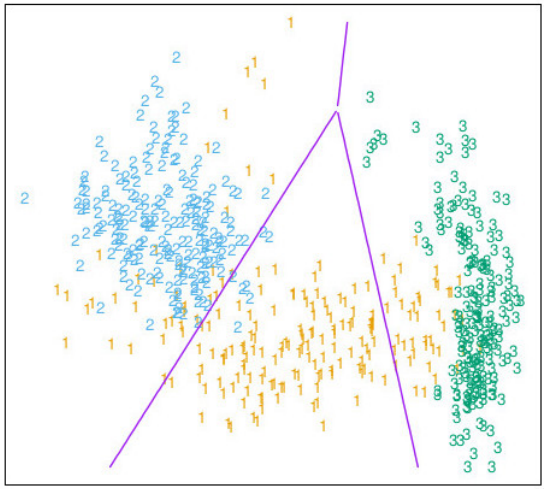
decision tree performs poorly



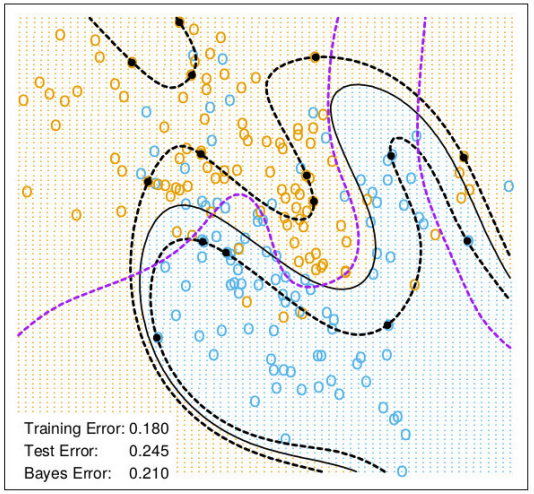
K-means classifier



generalized linear regression



support vector machine

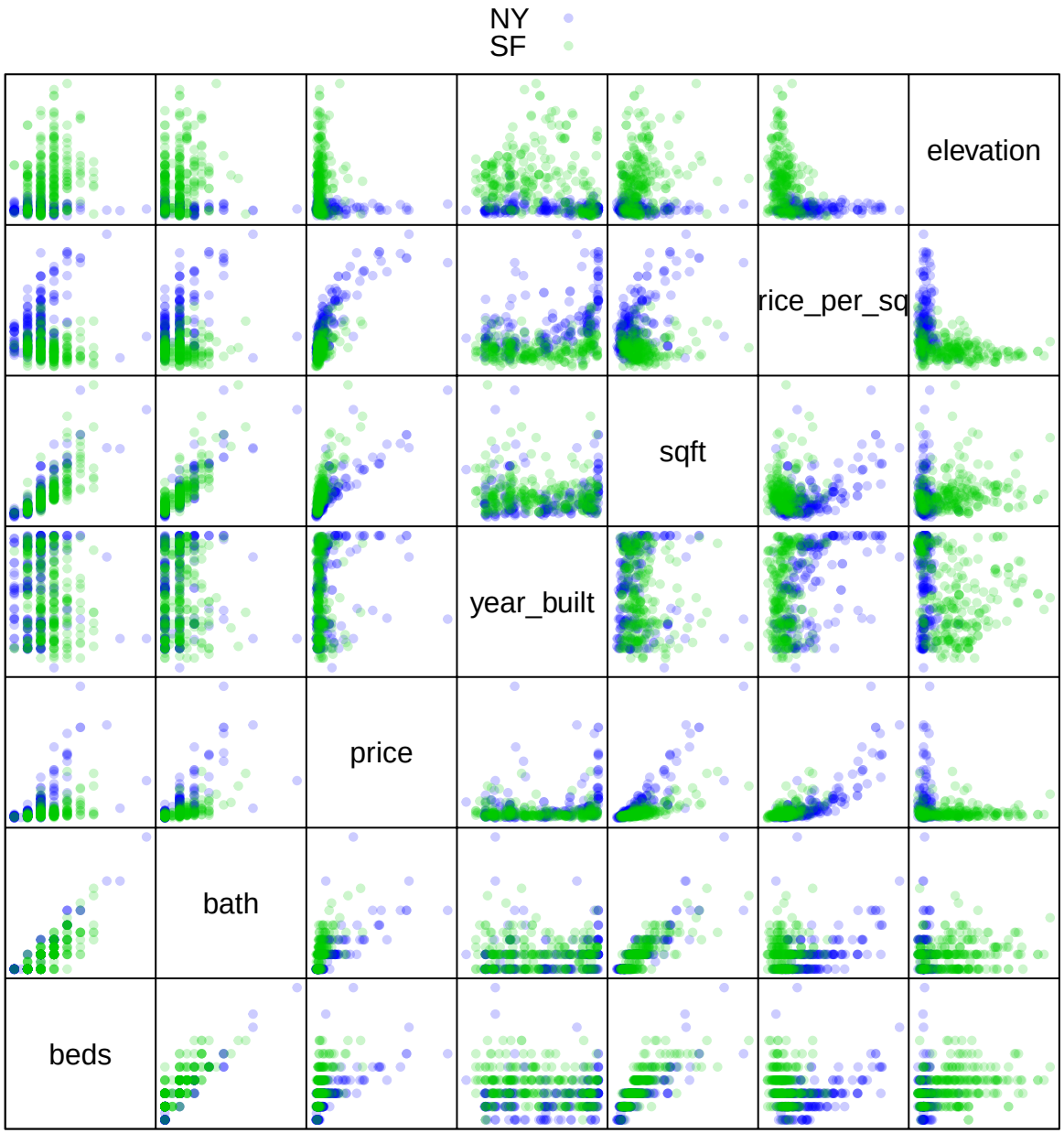


[8] https://web.stanford.edu/ hastie/Papers/ESLII.pdf

General Outline

1. design project (EN, ST)
2. collect and clean data (EN, PR, DSS, HPC)
3. explore data (EN, PR)
4. formulate task (EN, ST)
5. build models (EN, ST)
6. fit/learn/train the model(s) (PR, LIB, HPC)
7. select best model(s) (ST, PR, LIB, HPC)
8. apply best model to test data (PR, LIB)
9. interpret and report results (EN, ST, DOC)

All Inputs

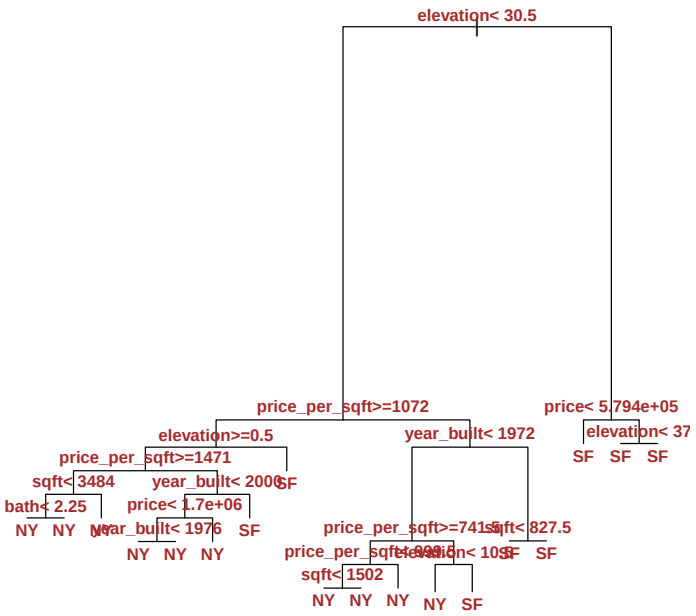


Scatter Plot Matrix

Fitted Decision Tree(s)[10]

Several related trees may be fitted.

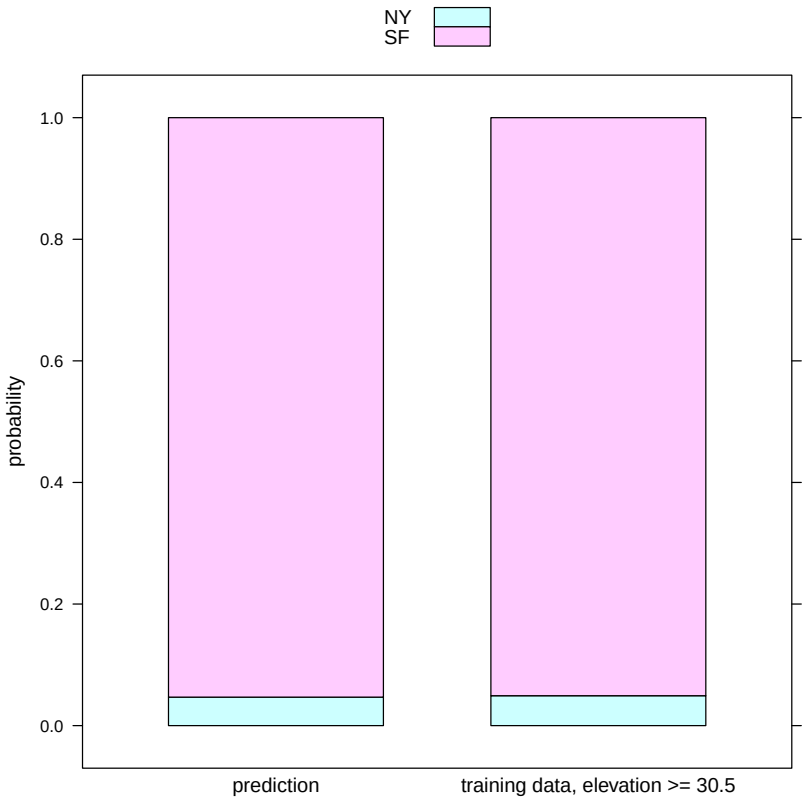
This one is rather complex.

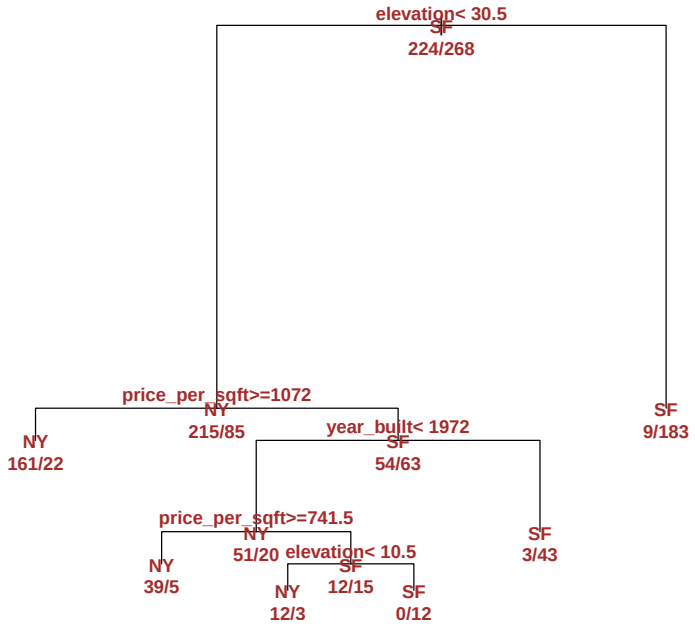


[10] https://attilagk.github.io/R-you-experienced/2017-10-16-fixed-and-mixed-models.html

Classifying Homes as NY or SF

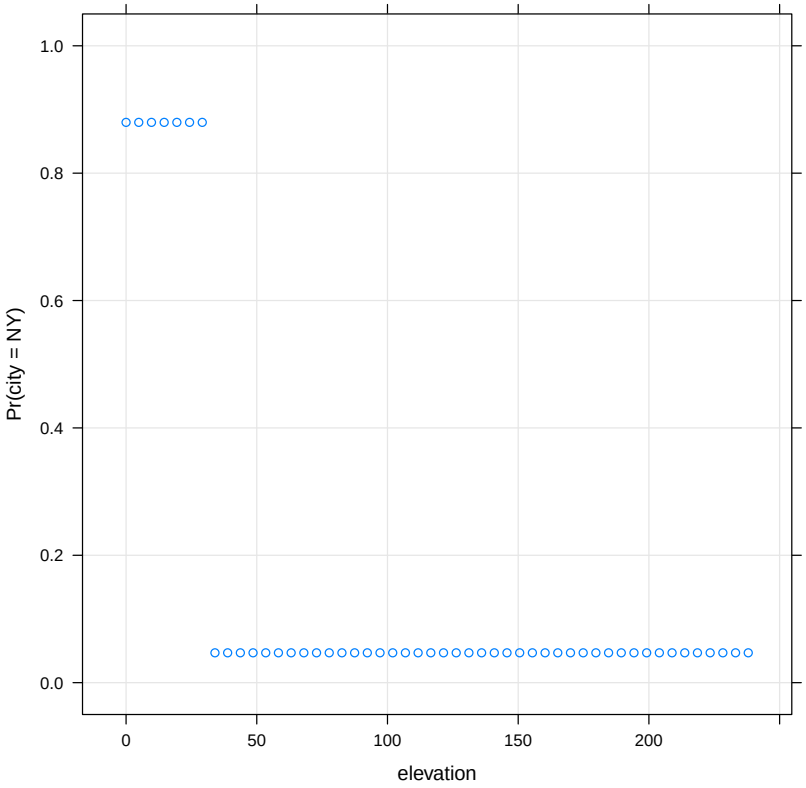
at "the average of traning data"

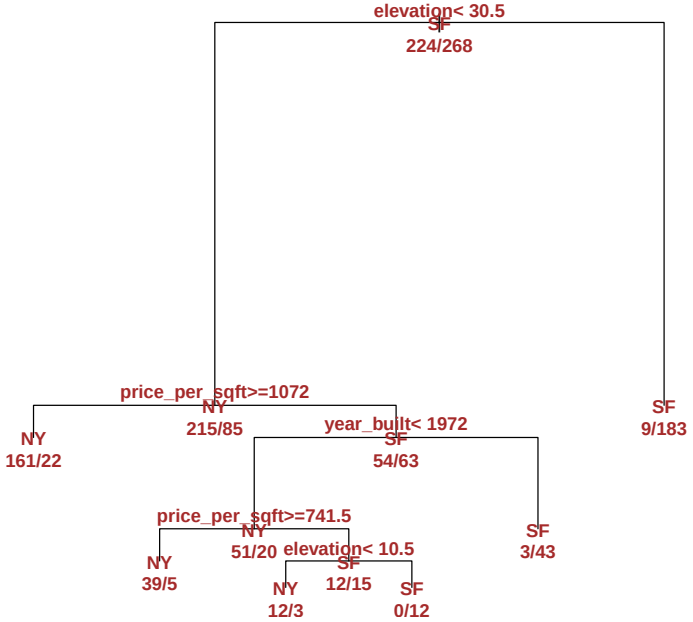




Classifying Homes as NY or SF

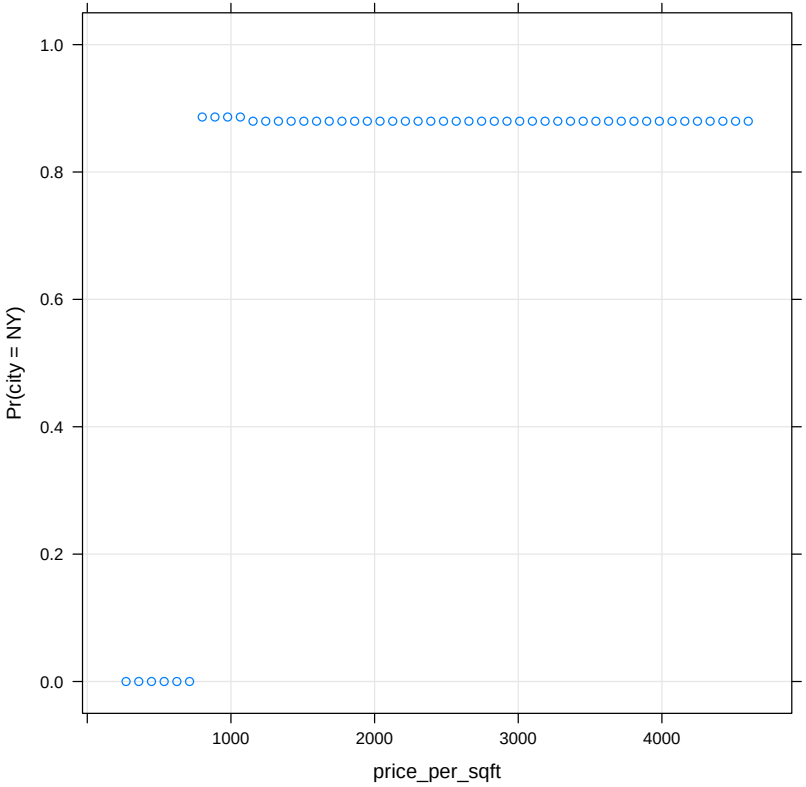
at varying elevation

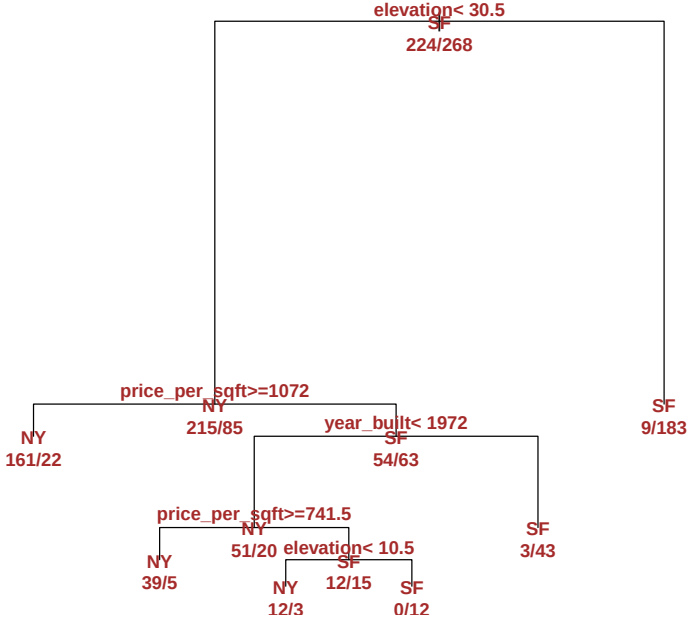




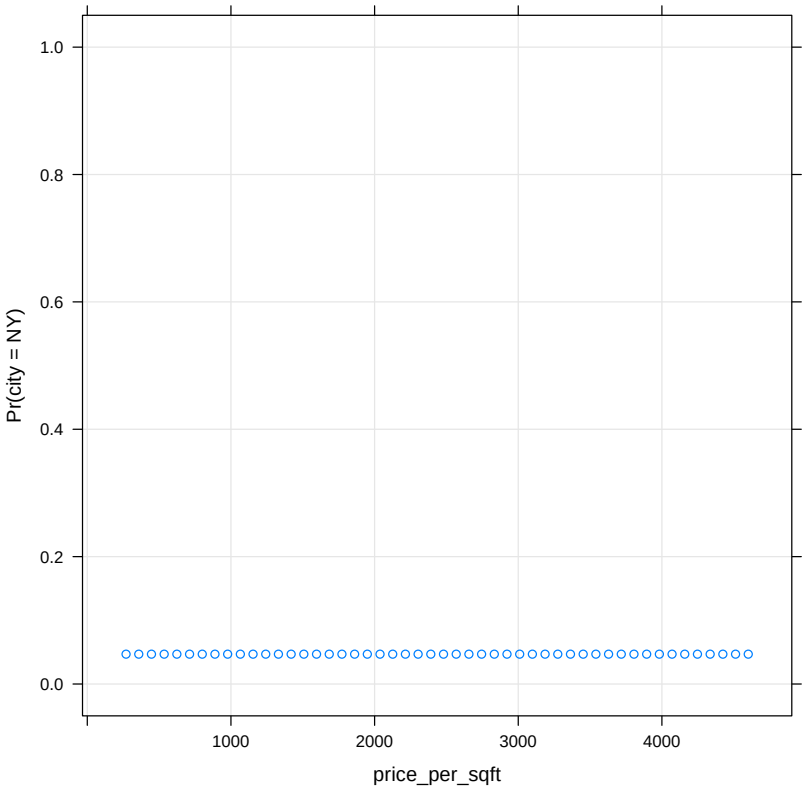
Classifying Homes as NY or SF

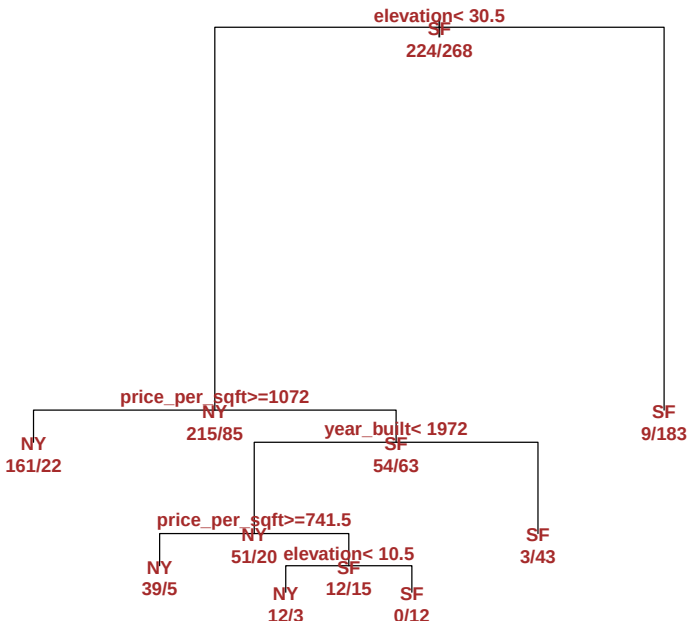
at varying price/sqft and average (40m) elevation





Classifying Homes as NY or SF





at varying price/sqft and 30m elevation

Conclusion: Machine Learning and You

1. understanding it
   - learn concepts not cooking
   - collaboration, interpretation
2. doing it
   - `Hello World!` is easy but useless
   - obtaining skills takes years but then pays off
3. Resources
   - The Elements of Statistical Learning
     https://web.stanford.edu/ hastie/Papers/ESLII.pdf
   - Machine Learning
     https://www.cs.ubc.ca/ murphyk/MLbook/
   - An Introduction to R
     https://cran.r-project.org/doc/manuals/r-release/R-intro.html
   - https://attilagk.github.io/R-you-experienced

Conclusion: Machine Learning and You

1. understanding it
   - learn concepts not cooking
   - collaboration, interpretation
2. doing it
   - `Hello World!` is easy but useless
   - obtaining skills takes years but then pays off
3. Resources
   - The Elements of Statistical Learning
     https://web.stanford.edu/ hastie/Papers/ESLII.pdf
   - Machine Learning
     https://www.cs.ubc.ca/ murphyk/MLbook/
   - An Introduction to R
     https://cran.r-project.org/doc/manuals/r-release/R-intro.html
   - https://attilagk.github.io/R-you-experienced