Visual validation of learning methods

Visual Analysis of Measurement Data

László Gönczy

2019.11.21.

Budapest University of Technology and Economics
Fault Tolerant Systems Research Group

- Remember the Anscombe quartett…
- Some of the common goals
  - Sensitivity of the results to a parameter
  - Correctness/preciseness of the results
  - Identifying outliers and reasons for mistakes
  - „Approval” and understanding
  - Compare alternatives

Reminder: what is the goal

- Classification
  - Binary/multiple class
- Regression
- Clustering
- Association rules…
- In details in other classes
- Important: „MLaaS”

Data mining „brickstones”

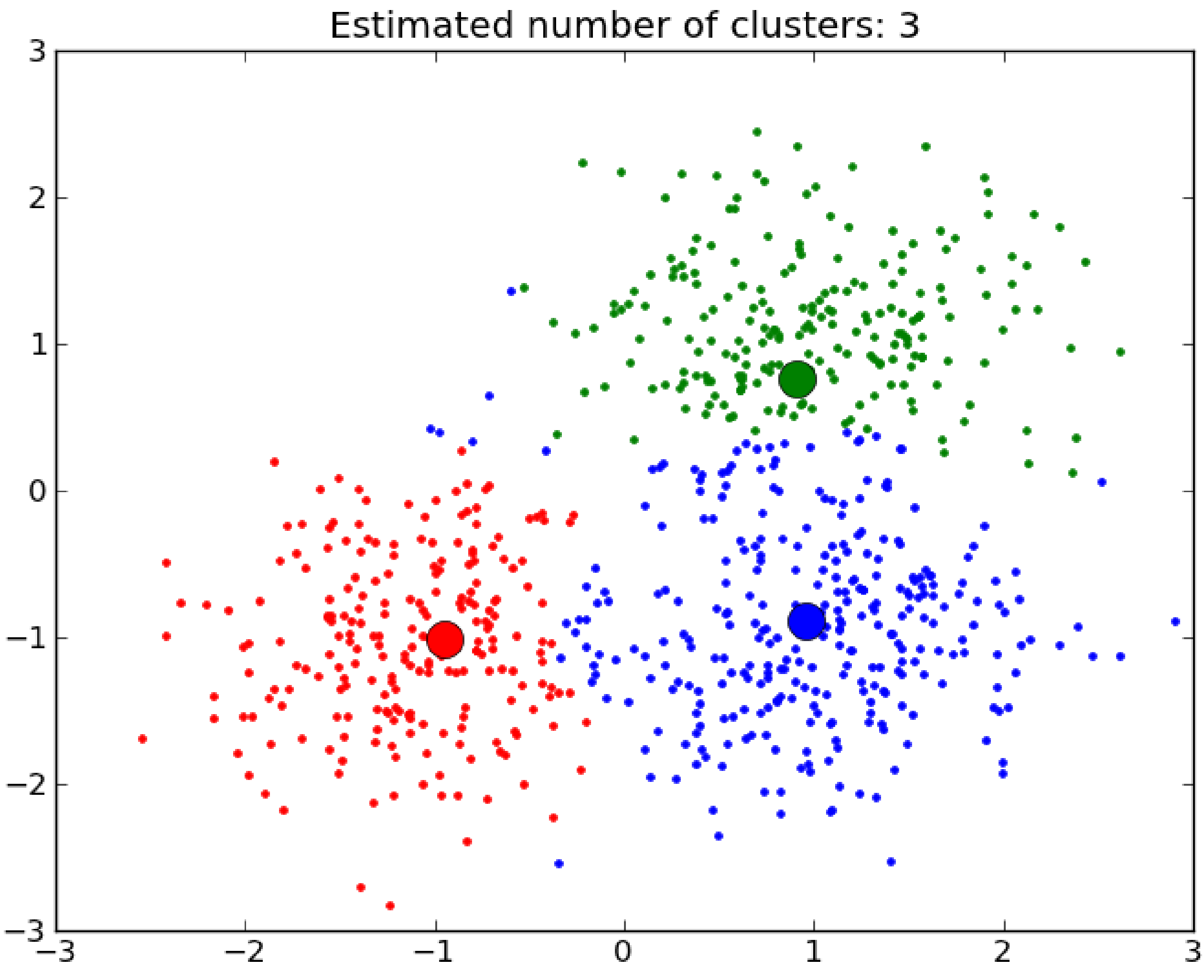
Clustering

Classification

Association rules

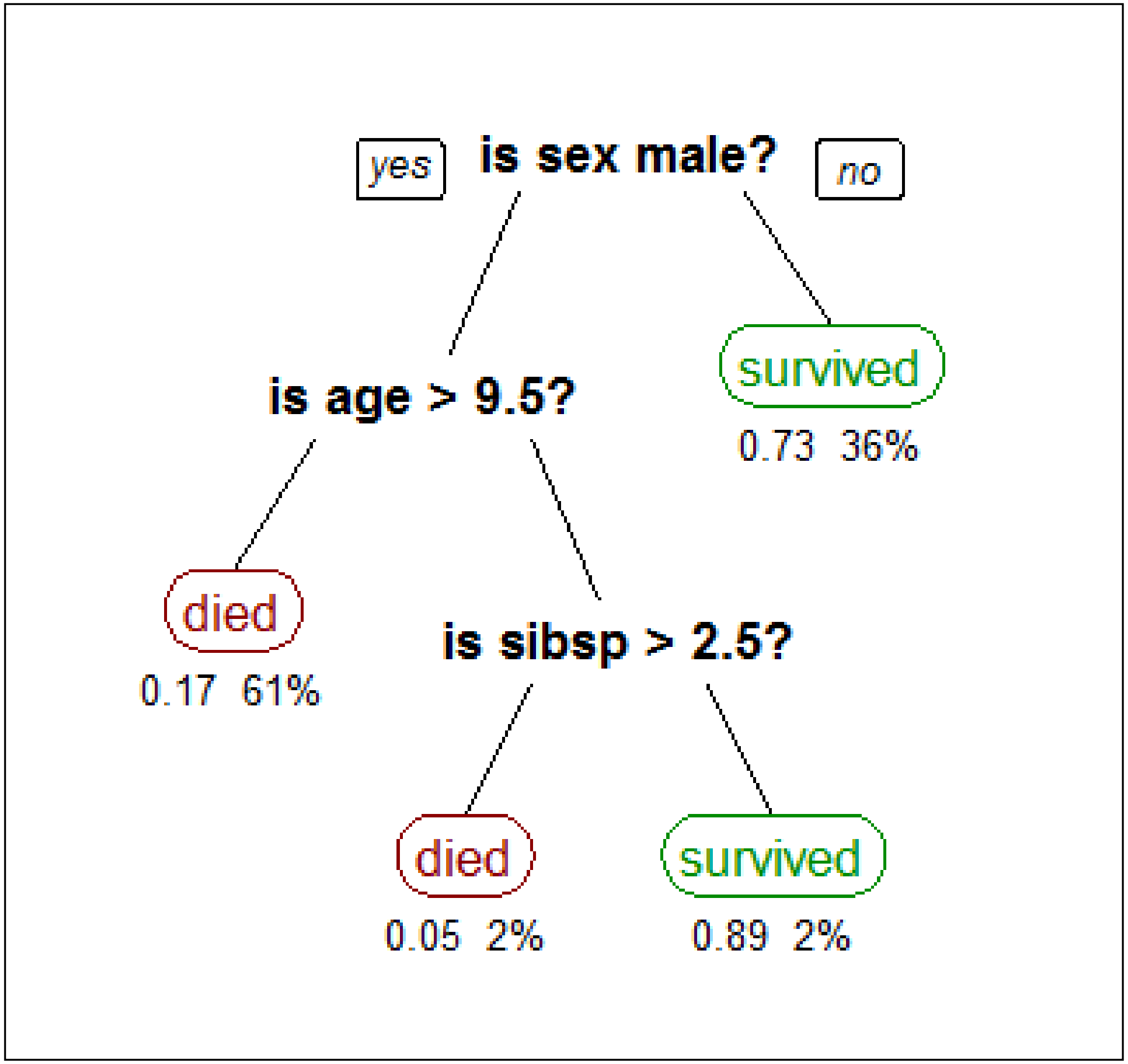
Regression

Clustering

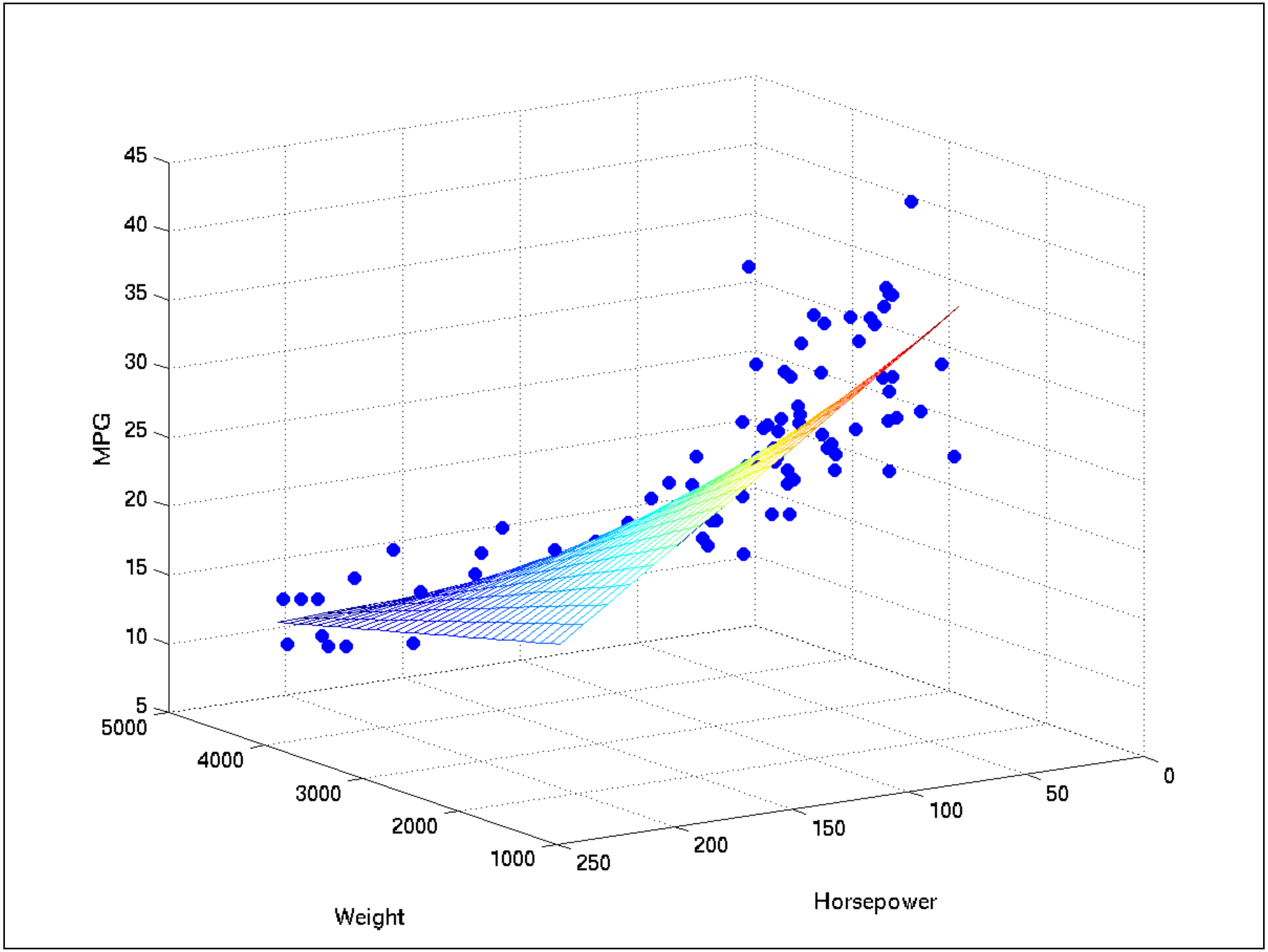


Association rules

Classification



Regression


MPG
45
40
35
30
25
20
15
10
5
5000
4000
3000
2000
1000
Weight
250
200
150
100
50
0
Horsepower

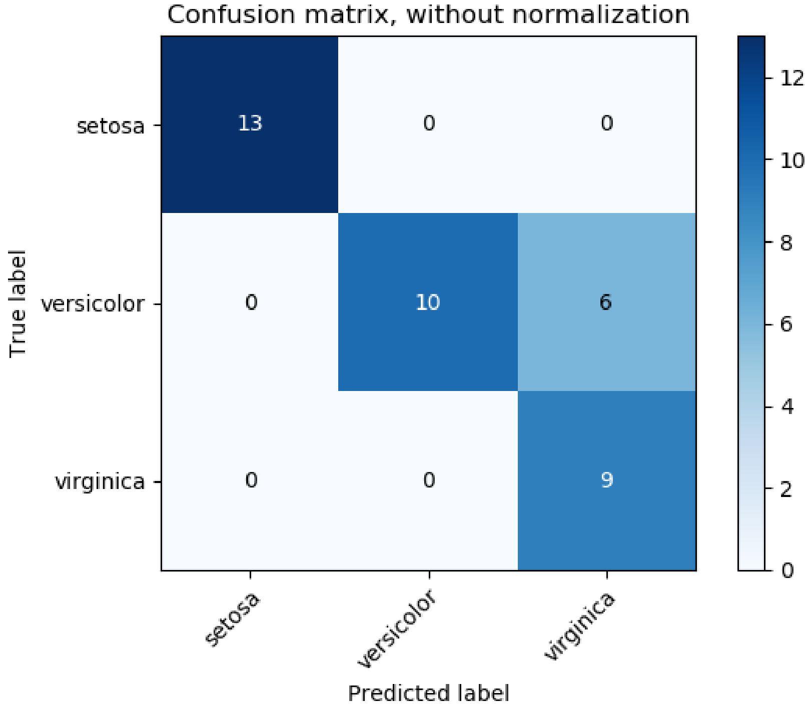
Basics

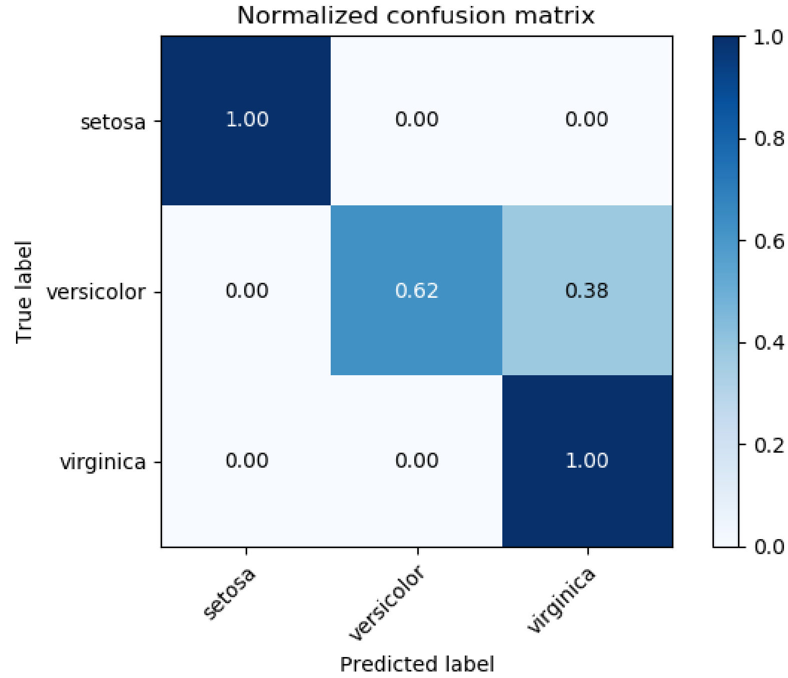
- How to classify one prediction?

		Real value	
		Condition true	Condition false
Predicted value	Condition true	True positive	False positive
	Condition false	False negative	True negative

Confusion matrix

- Generalization for C categories





https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Measures

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

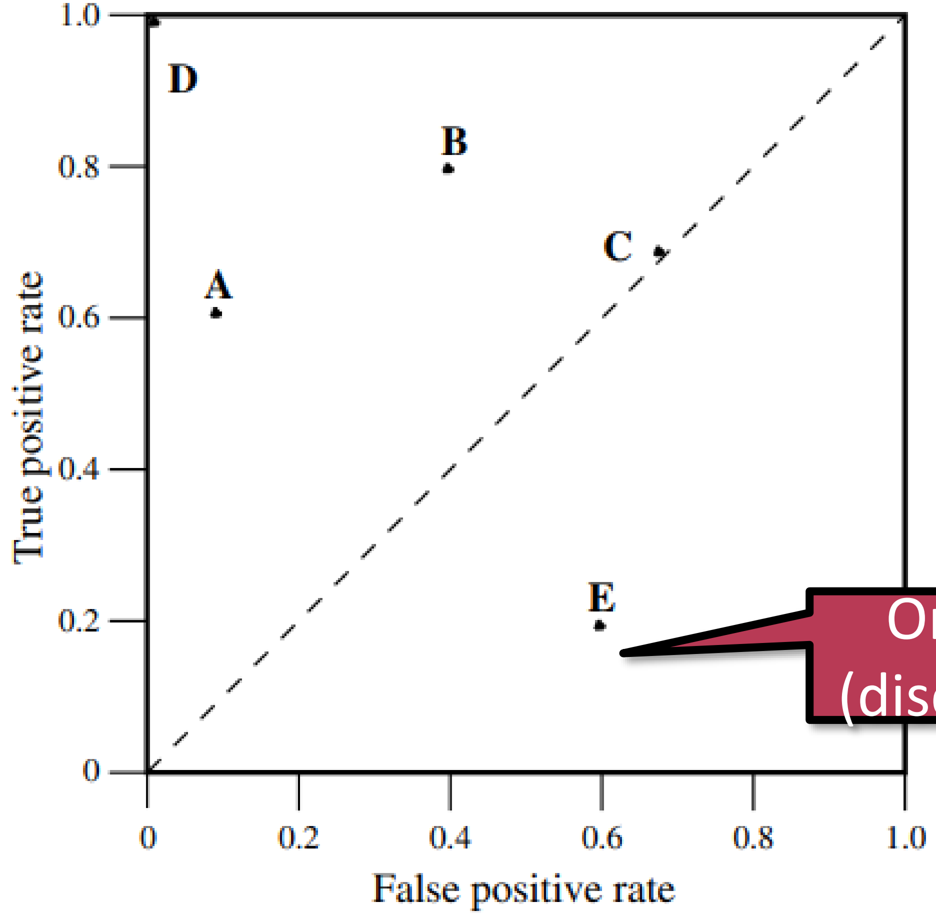
$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

ROC

- Receiver Operating Characteristic Plot

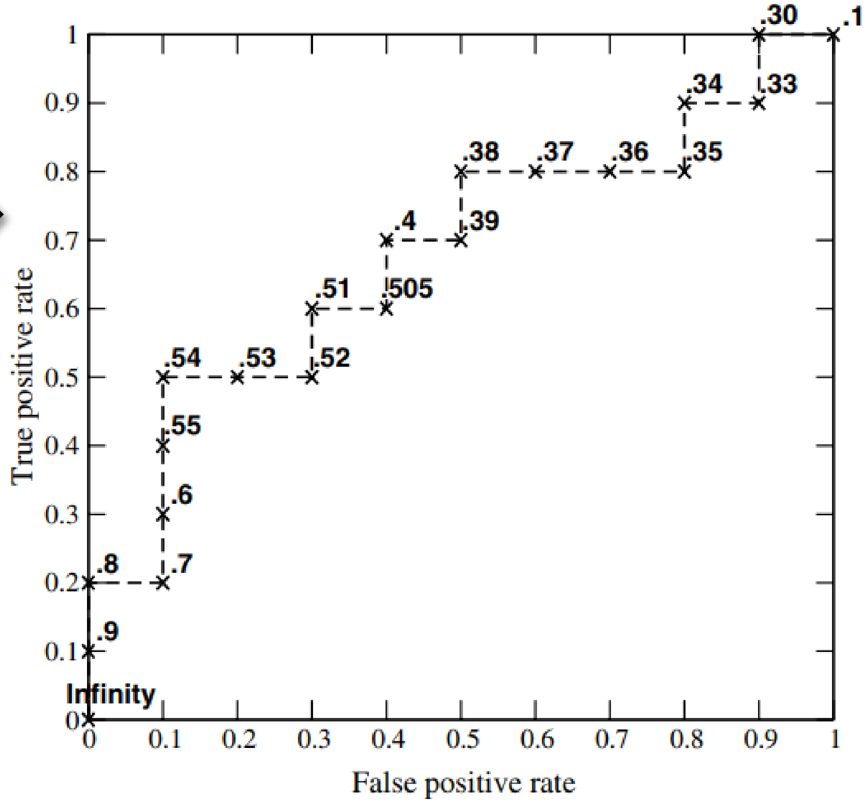


One point: one (discrete) classifier

ROC curve

- Classifiers with scores/probabilities
- Step function for given points

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1



Insensitive for class skewness

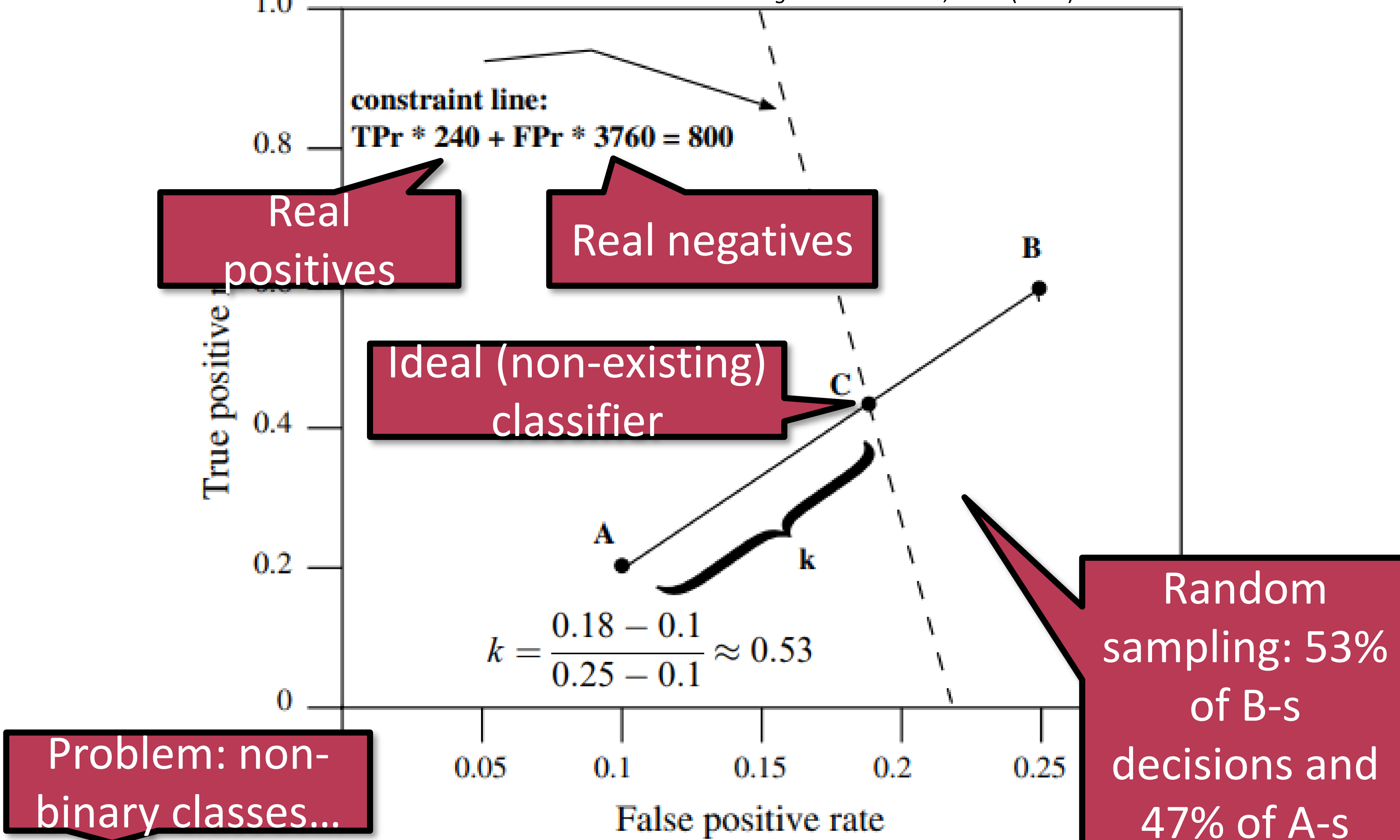
Interpolating classifiers

- Discrete number of discrete classifiers
- Example: CoIL Challenge 2000
  - 4000 clients, budget for 800
  - Expected responders: 6%
- Classifier A: 424 client candidates
- Classifier B: 1084 candidates

Fawcett, Tom. "An introduction to ROC analysis."
Pattern recognition letters 27, no. 8 (2006): 861-874.

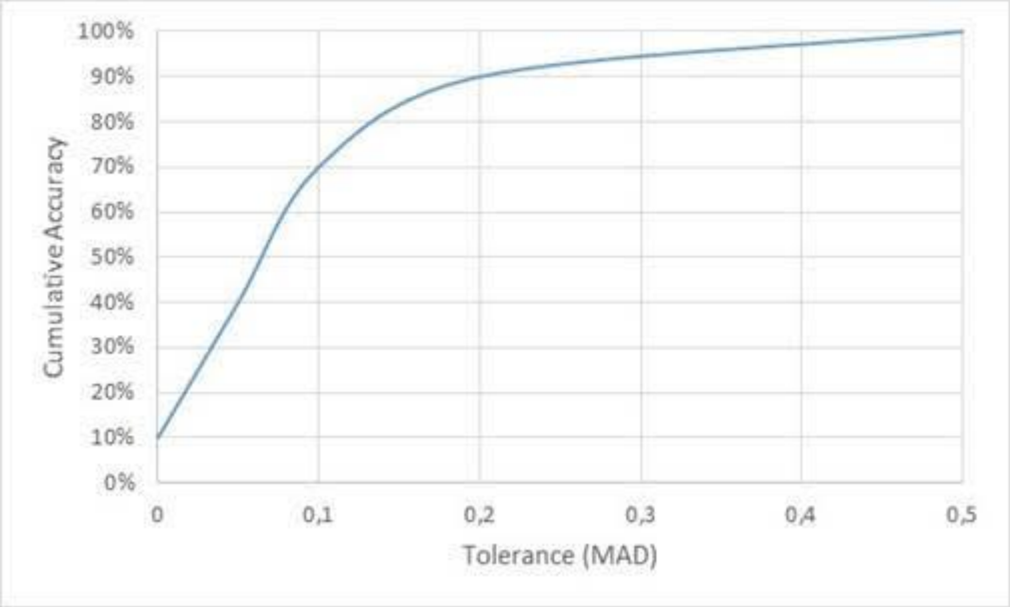
Interpolating classifiers

Fawcett, Tom. "An introduction to ROC analysis." *Pattern recognition letters* 27, no. 8 (2006): 861-874.



Extensions

- Regression Error Curve
  - Error: squared residual, absolute deviation, etc.
  - Releative position of alternatives will not change



https://www.dataminingapps.com/2016/07/what-is-a-rec-curve-in-a-regression-context/

Regression ROC

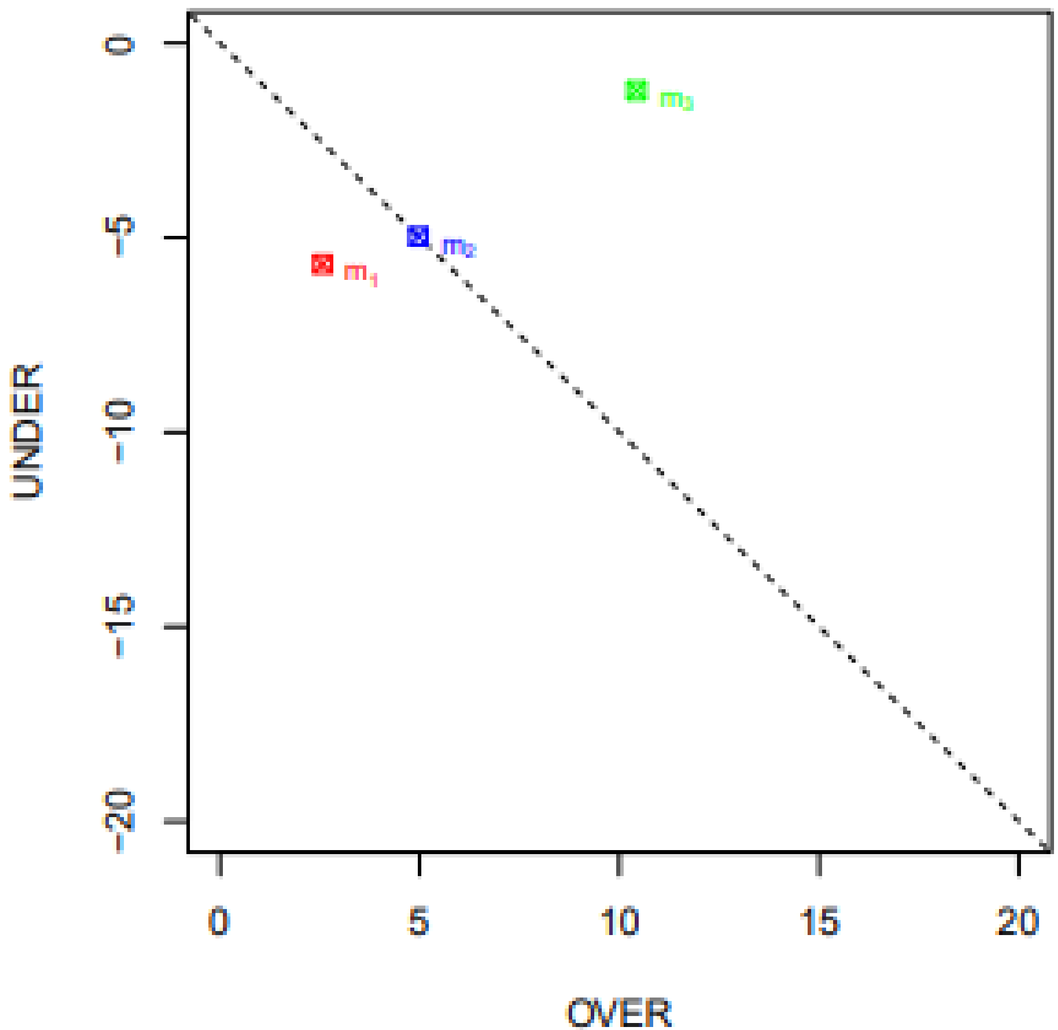
	1	2	3	4	5	6	7	8	9	10
$\hat{y}$	-0.082	3.323	2.320	1.080	7.893	4.983	5.121	3.442	2.083	1.112
y	0.211	2.725	1.933	3.242	7.858	6.061	7.173	3.082	0.894	1.203
e	-0.293	0.598	0.387	-2.162	0.035	-1.078	-2.052	0.360	1.189	-0.091

Predicted vs actual value, error

OVER vs UNDER estimation

Hernández-Orallo, José. "A graphical analysis of cost-sensitive regression problems." (2012).
https://arxiv.org/pdf/1211.2359.pdf

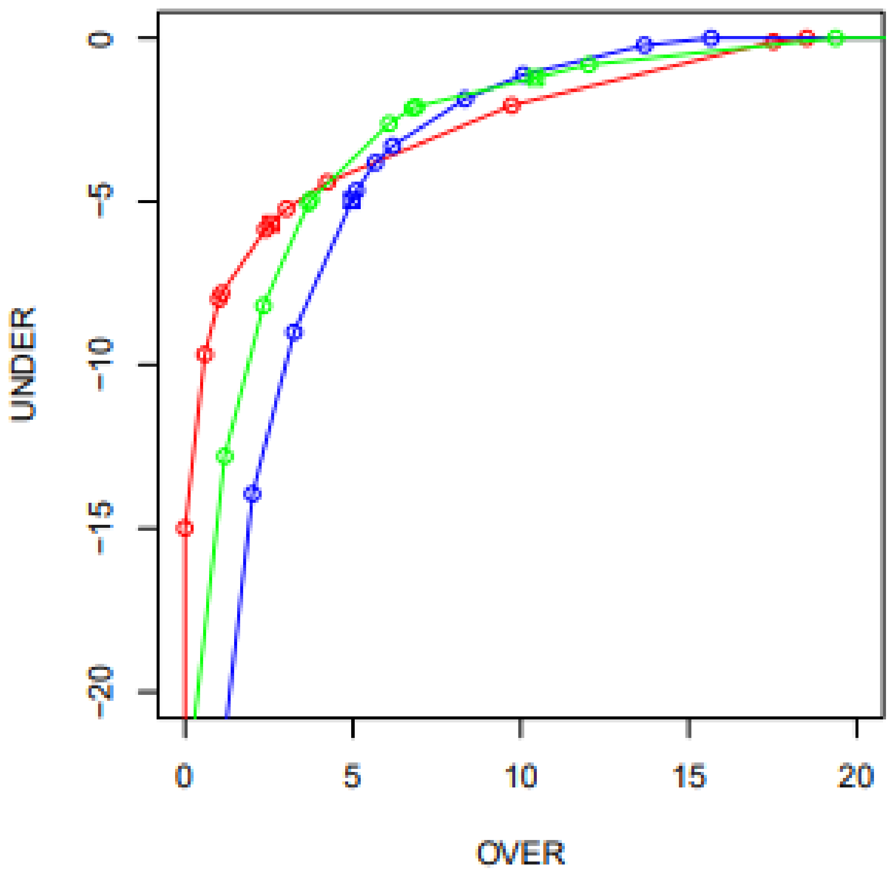
RROC example



Hernández-Orallo, José. "A graphical analysis of cost-sensitive regression problems." (2012).
https://arxiv.org/pdf/1211.2359.pdf

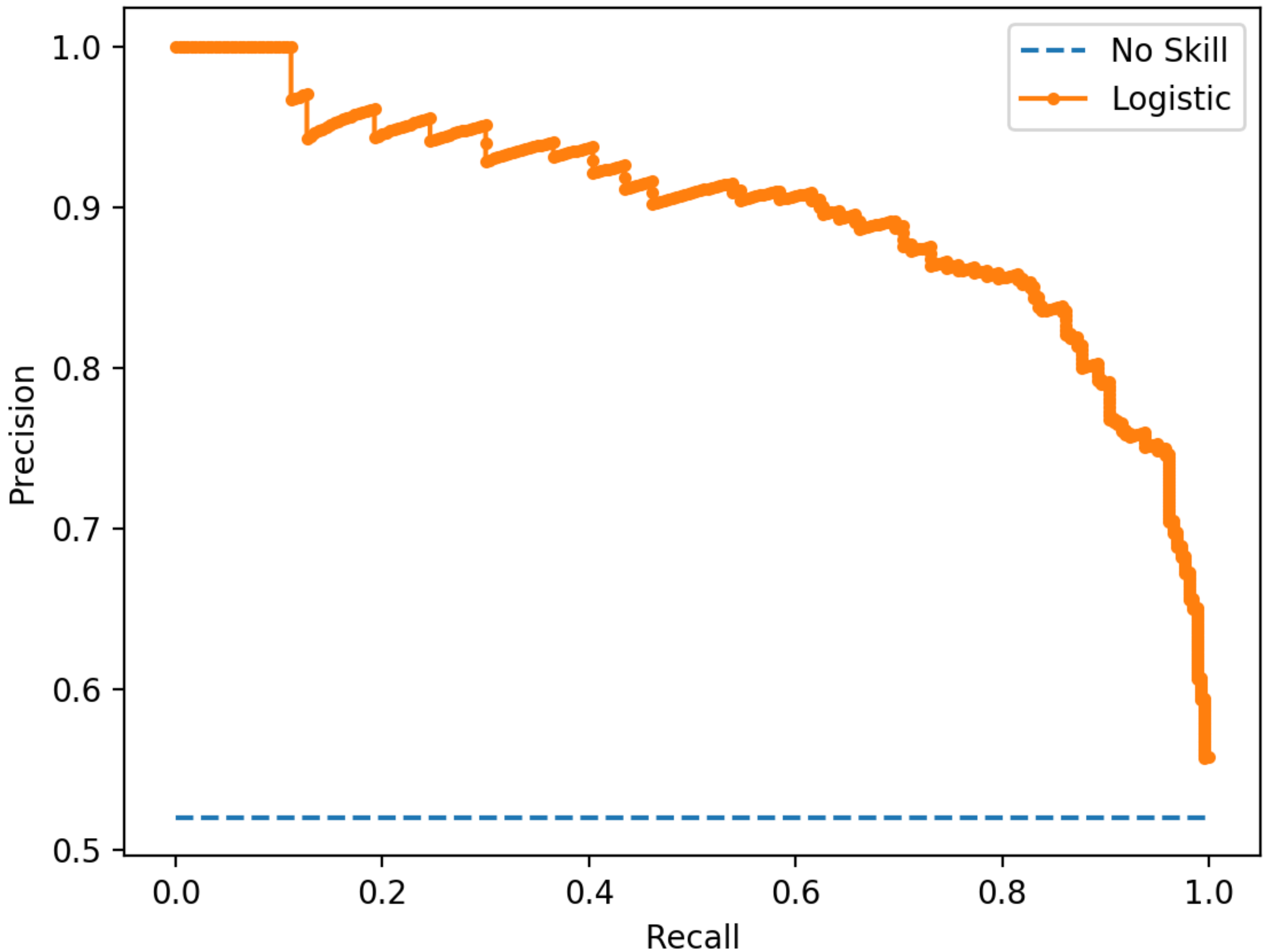
RROC curve

- Instead of threshold: shift (add s to the prediction)



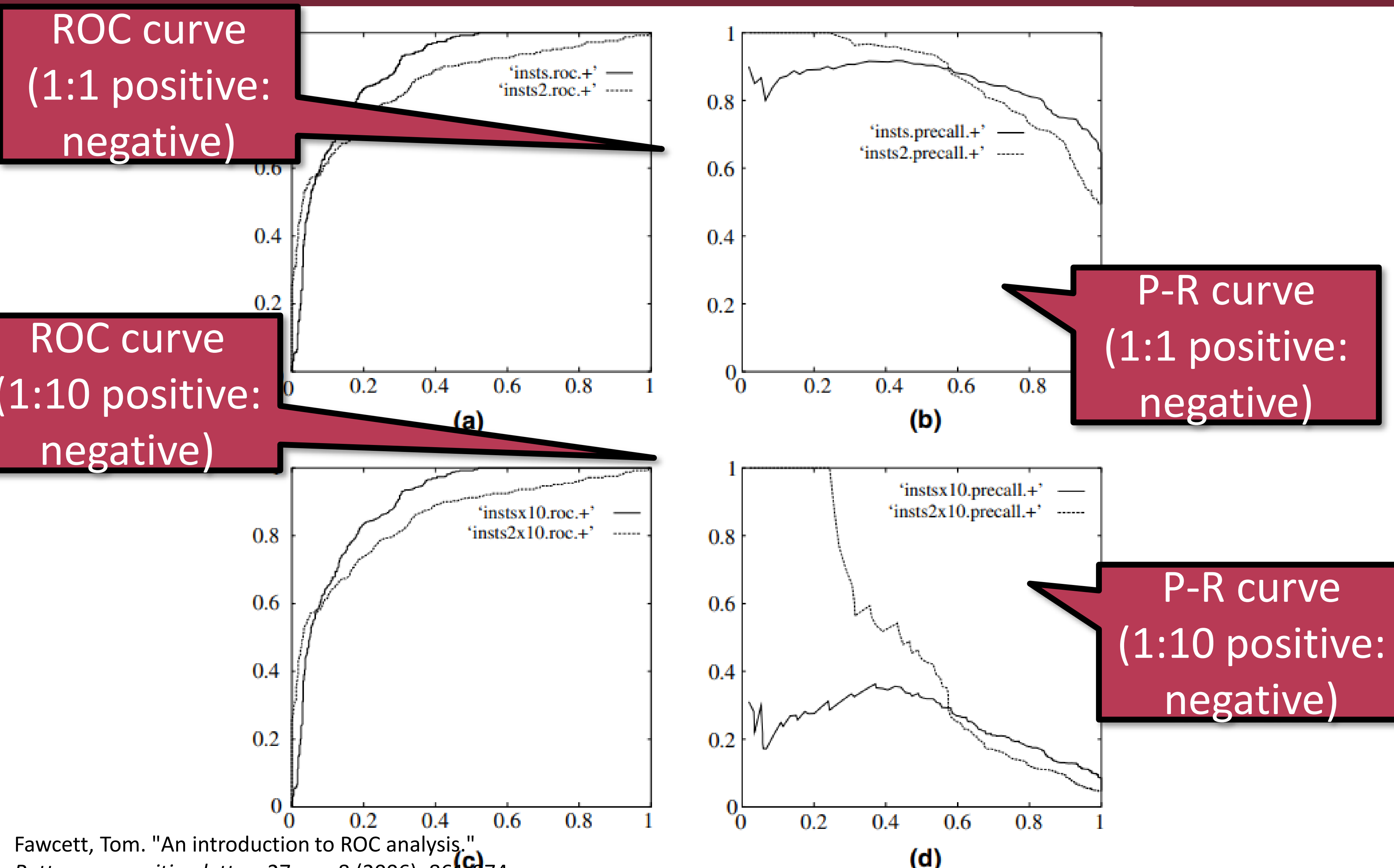
Hernández-Orallo, José. "A graphical analysis of cost-sensitive regression problems." (2012).
https://arxiv.org/pdf/1211.2359.pdf

Precision-recall plots



https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/

Comparison of two classifiers

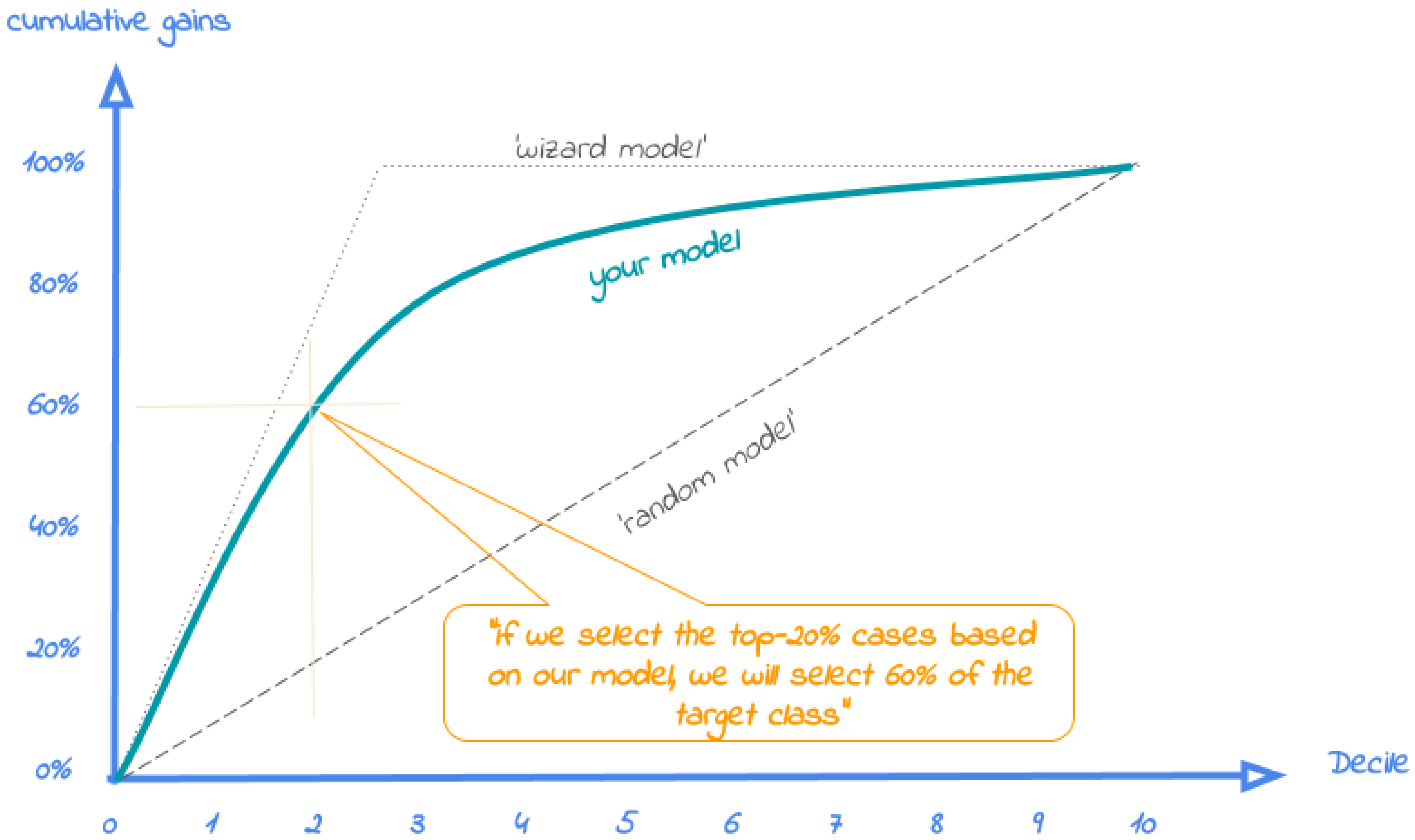


Fawcett, Tom. "An introduction to ROC analysis." *Pattern recognition letters* 27, no. 8 (2006): 861-874.

Understanding the results

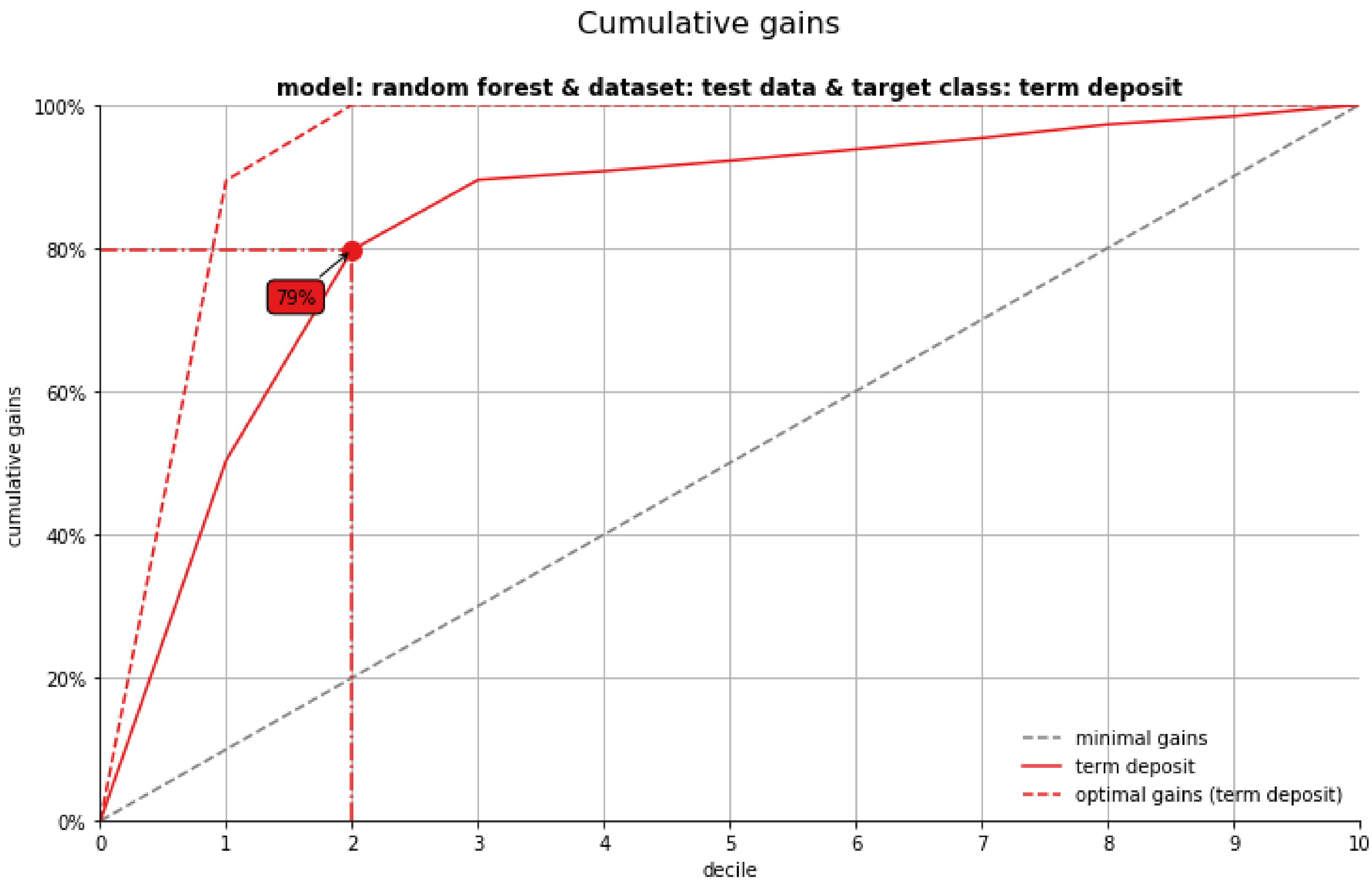
- How good is a model?
- How to compare models?
- Inputs
  - Predicted probability of the target class
  - Equally sized groups based on this probability
    - Typcially deciles
  - Observations/real values of the target class
- → model performance metrics
- Gains, lift, response
- From https://modelplot.github.io/intro_modelplotpy.html

Cumulative gains plot



https://modelplot.github.io/intro_modelplotpy.html

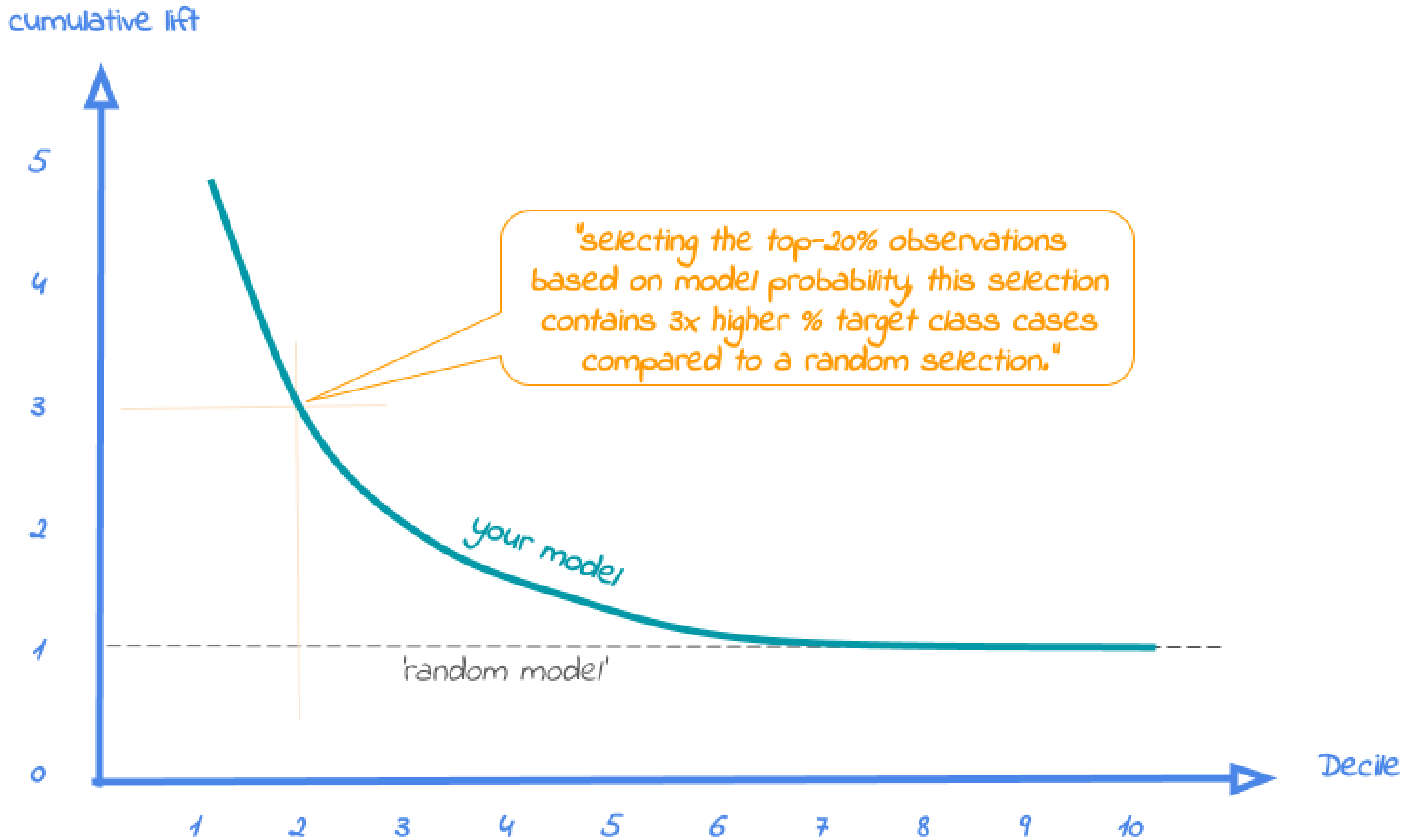
Our model vs perfect predictor



When we select 20% with the highest probability according to model random forest, this selection holds 79% of all term deposit cases in dataset test data.

https://modelplot.github.io/intro_modelplotpy.html

Cumulative lift plot

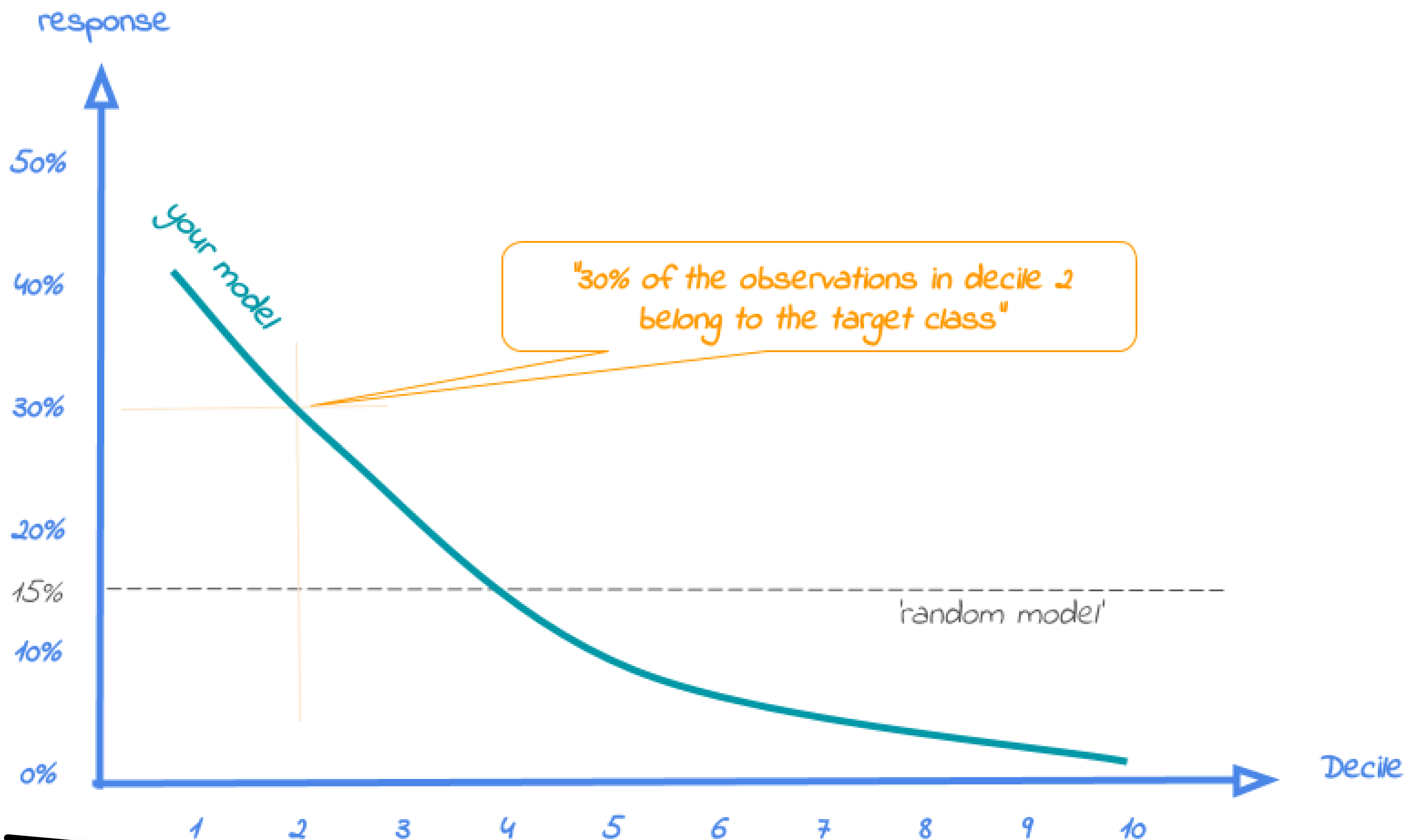


https://modelplot.github.io/intro_modelplotpy.html

Example

https://modelplot.github.io/intro_modelplotpy.html

Response plot

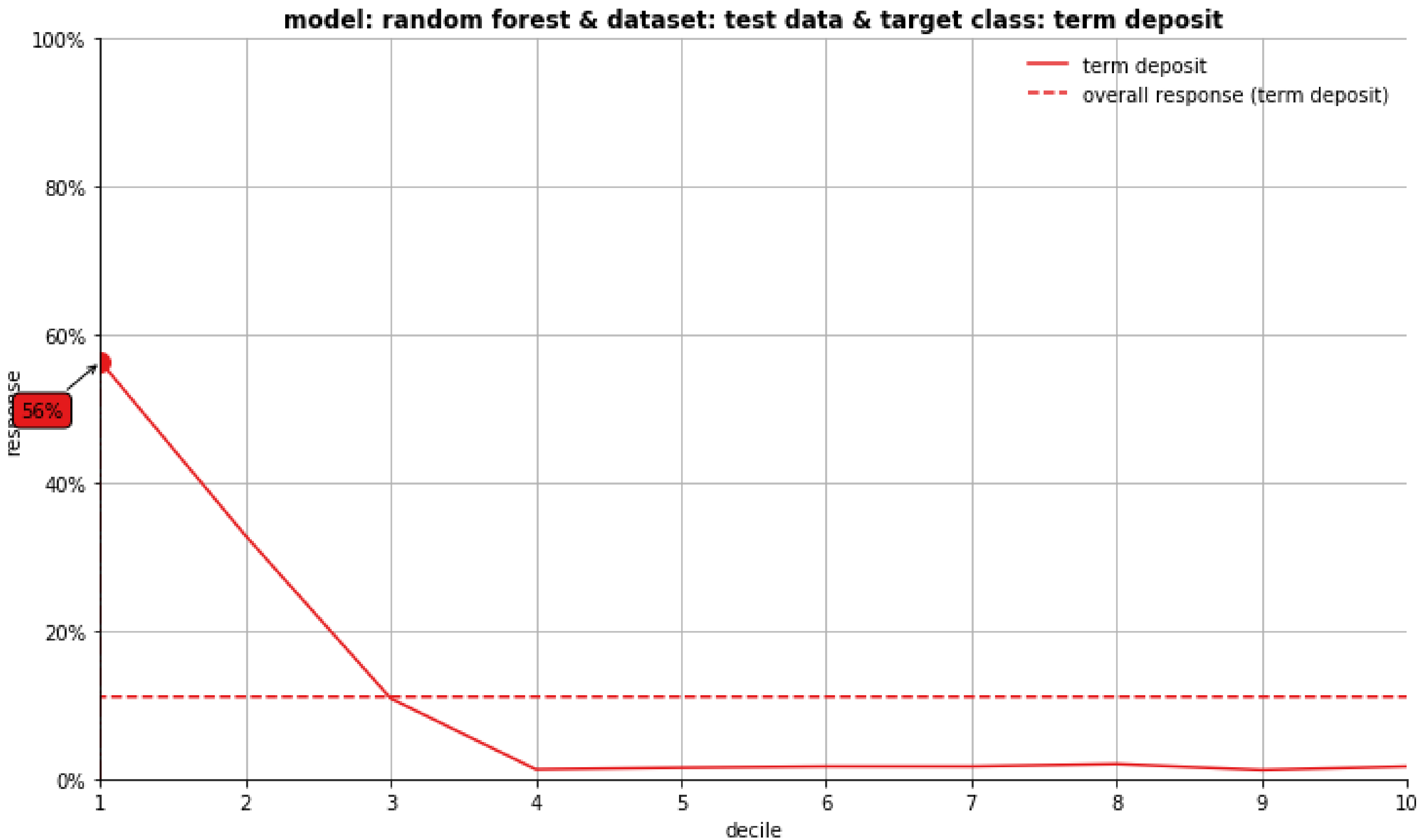


https://modelplot.github.io/intro_modelplotpy.html

How many of the observations belong to the target in a specific decile?

Example

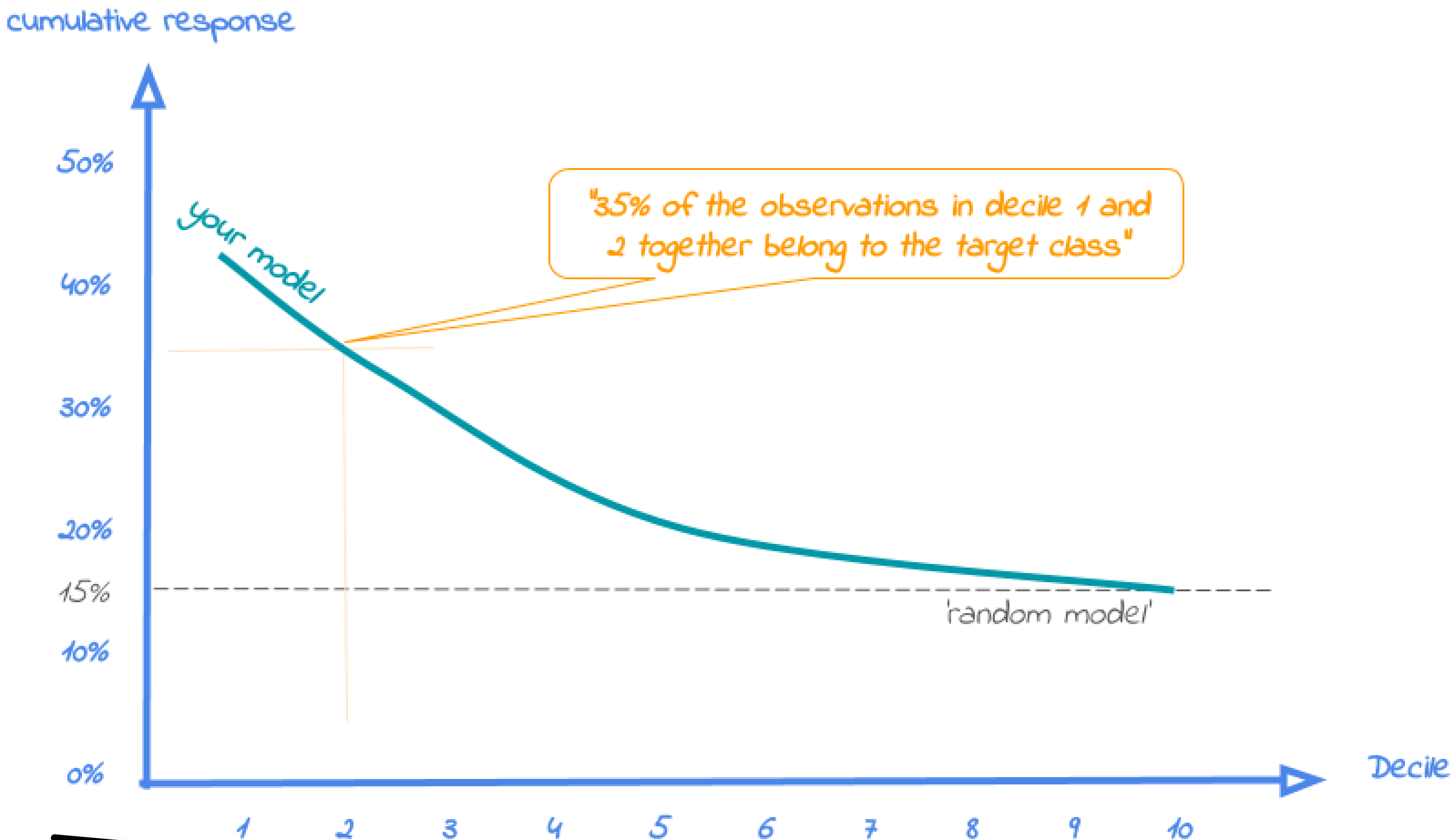
Response



https://modelplot.github.io/intro_modelplotpy.html

When we select decile 1 from model random forest in dataset test data the percentage of term deposit cases in the selection is 56%.

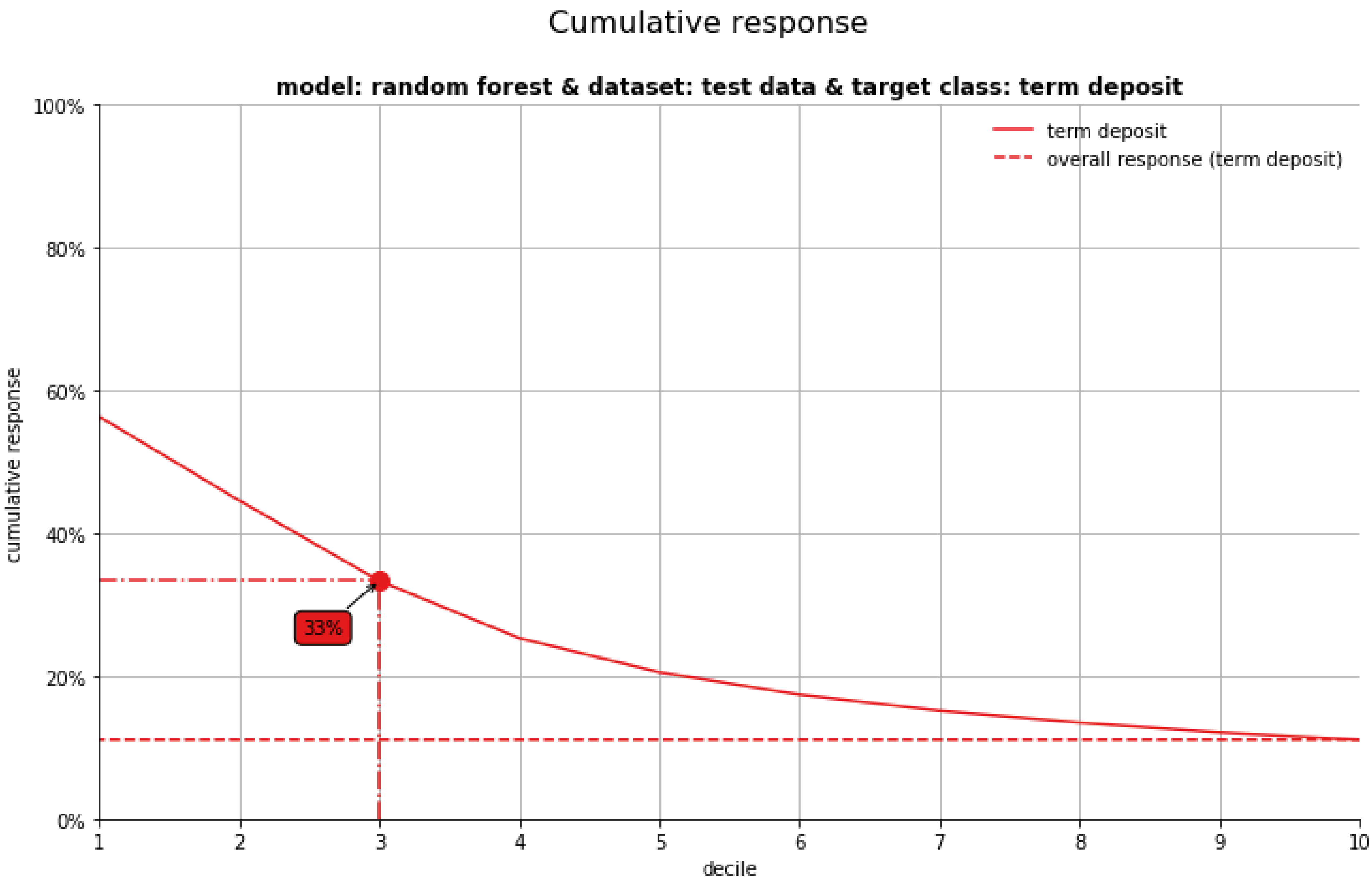
Cumulative response



https://modelplot.github.io/intro_modelplotpy.html

If we select the top N deciles, how many observations will belong to target class?

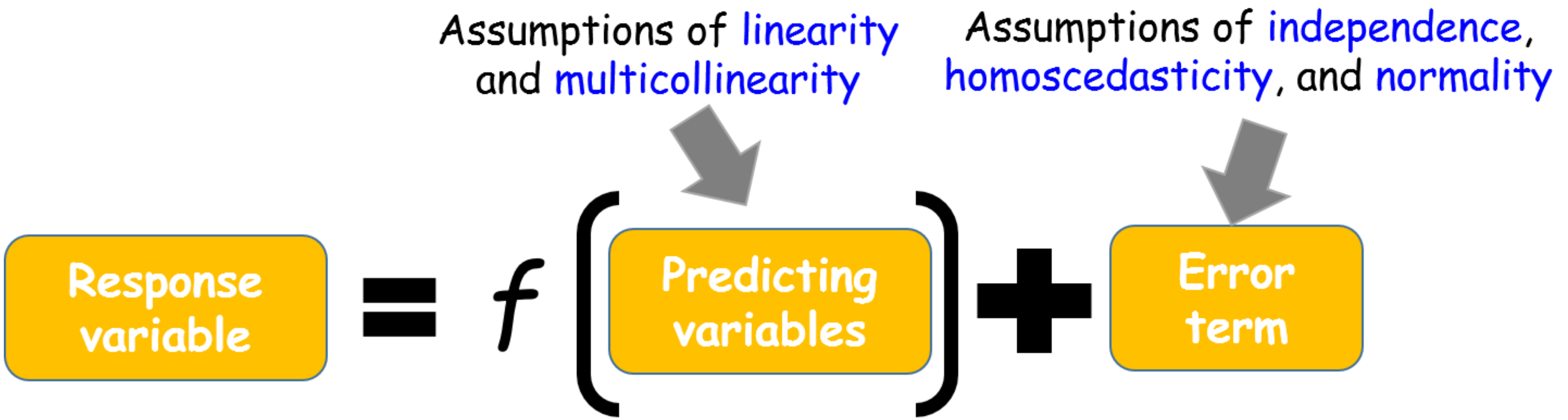
Example



When we select deciles 1 until 3 according to model random forest in dataset test data the percentage of term deposit cases in the selection is 33%.

https://modelplot.github.io/intro_modelplotpy.html

Reminder: regression

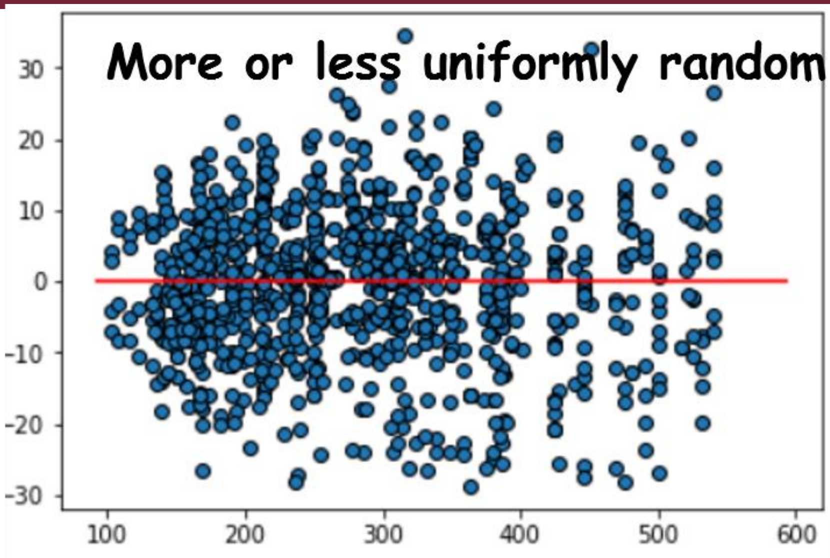


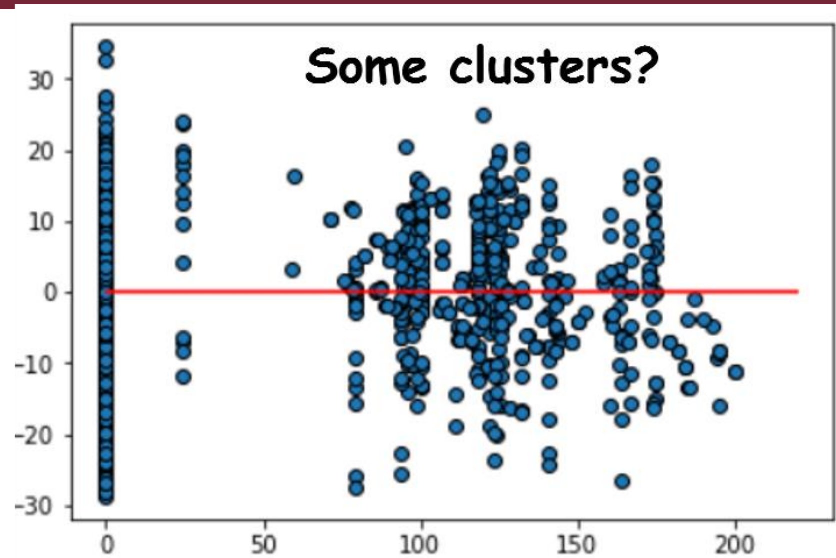
Reminder: statsmodel in Python

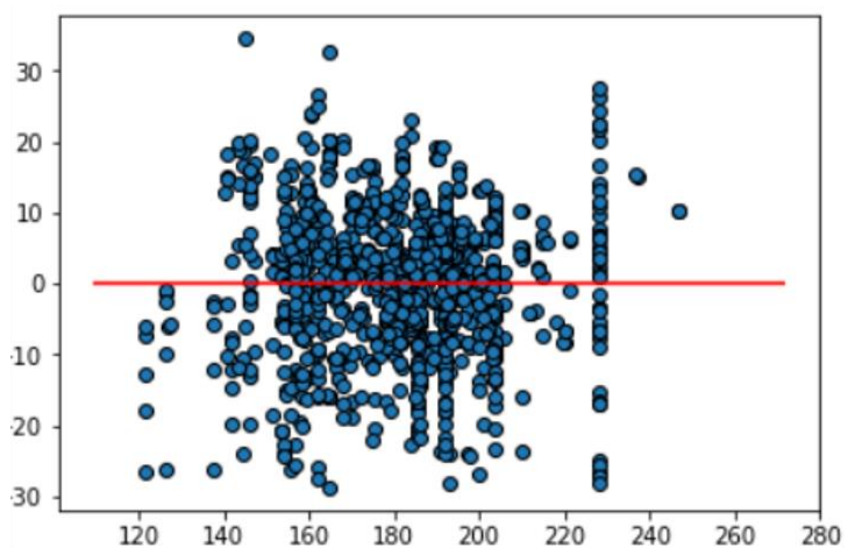
Assumptions

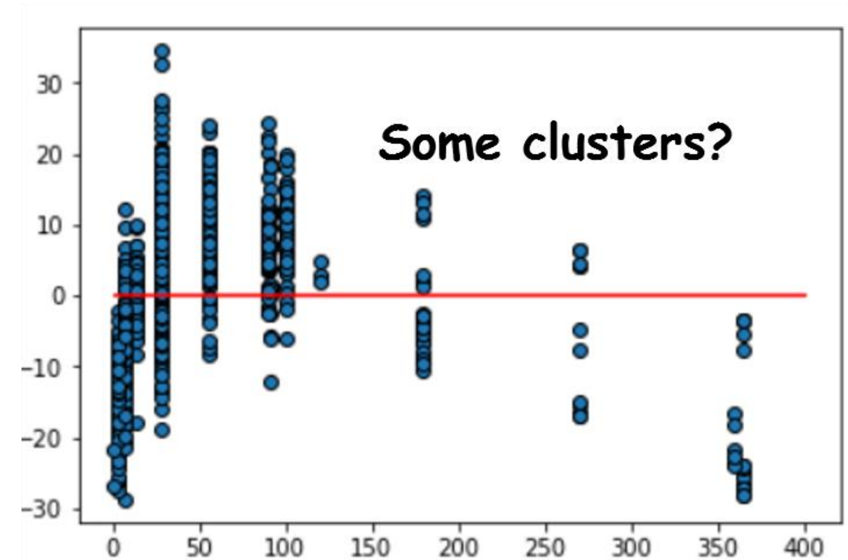
- Linearity
- Independence
- Constant variance (homoscedasticity)
- Normal distribution of errors

Analysing residuals (errors)



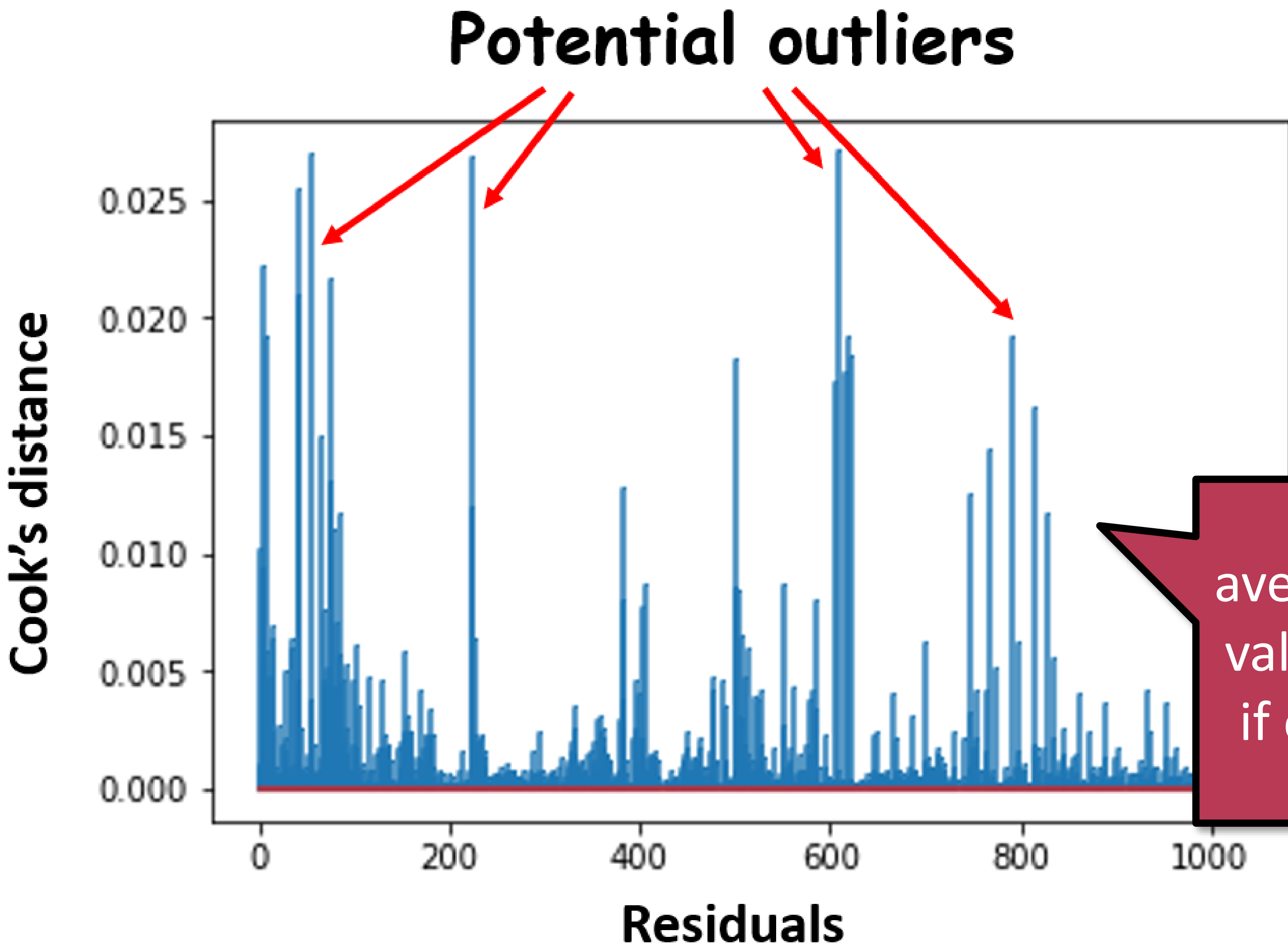






https://towardsdatascience.com/how-do-you-check-the-quality-of-your-regression-model-in-python-fa61759ff685

Cook's distance



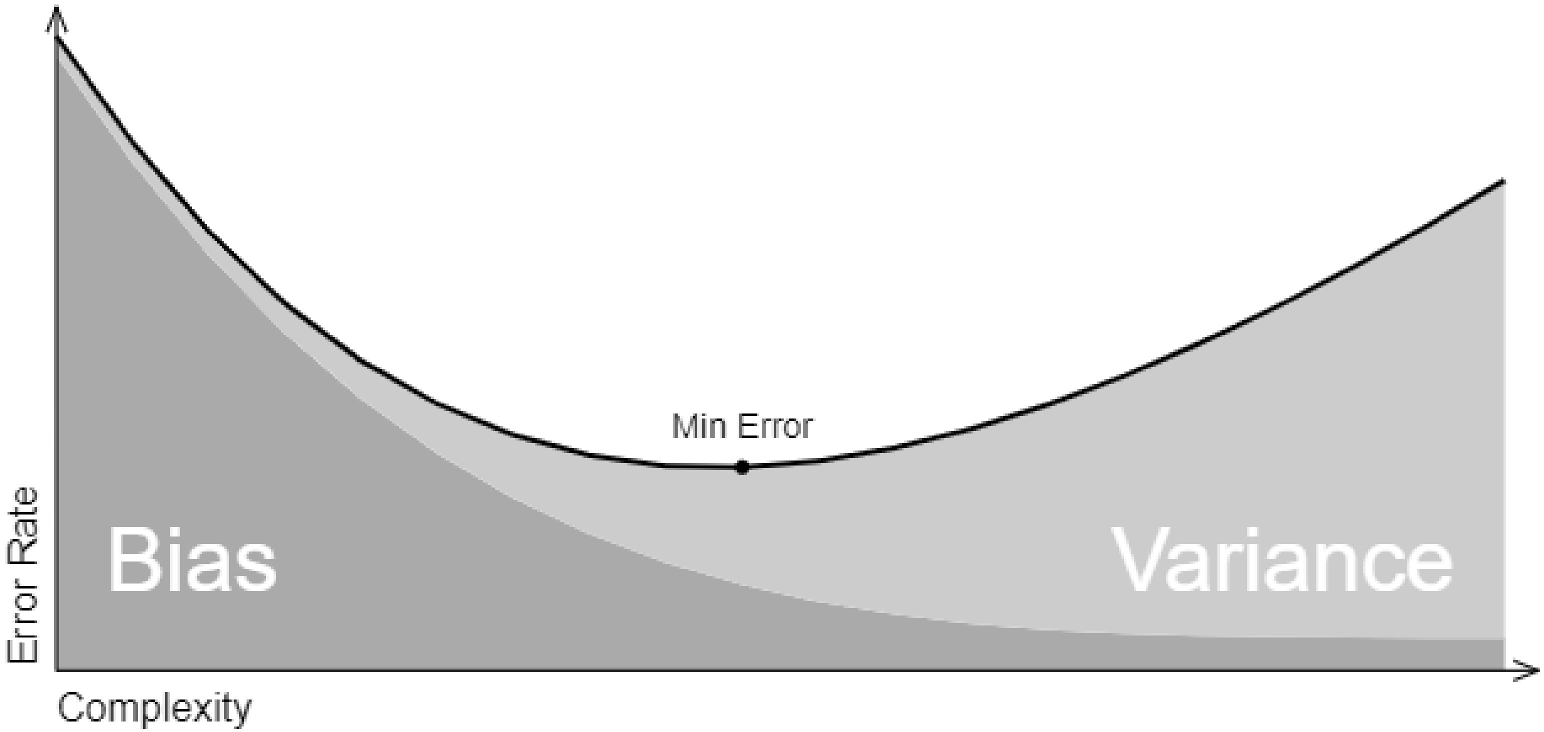
How far an average predicted value will change if observation is dropped?

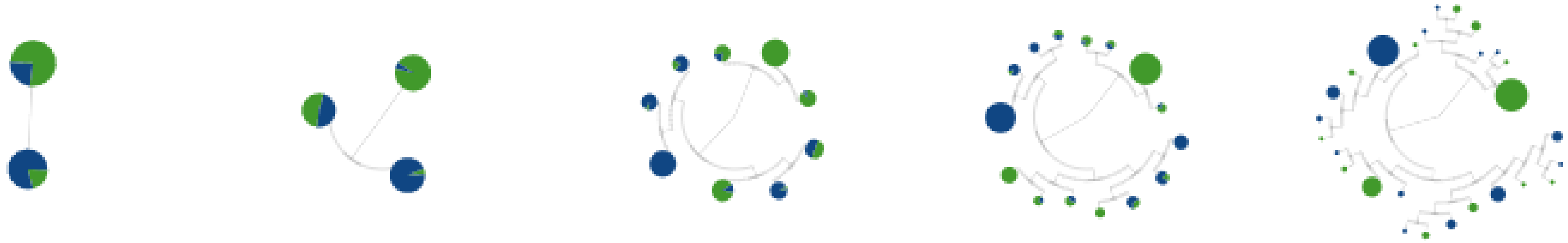
https://towardsdatascience.com/how-do-you-check-the-quality-of-your-regression-model-in-python-fa61759ff685

DECISION TREES

http://www.r2d3.us/visual-intro-to-machine-learning-part-2/?from=@

Bias vs variance

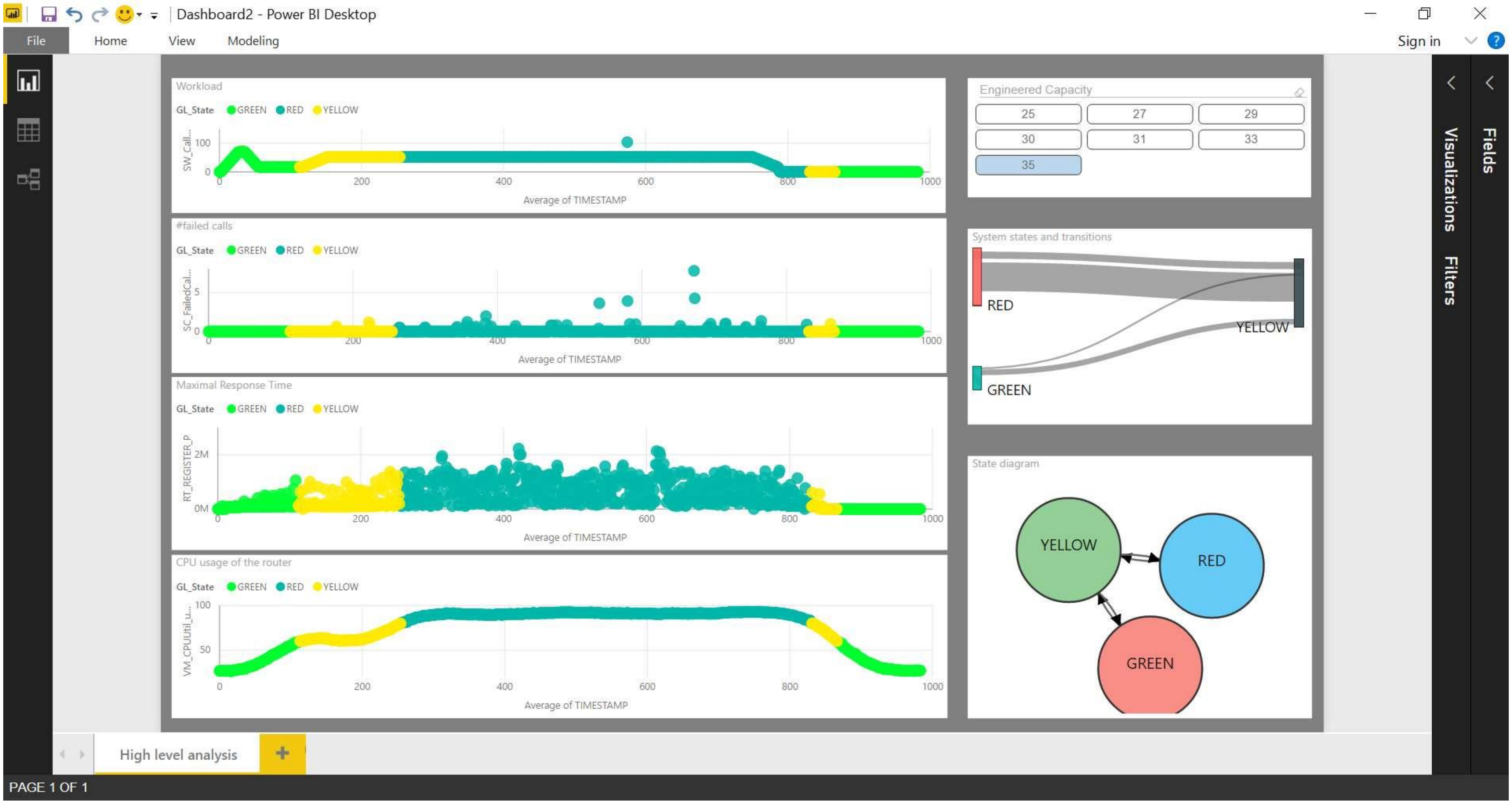




http://www.r2d3.us/visual-intro-to-machine-learning-part-2/?from=@

OUTLOOK FOR NEXT CLASS

Examples: bad visualizations



Examples: bad visualizations