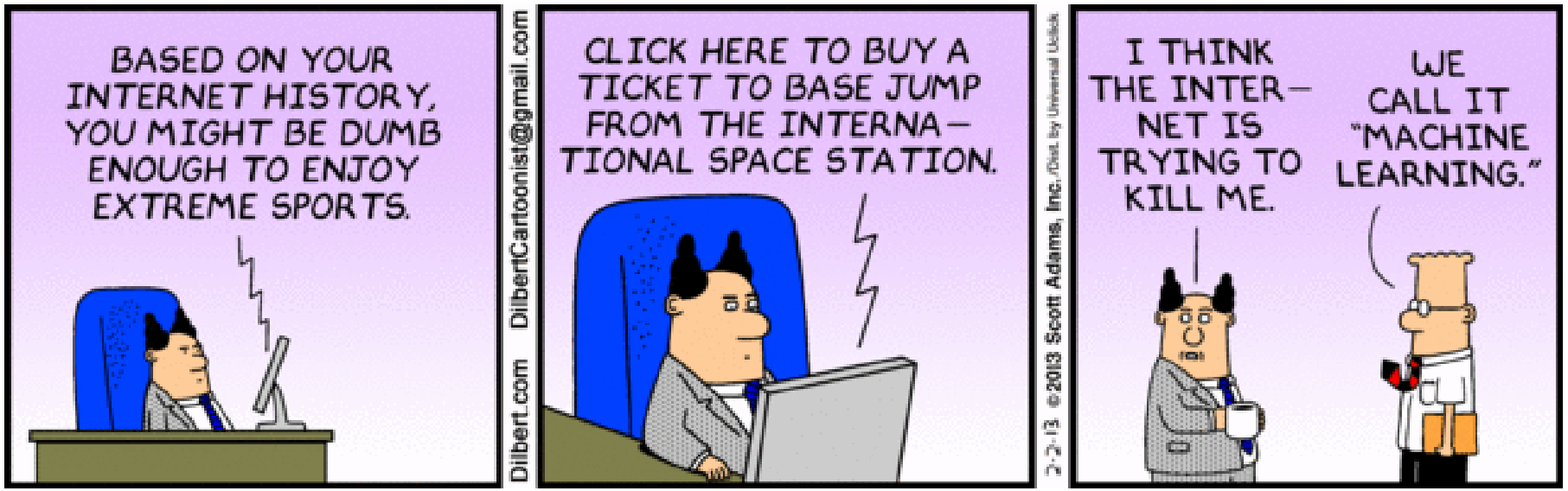


http://dilbert.com/

Harikrishna Narasimhan

Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Learning from 'Binary Labels':

A Plethora of Performance Measures, A Plethora of Algorithms Needed

A Software Engineer's Plight to Build a Machine Learning System that Works!

Machine Learning

Machine Learning

Data Mining

Machine Learning

Data Mining

Big Data! ☺

Machine Learning

Data Mining

Big Data! ☺

Predictive Analytics

....

Spam Filter

Google in:spam

Gmail ▾ | Delete forever | Not spam | More ▾

COMPOSE

Inbox (12)
Chats
Sent Mail
Drafts (229)
Spam (176)
Circles
Personal
Project K
Work
More ▾

Time travelers PLEASE HELP!!!!!! Spam x

<xxxxxxx@yyy.com> 10:29 PM (1 minute ago)

to me

If you are a time traveler or alien disguised as human and or have the technology to travel physically through time I need your help!

My life has been severely tampered with and cursed!!

I have suffered tremendously and am now dying!

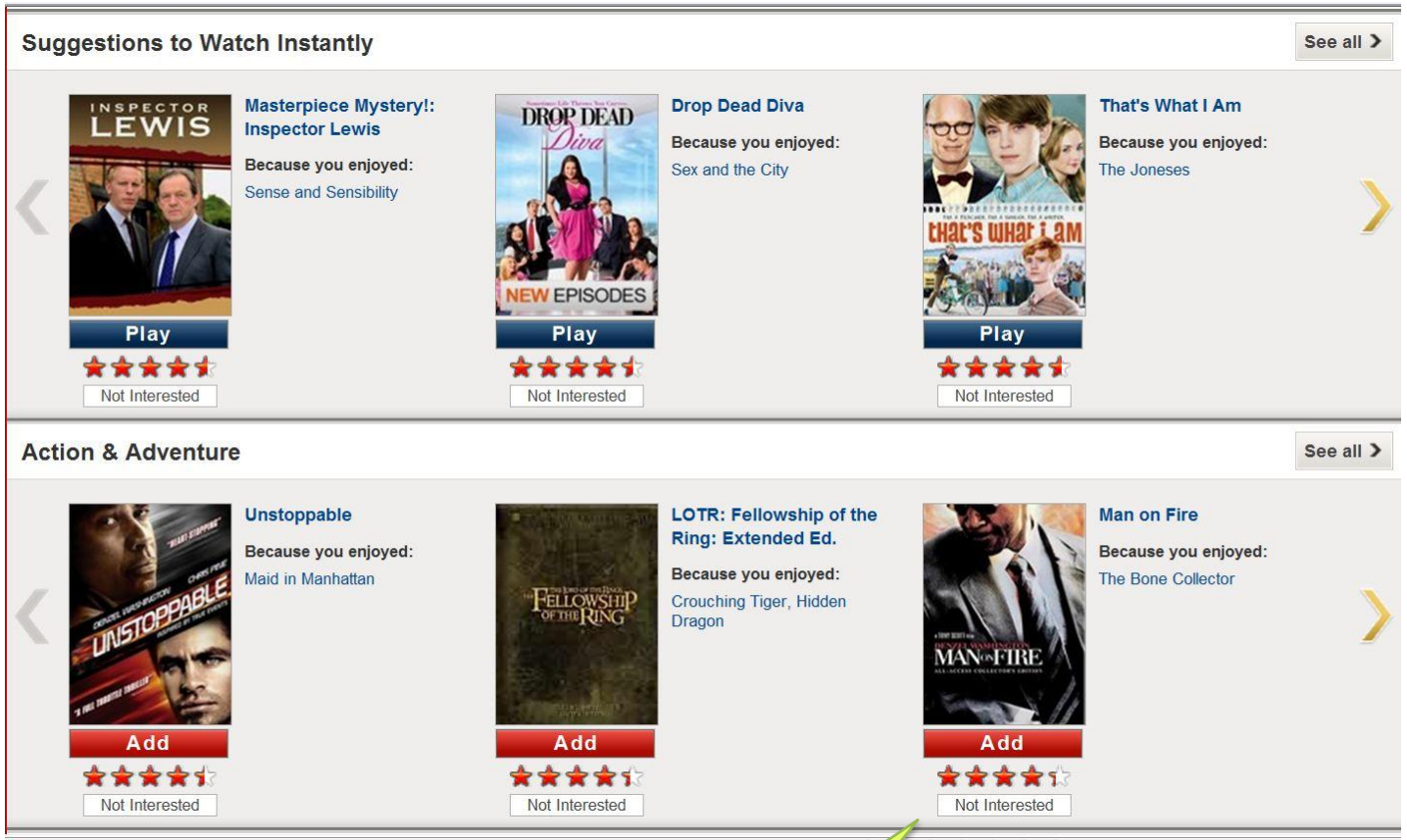
I need to be able to:

Travel back in time.

Spam or Non-spam?

http://gmail.com/

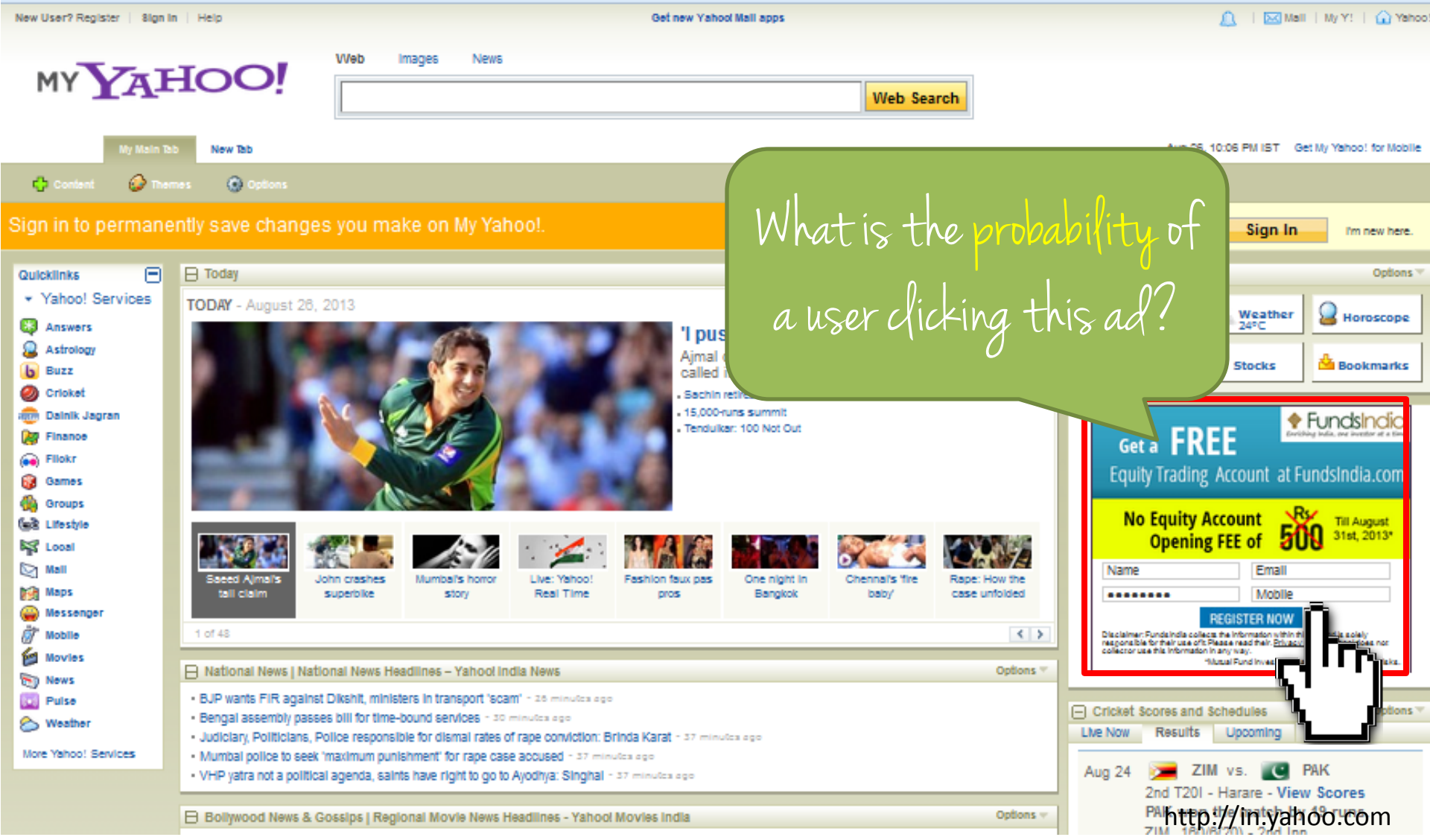
Movie Recommendation



What rating will user X give movie Y?

http://techstuff123.wordpress.com/
http://www.netflix.com/

Click Prediction

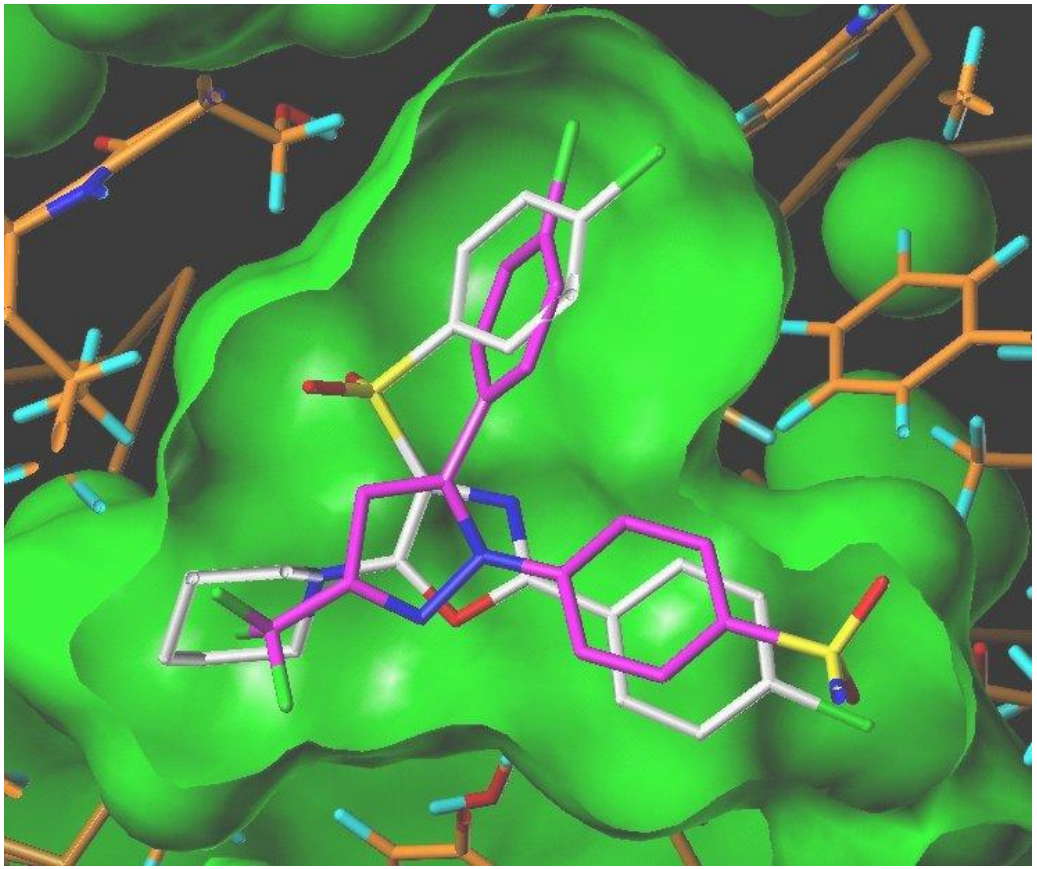


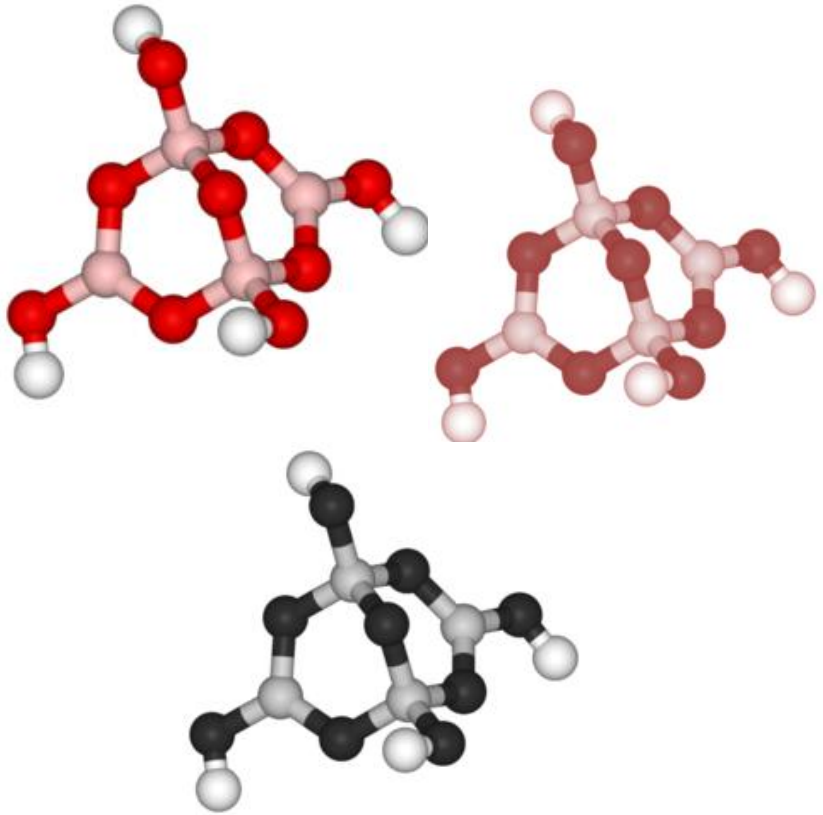
Hospital Readmission

What is the probability of a patient being readmitted?

http://www.nytimes.com/

Drug Discovery

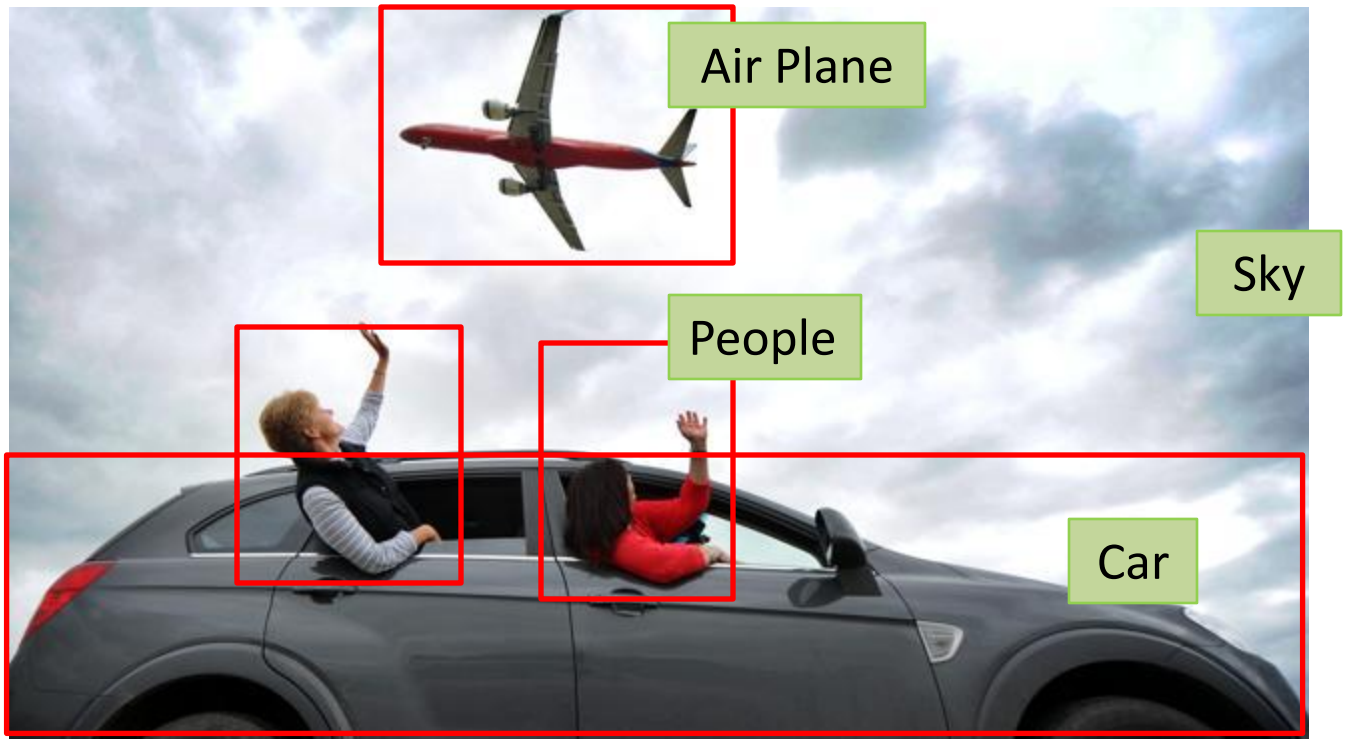




Millions of Chemical Compounds!!

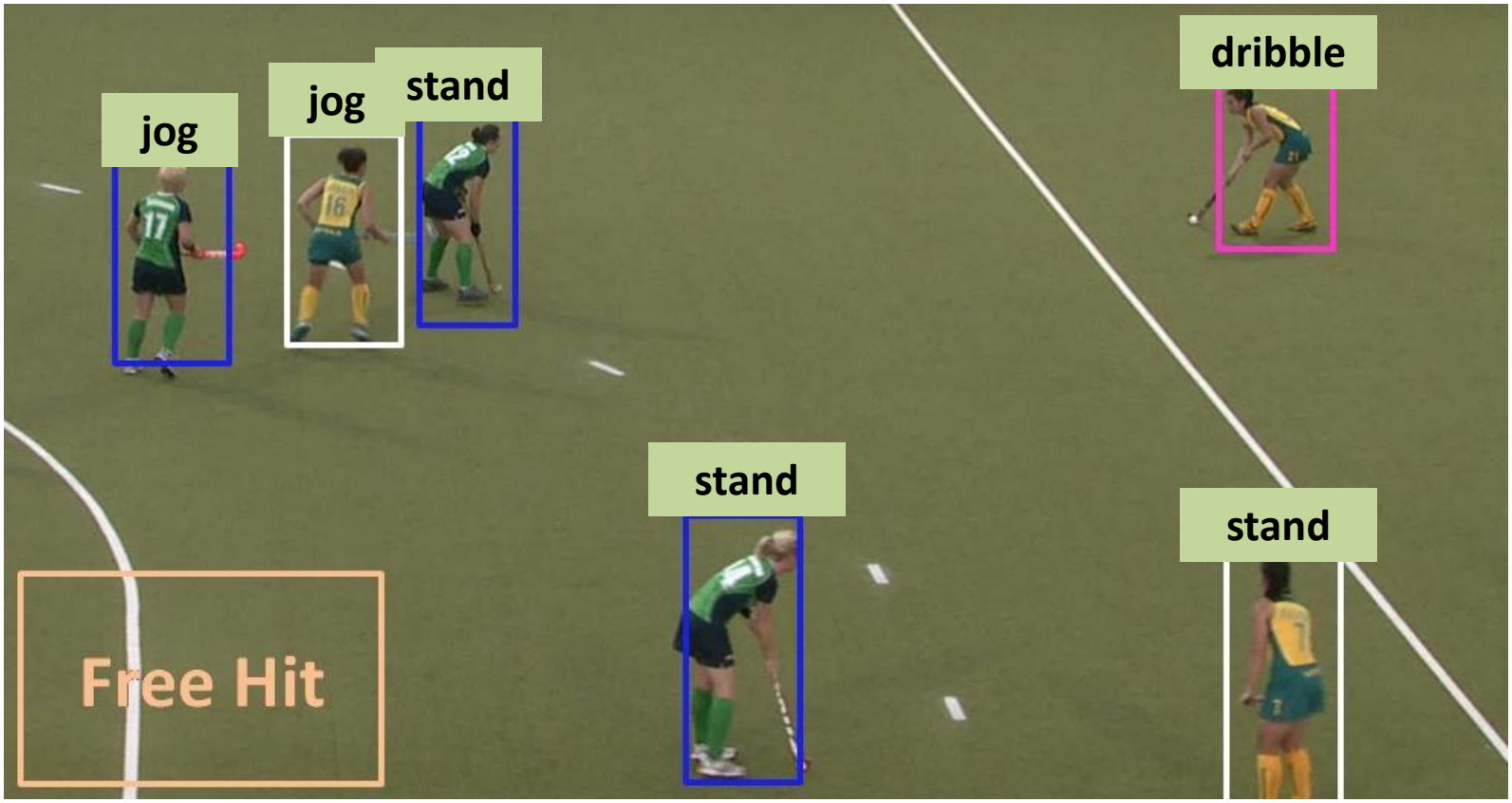
http://www.google.co.in/imghp

Image Annotation

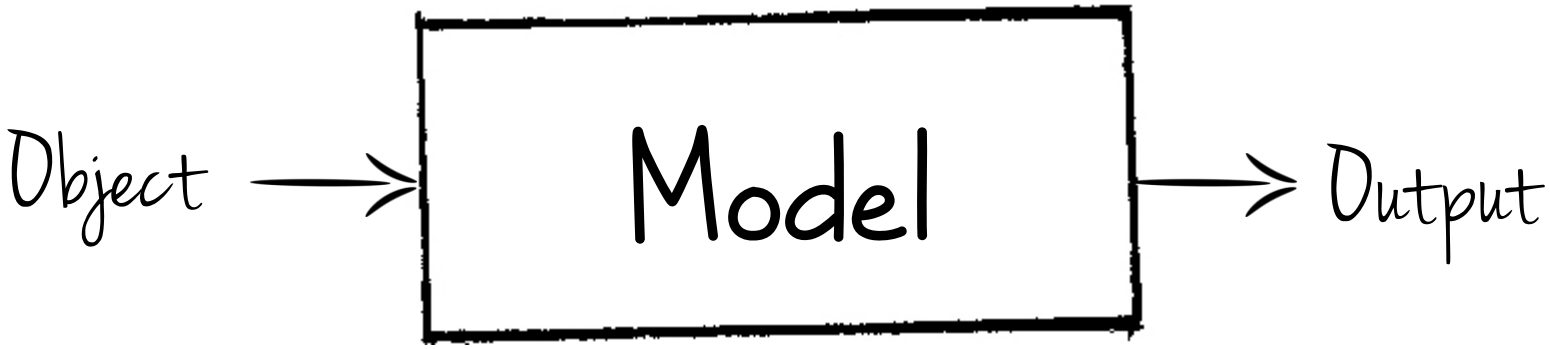


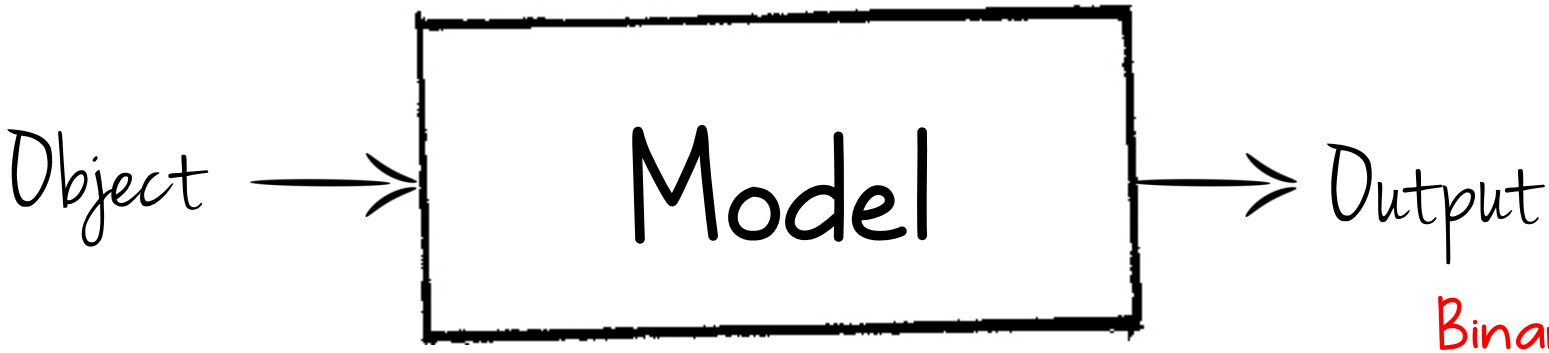
http://www.heraldsun.com.au

Activity Recognition

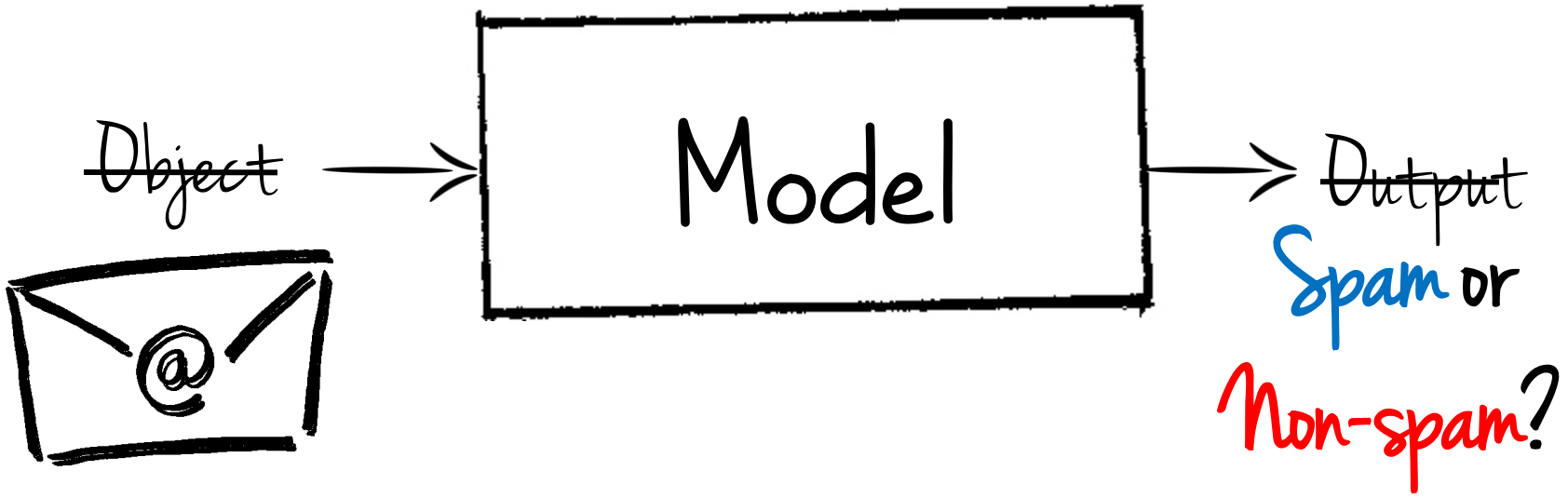


http://cs.brown.edu/~ls


Object
Model
Output

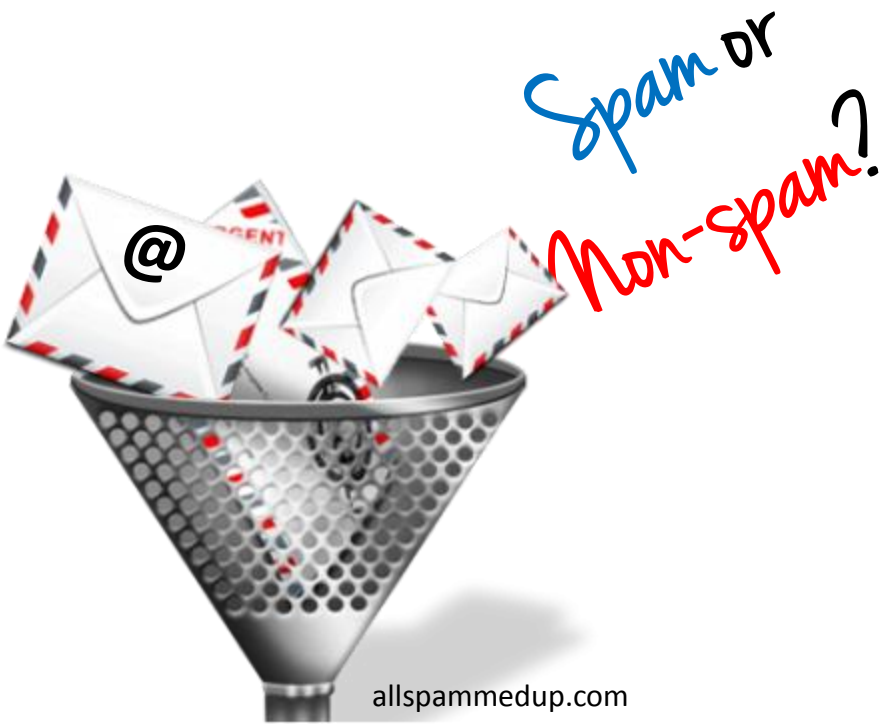


Binary
Categorical
Real Valued
Multi-label
Structured
....


~~Object~~
Model
~~Output~~
Spam or
Non-spam?

Learning from 'Binary Labels'

Spam or Non-spam?


Spam or
Non-spam?
@
allspammedup.com


Disease or
No Disease?
pascal-network.org

Spam or Non-spam?

allspammedup.com

Disease or No Disease?

pascal-network.org

Relevant or Irrelevant?

fusionsedge.com

Spam or Non-spam?

allspammedup.com

Disease or No Disease?

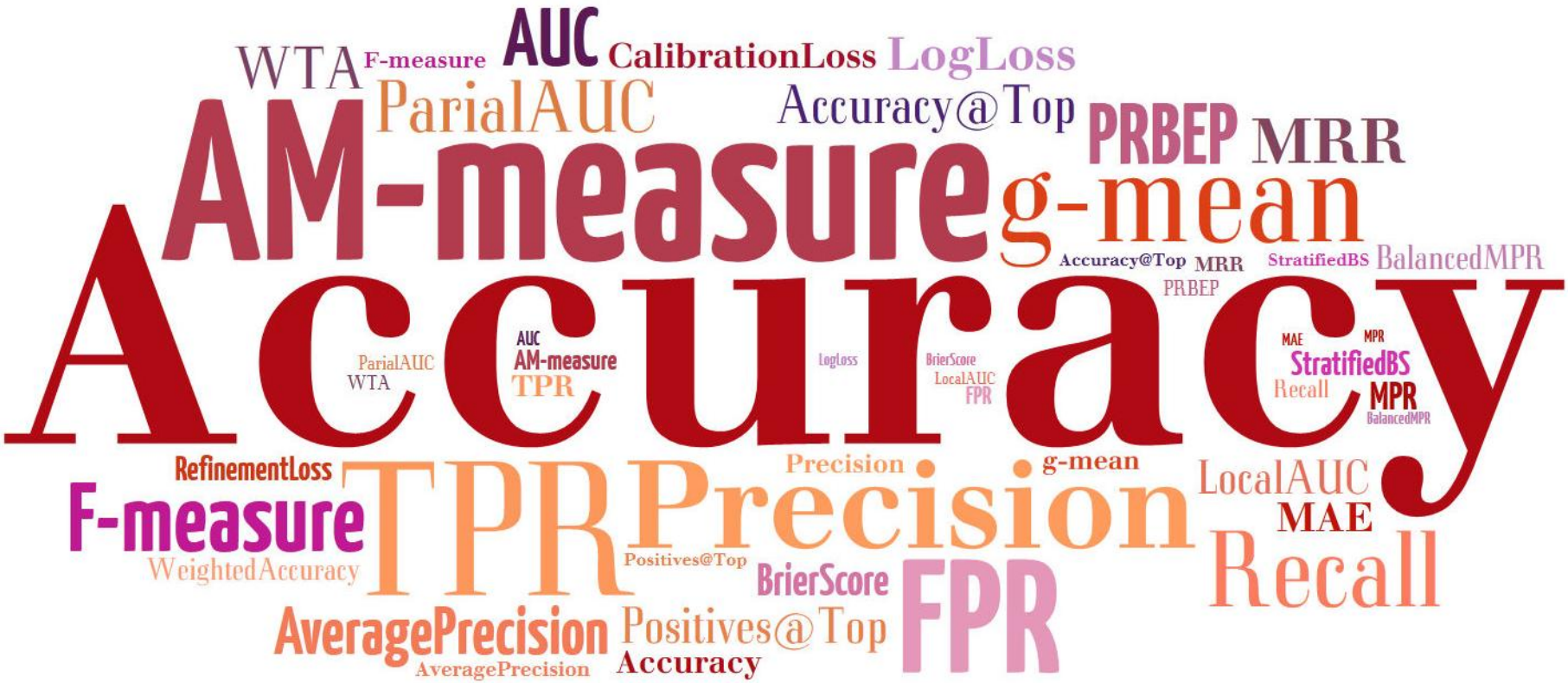
pascal-network.org

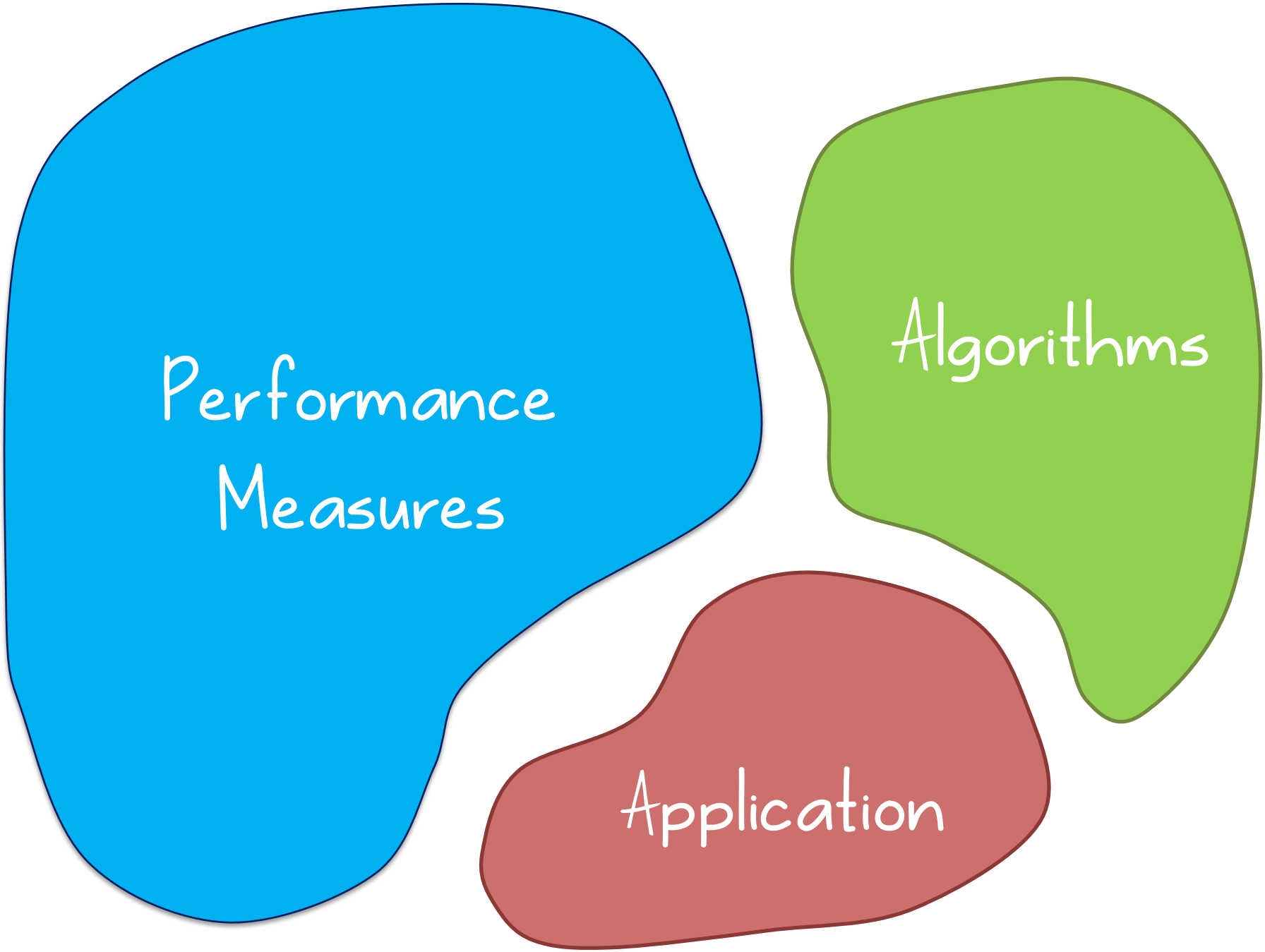
Relevant or Irrelevant?

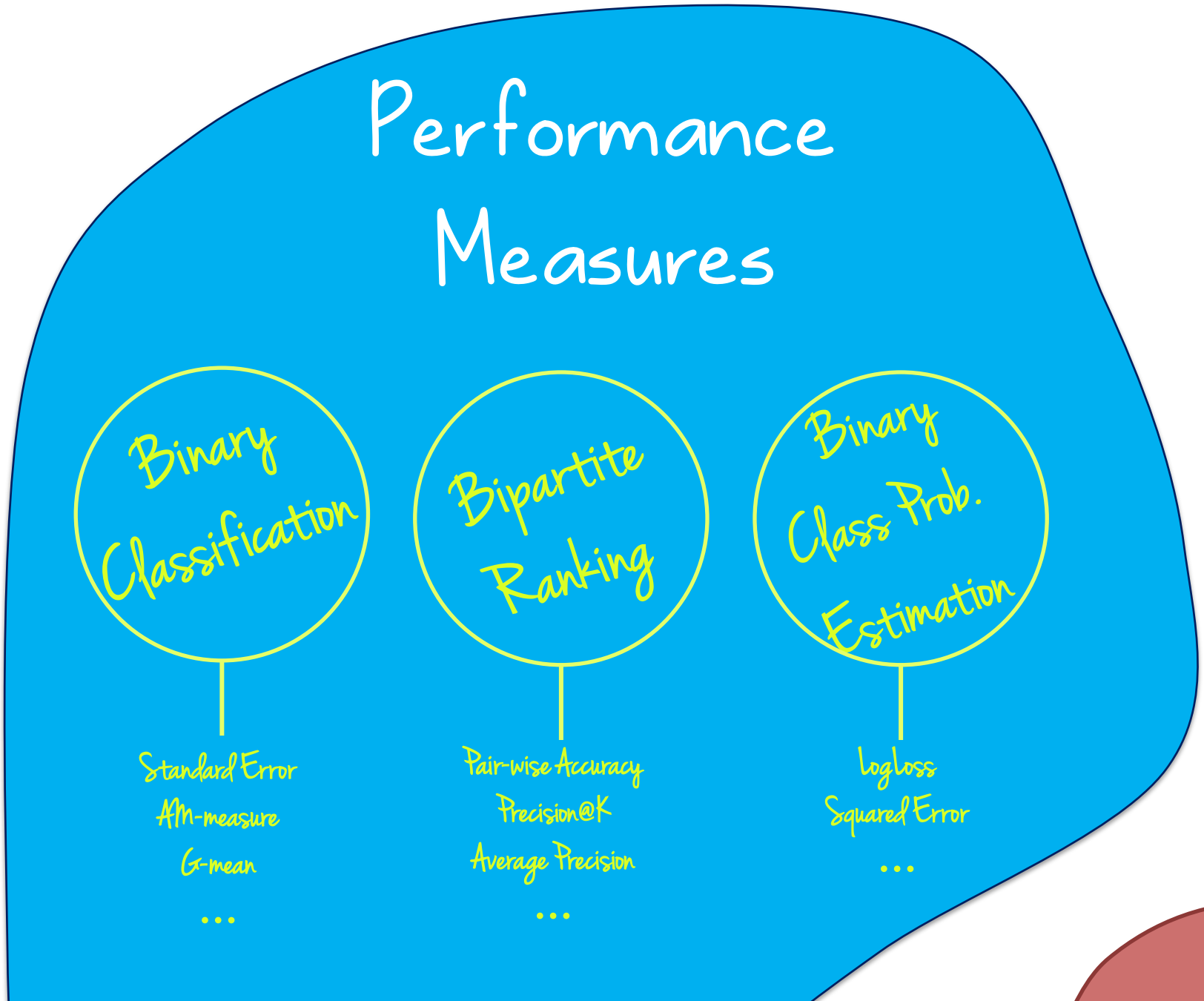
fusionsedge.com

Click or No Click?

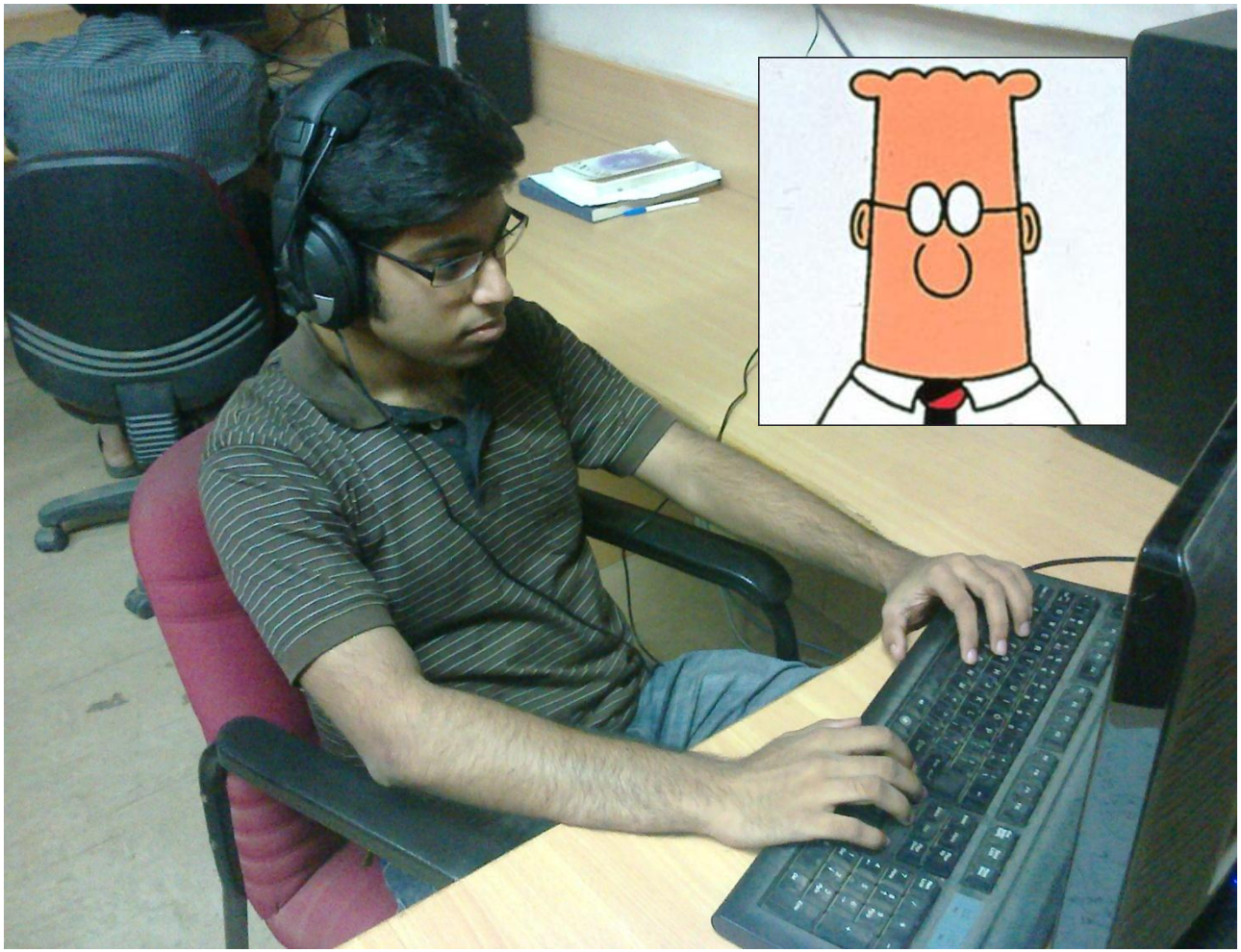
optimum7.com



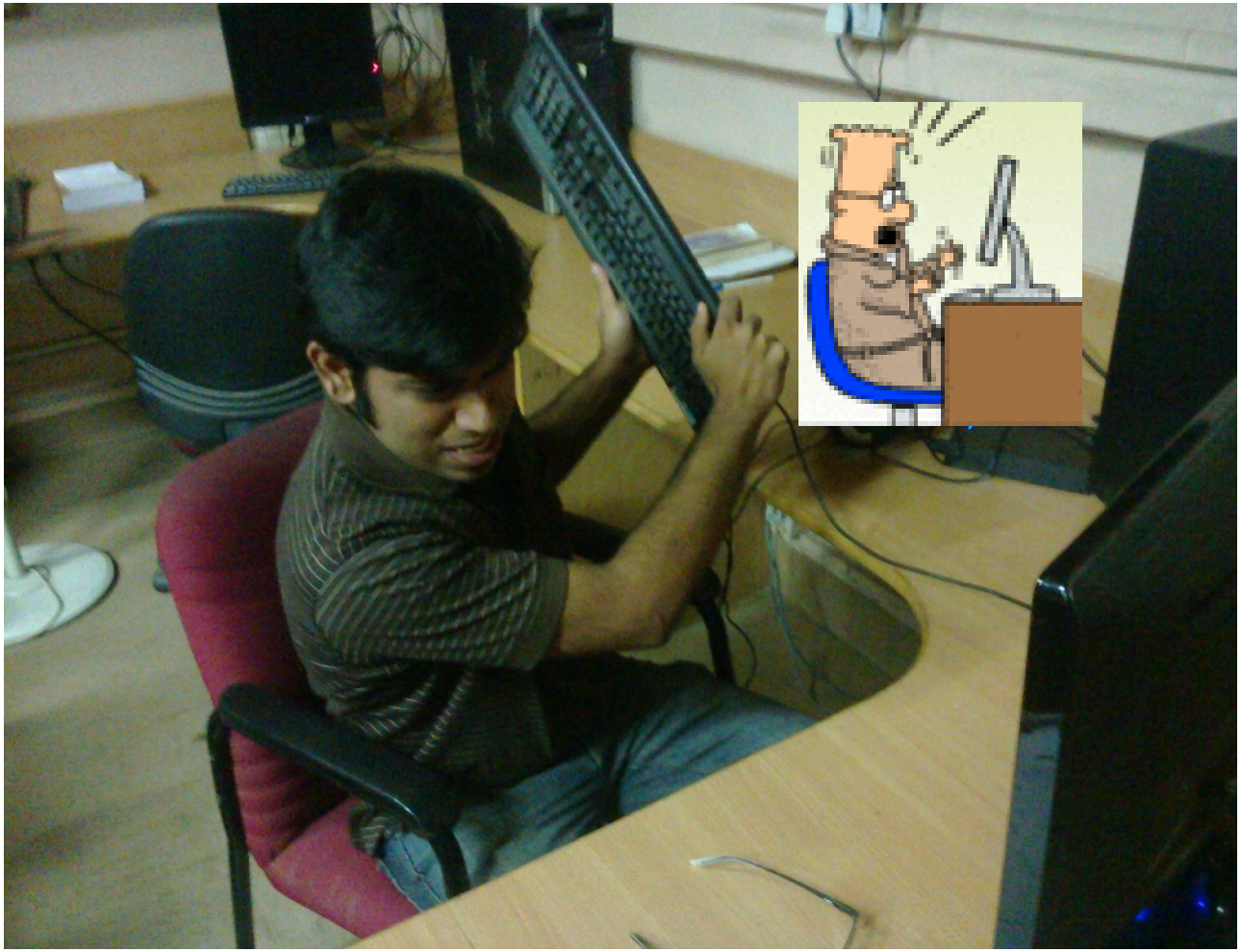

Performance Measures
Algorithms
Application


Performance Measures
Binary Classification
Standard Error
AM-measure
G-mean
...
Bipartite Ranking
Pair-wise Accuracy
Precision@K
Average Precision
...
Binary Class Prob. Estimation
Logloss
Squared Error
...

Our 'Desi' Dilbert Then



Our 'Desi' Dilbert Now




INSURANCE
CLAIM FORM
REJECTED

Genuine

Fraud

.....

Genuine

Fraud or

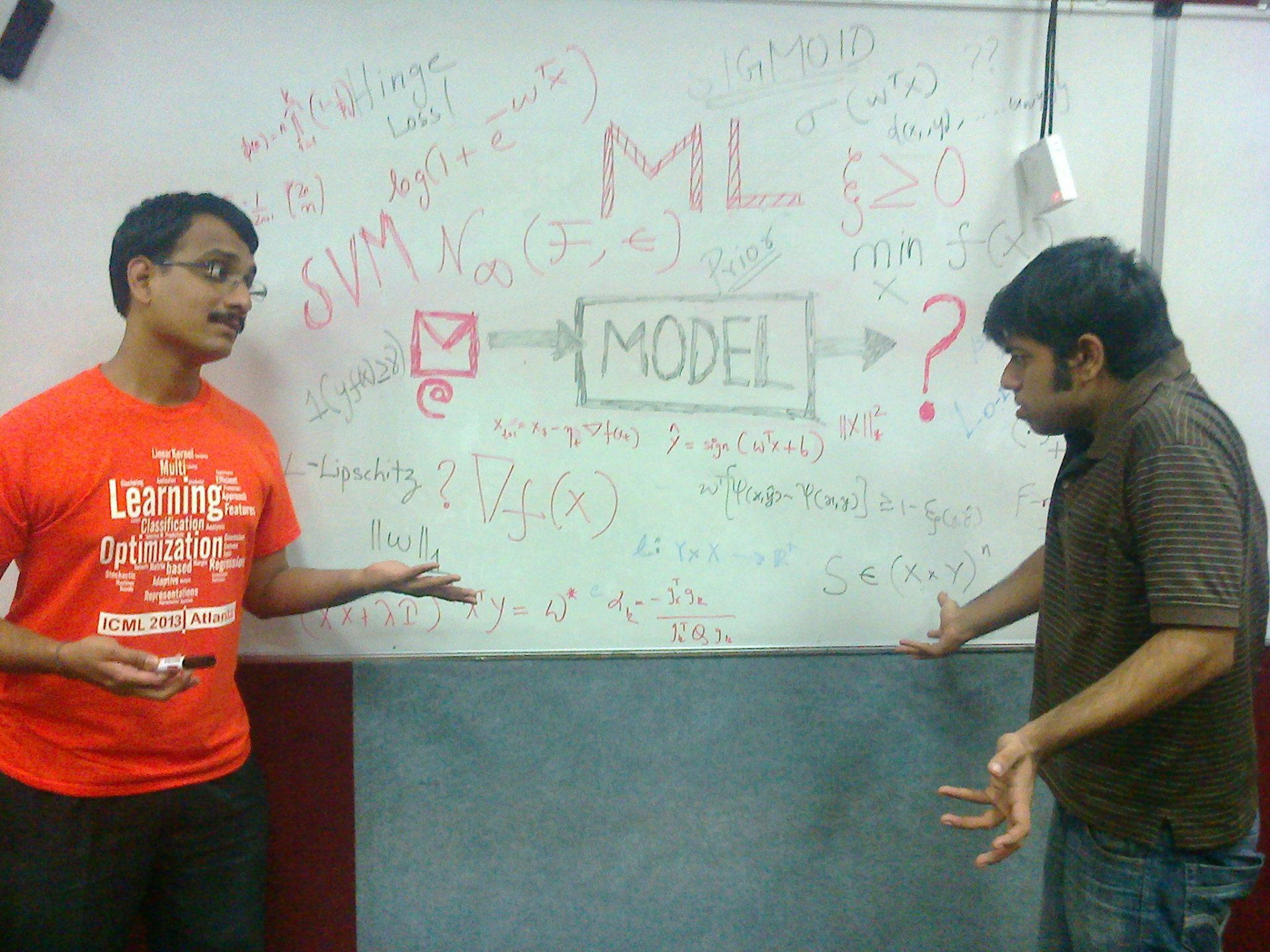
Genuine ?


ML
MODEL
Learning
Optimization


?#?#!!??
Surrogate Risk Minimization with a classification-calibrated loss function and appropriate regularization
Learning
Optimization
Classification
Multi
Features
Regression
Representations
ICML
Atlanta

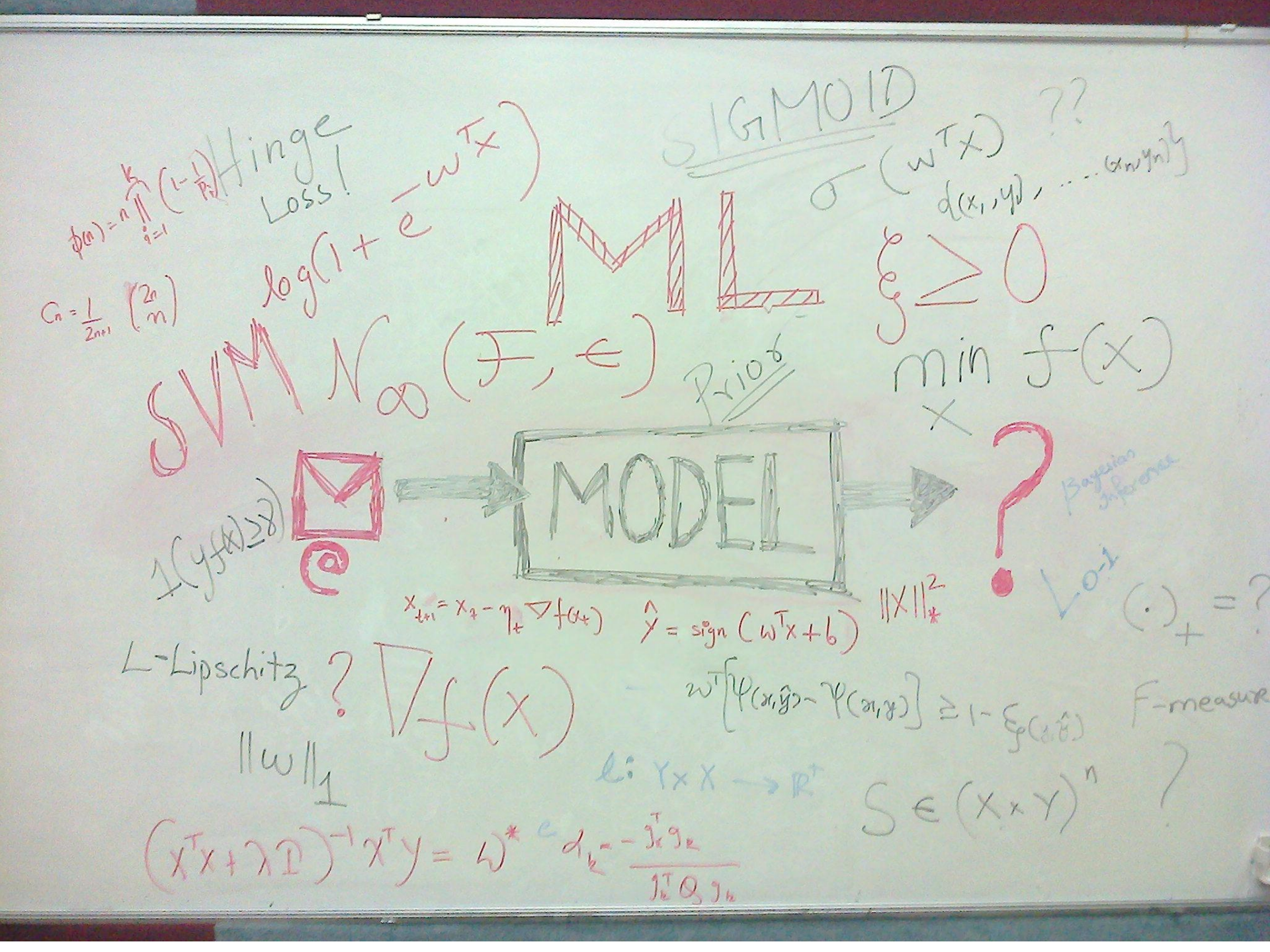

ML
MODEL
Learning
Optimization

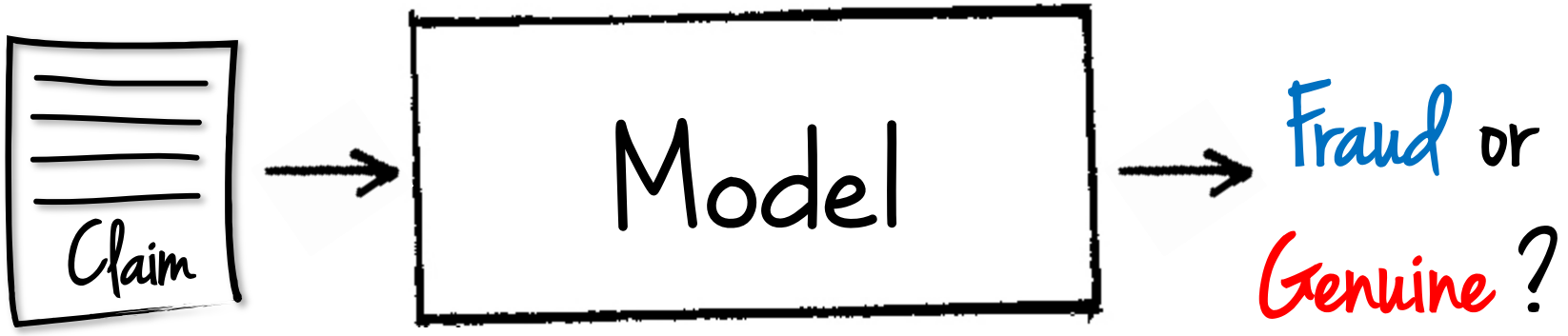

ML
MODEL
?
Learning
Optimization
ICML 2013

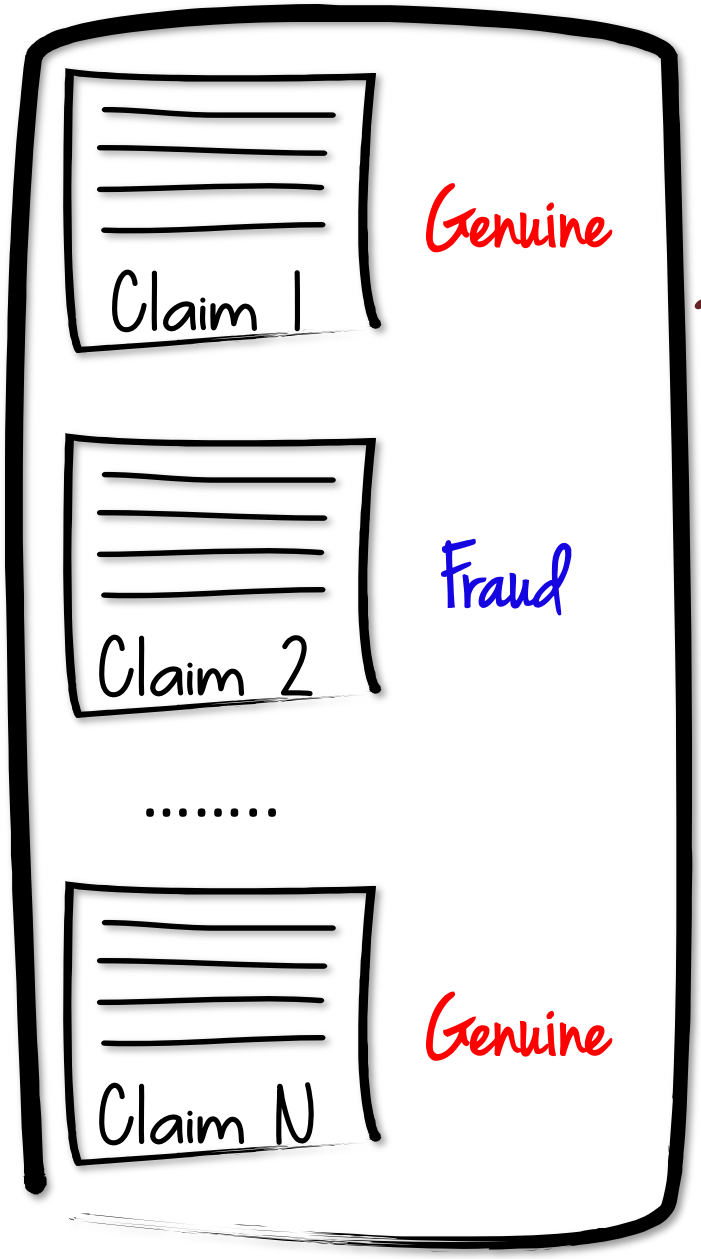

Learning
Optimization
ICML 2013 Atlanta
MODEL
ML
SVM
SIGMOID
Hinge Loss
?


ML
SVM
MODEL
SIGMOID
Learning
Optimization

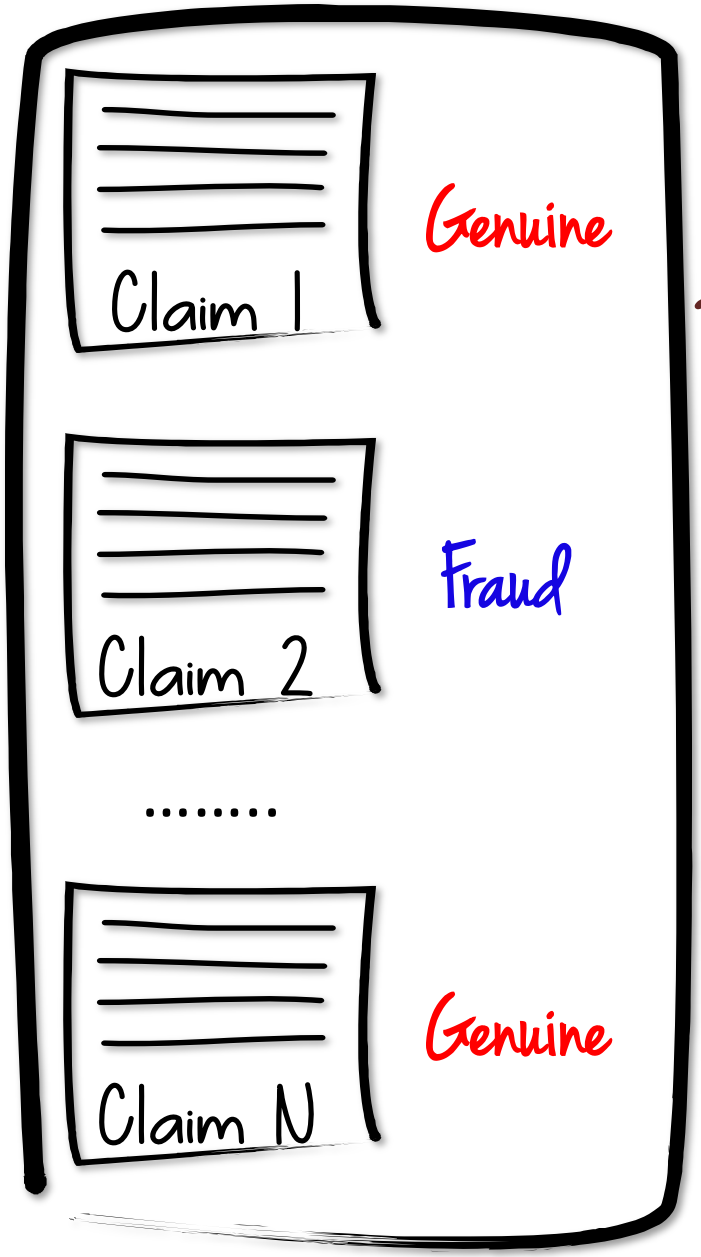

SVM
ML
SIGMOID
MODEL
?
Learning
Optimization
ICML 2013


SIGMOID
Hinge Loss!
$\log(1+e^{-w^T x})$
$\sigma(w^T x)$??
$d(x_1, y_1), \ldots (x_n, y_n)\}$
$\xi \geq 0$
$\min f(x)$
$\times$
SVM
$N_\infty(\mathcal{F}, \epsilon)$
ML
Priors
MODEL
?
$\mathbb{1}(y f(x) \geq 0)$
$x_{t+1} = x_t - \eta_t \nabla f(x_t)$
$\hat{y} = \text{sign}(w^T x + b)$
$\|X\|_*^2$
Bayesian Inference
L-Lipschitz ?
$\nabla f(x)$
$\|w\|_1$
$w^T[\psi(x,\hat{y}) - \psi(x,y)] \geq 1 - \xi_{(x,\hat{y})}$
$\ell: Y \times X \to \mathbb{R}^+$
$S \in (X \times Y)^n$?
$(X^T X + \lambda I)^{-1} X^T y = w^*$
$d_k = -\frac{J_k^T g_k}{J_k^T Q_k J_k}$
$(\cdot)_+ = ?$
F-measure


Claim
Model
Fraud or
Genuine ?



Training Set



Training Set

Goal?

Learn a model with 'minimum error' on training set!

Model

Age

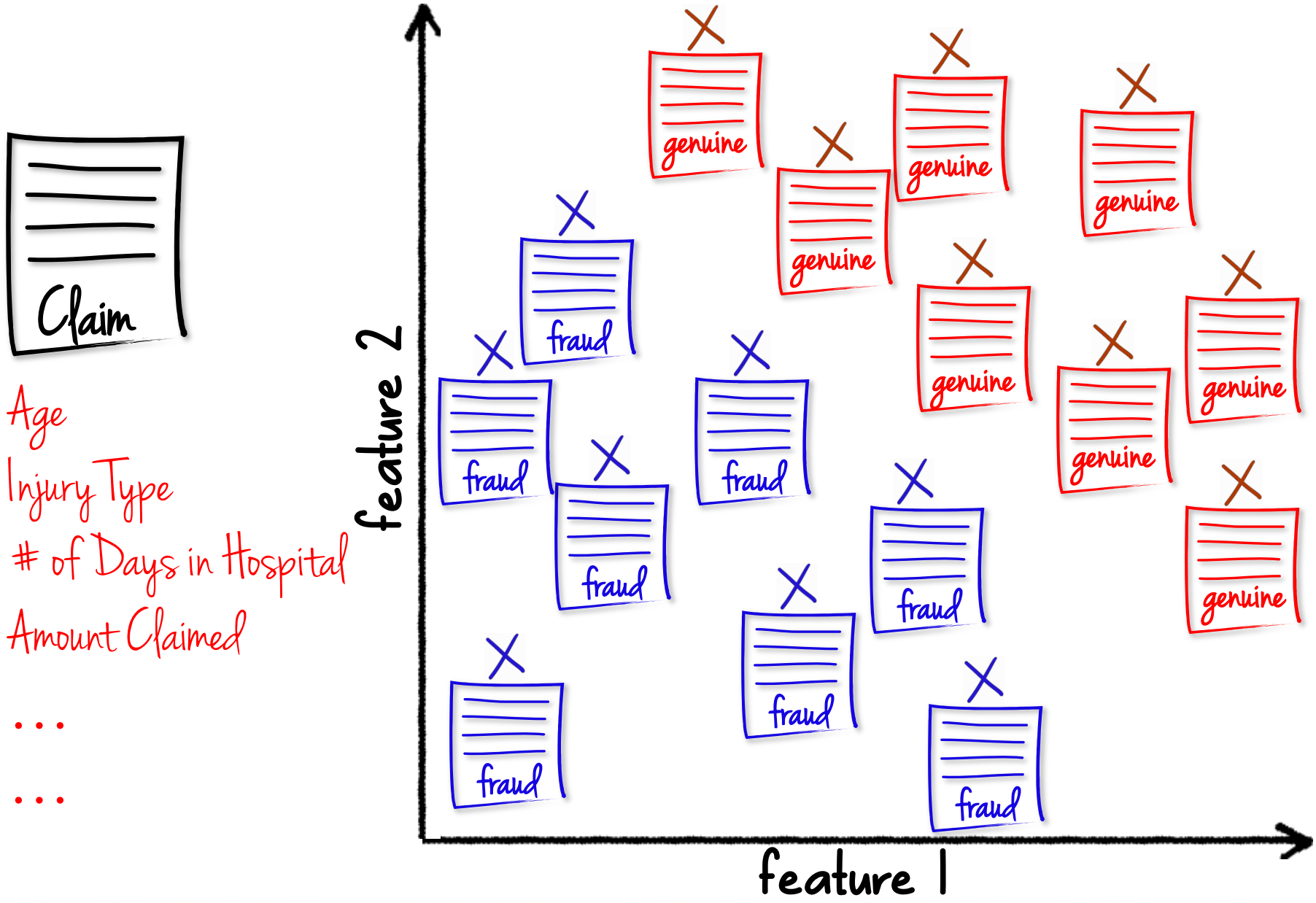
Injury Type

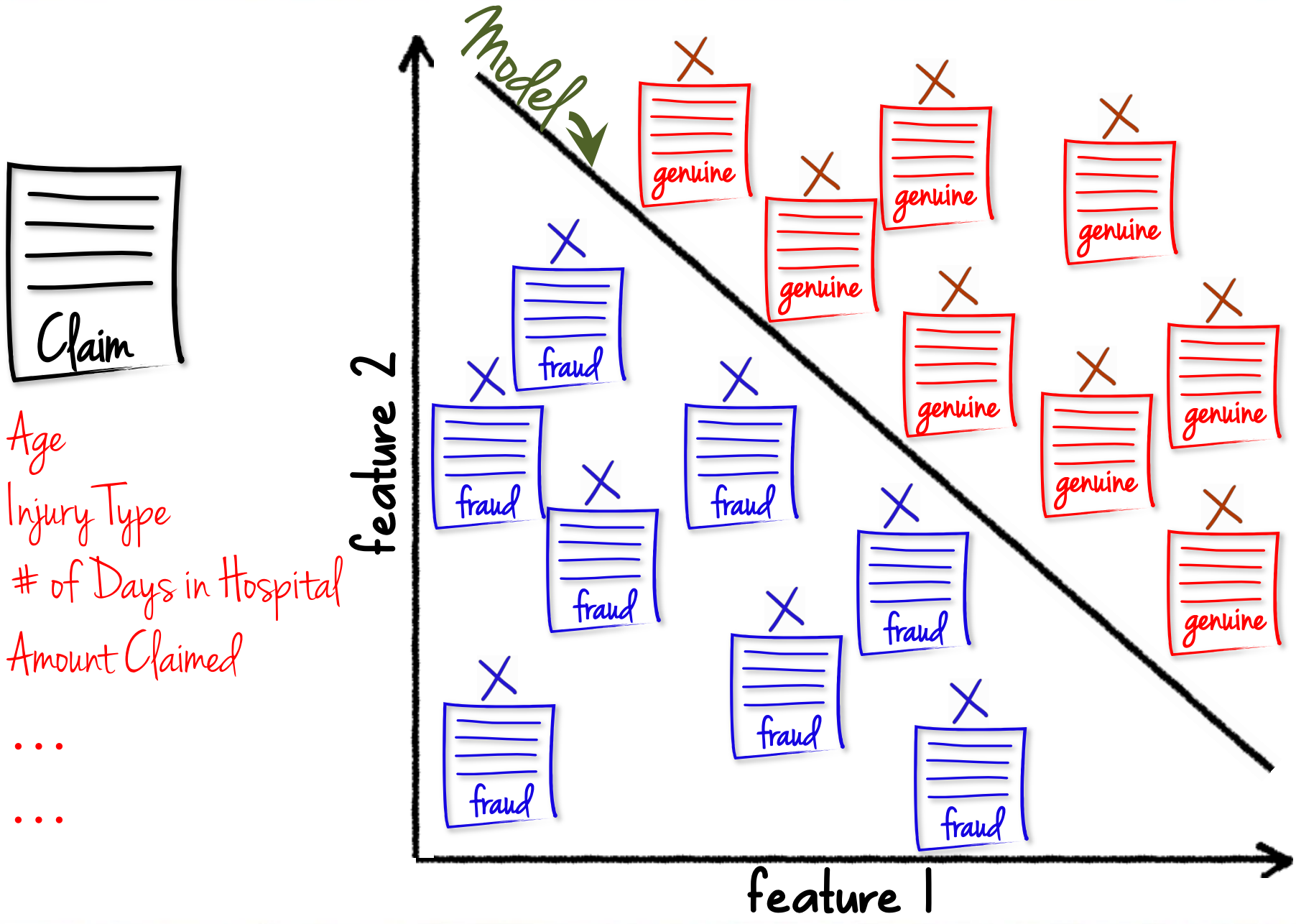
# of Days in Hospital

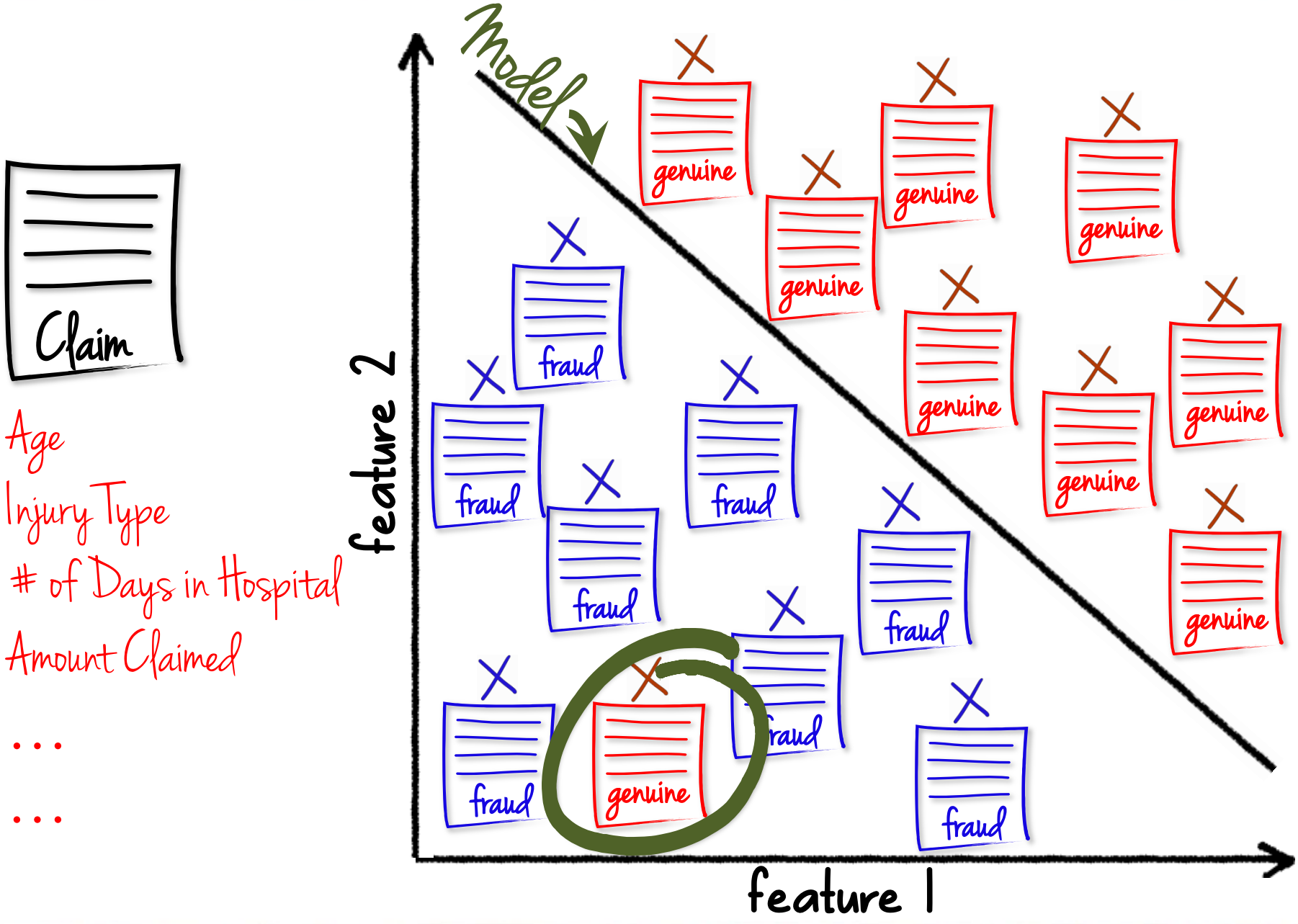
Amount Claimed

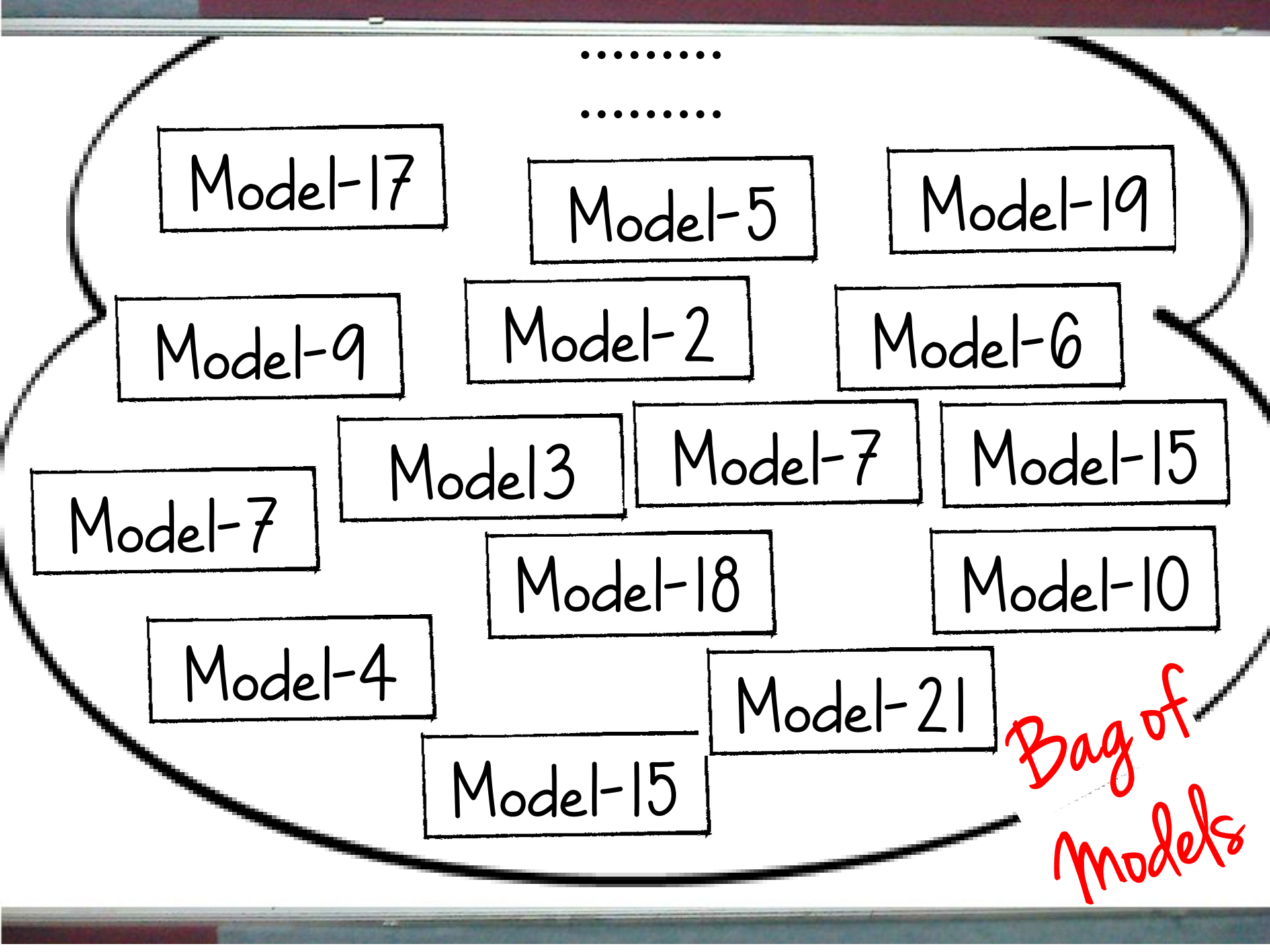
...

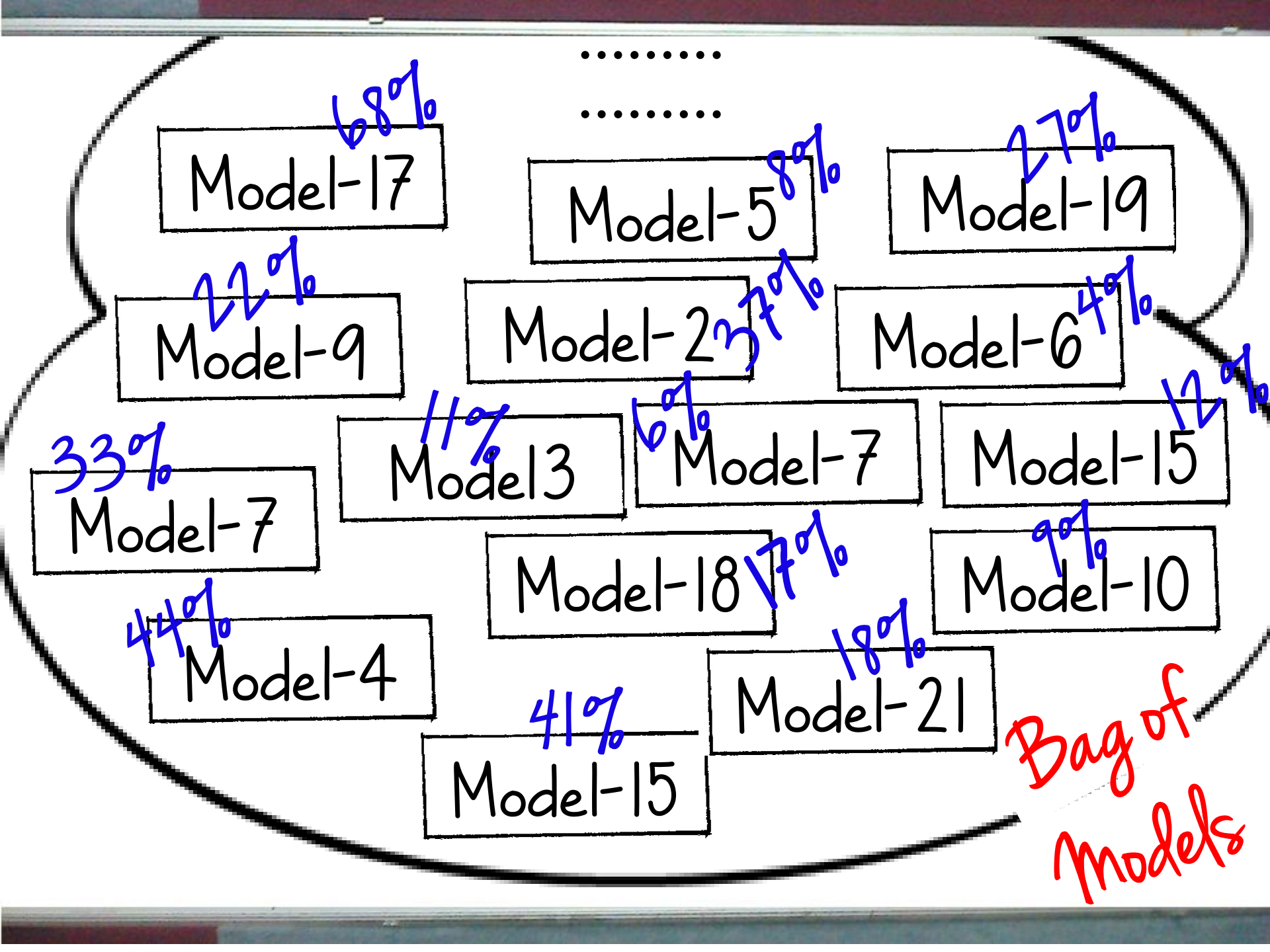
...

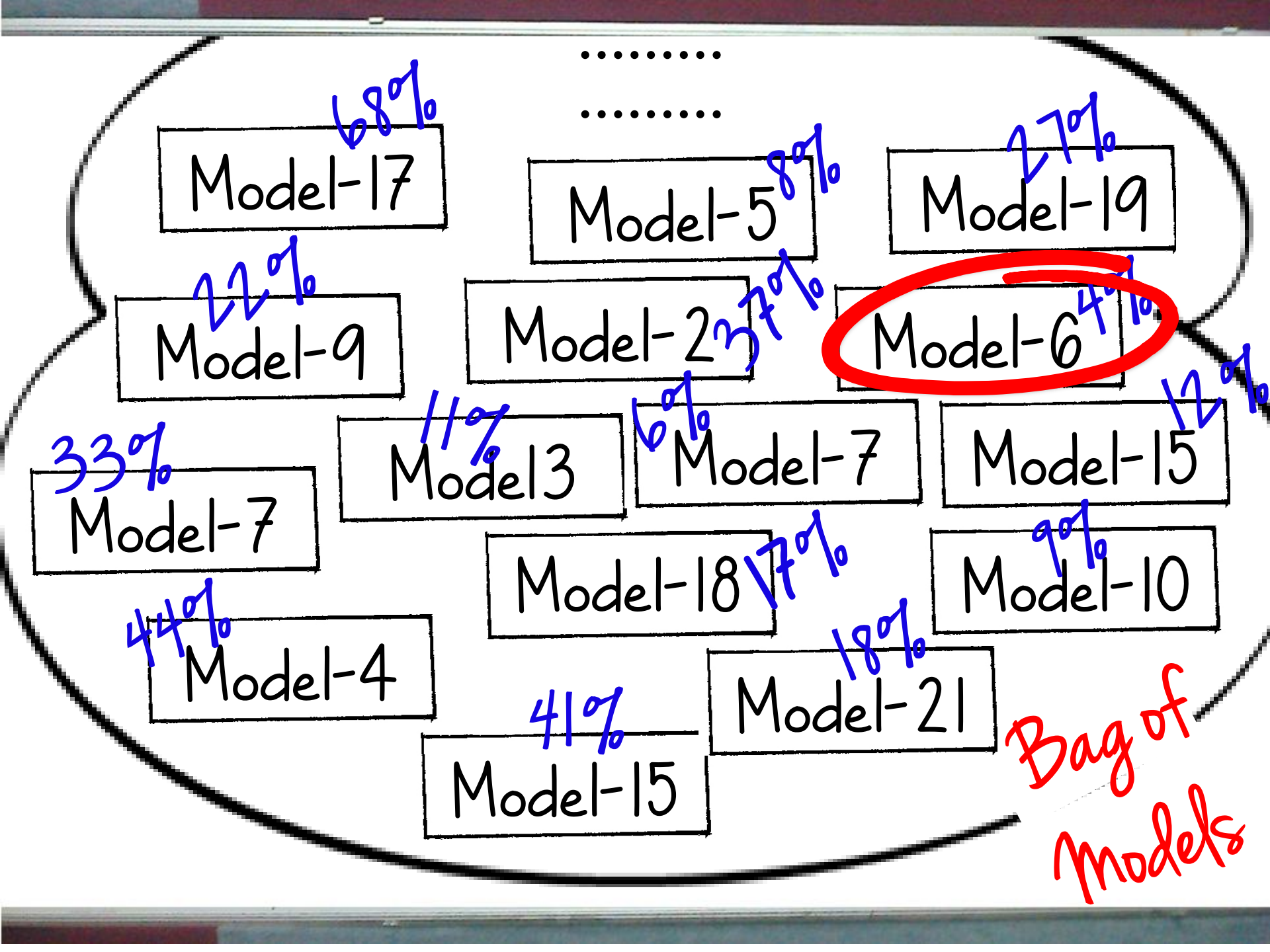

Claim
Age
Injury Type
# of Days in Hospital
Amount Claimed
...
...
feature 2
feature 1
genuine
fraud


Claim
Age
Injury Type
# of Days in Hospital
Amount Claimed
...
...
Model
feature 2
feature 1
genuine
fraud


Claim
Age
Injury Type
# of Days in Hospital
Amount Claimed
...
...
Model
feature 2
feature 1
genuine
fraud


.....
.....
Model-17
Model-5
Model-19
Model-9
Model-2
Model-6
Model13
Model-7
Model-15
Model-7
Model-18
Model-10
Model-4
Model-21
Model-15
Bag of Models


Model-17
68%
Model-5
8%
Model-19
27%
Model-9
22%
Model-23
37%
Model-6
4%
Model-7
33%
Model13
11%
Model-7
6%
Model-15
12%
Model-18
17%
Model-10
9%
Model-4
44%
Model-21
18%
Model-15
41%
Bag of Models


Model-17
68%
Model-5
8%
Model-19
27%
Model-9
22%
Model-2
37%
Model-6
4%
Model-7
33%
Model13
11%
Model-7
6%
Model-15
12%
Model-18
17%
Model-10
9%
Model-4
44%
Model-21
18%
Model-15
41%
Bag of Models

Optimization Problem

find model with 'minimum training error'

Optimization Problem

find model with 'minimum training error'

Intractable / Hard

Optimization Problem

find model with 'minimum training error'

Intractable / Hard

Proxy Optimization Problem

zation

blem

del with

raining error'

Logistic Regression

Support Vector Machine

AdaBoost

Neural Network

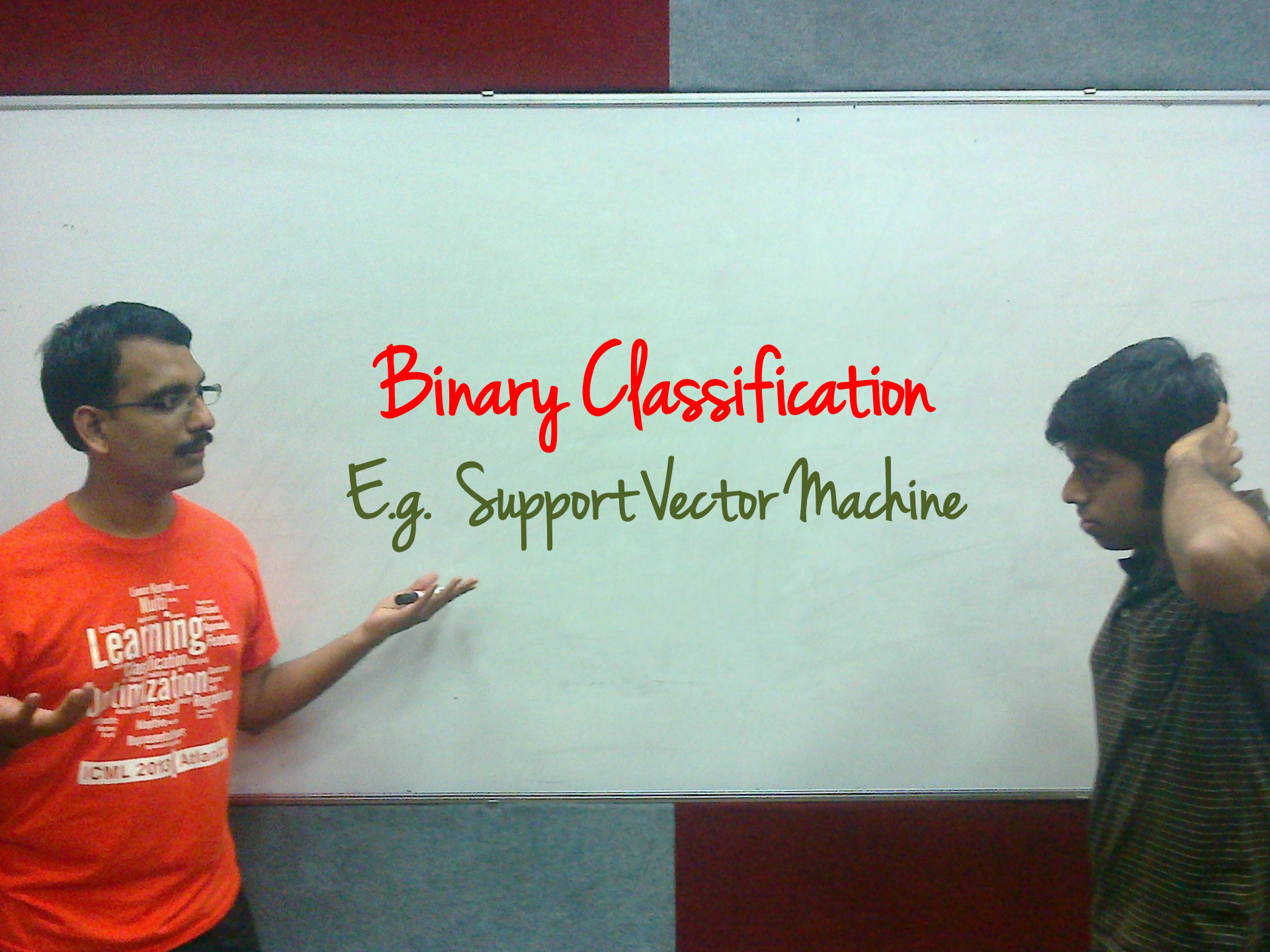
Decision Trees

...

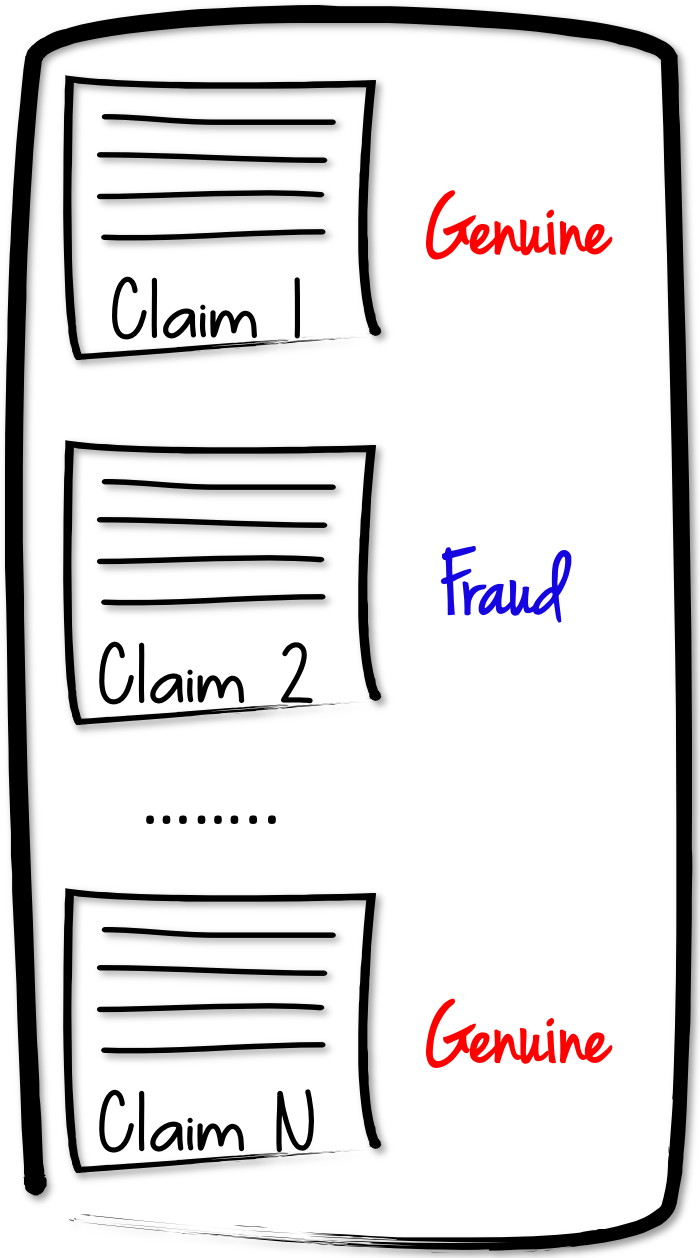
Proxy

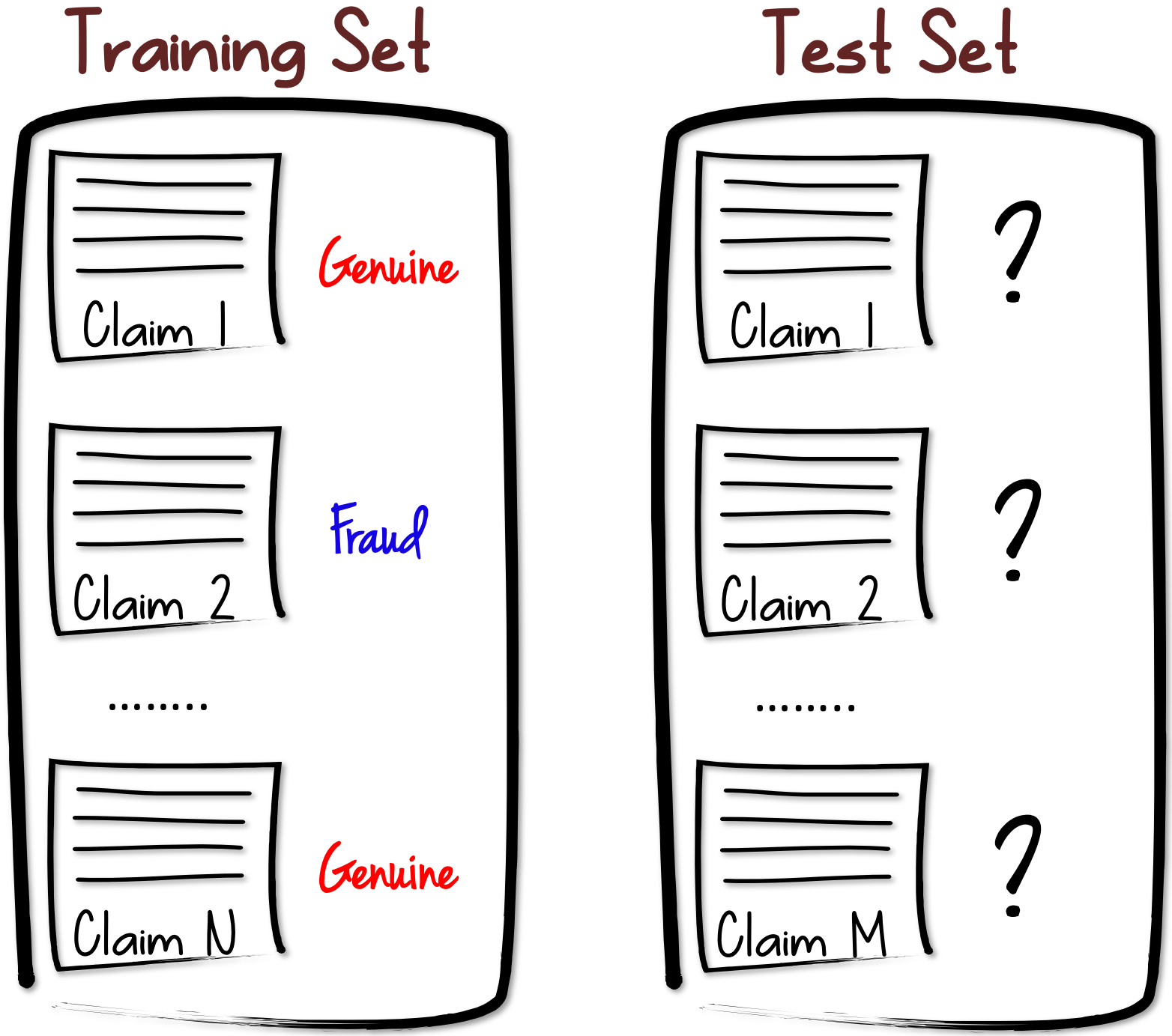
Optimiza

Proble


Binary Classification
E.g. Support Vector Machine

Training Set




Training Set
Claim 1
Genuine
Claim 2
Fraud
.....
Claim N
Genuine
Test Set
Claim 1
?
Claim 2
?
.....
Claim M
?

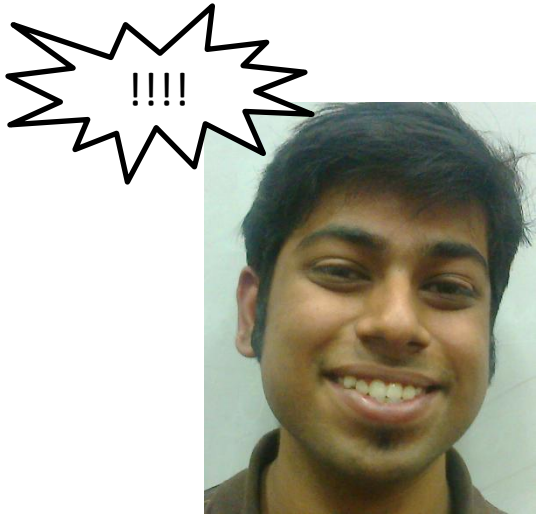
[Age | Injury Type | # of Days in Hospital | ... | ... | Amount Claimed]

Support Vector Machine

[7 Age | 23 Injury Type | 9 # of Days in Hospital | ... | ... | 24 Amount Claimed]

Support Vector Machine

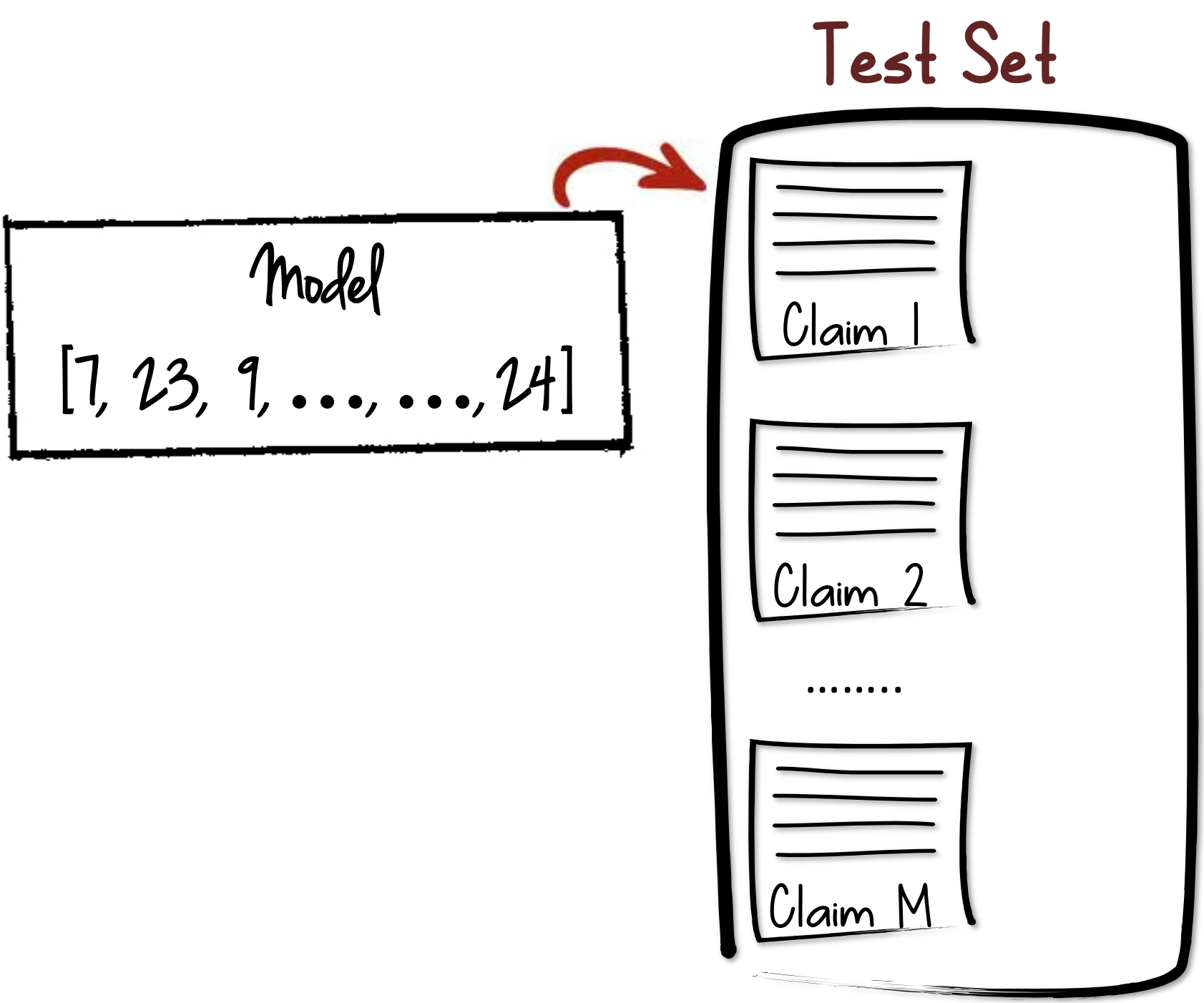
[7 Age | 23 Injury Type | 9 # of Days in Hospital | ... | ... | 24 Amount Claimed]



Model

[7, 23, 9, ..., ..., 24]

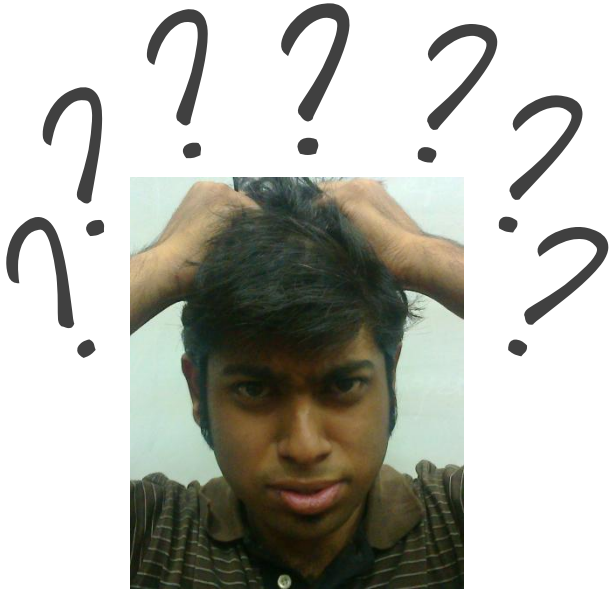
2% training error


Test Set
Model
[7, 23, 9, ..., ..., 24]
Claim 1
Claim 2
.....
Claim M

Test Set

Model

[7, 23, 9, ..., ..., 24]



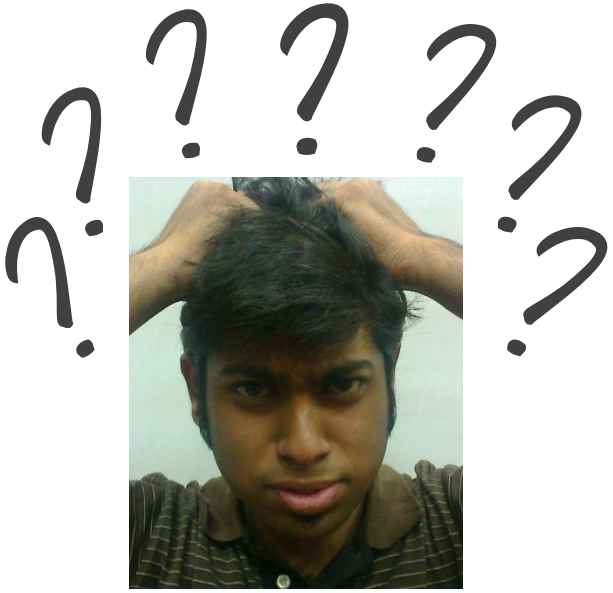
Claim 1 — Genuine

Claim 2 — Genuine

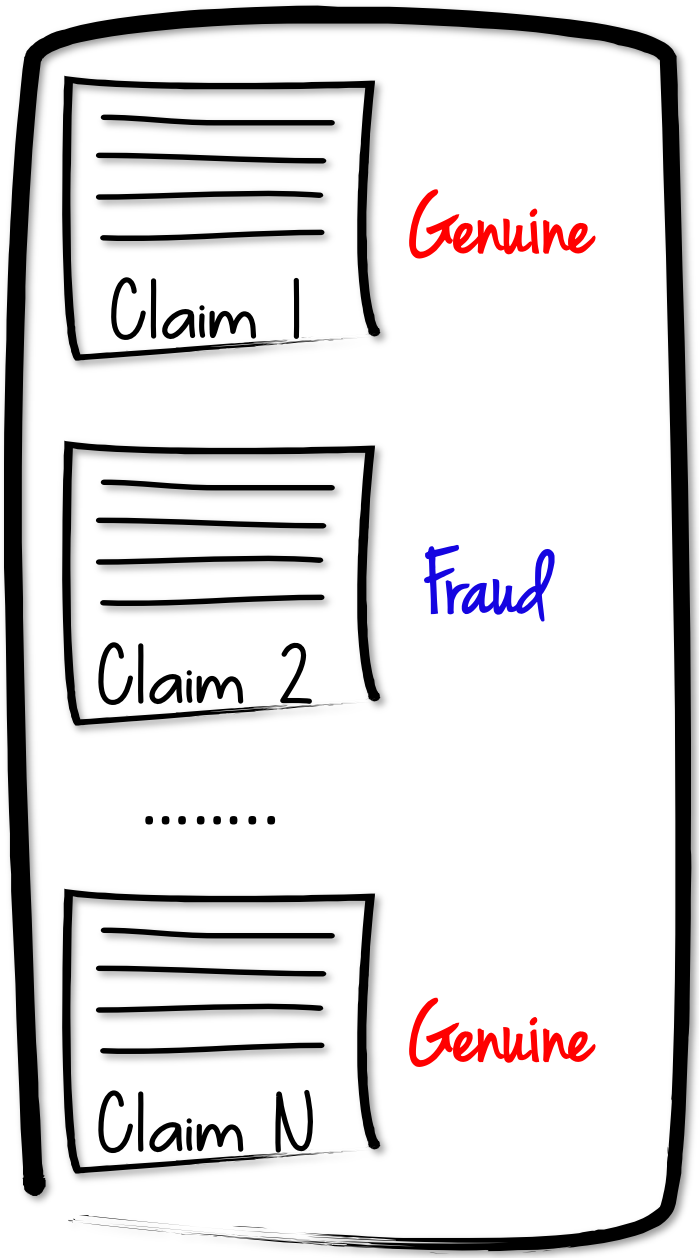
.....

Claim M — Genuine

Only 2% of the training examples are 'fraud'

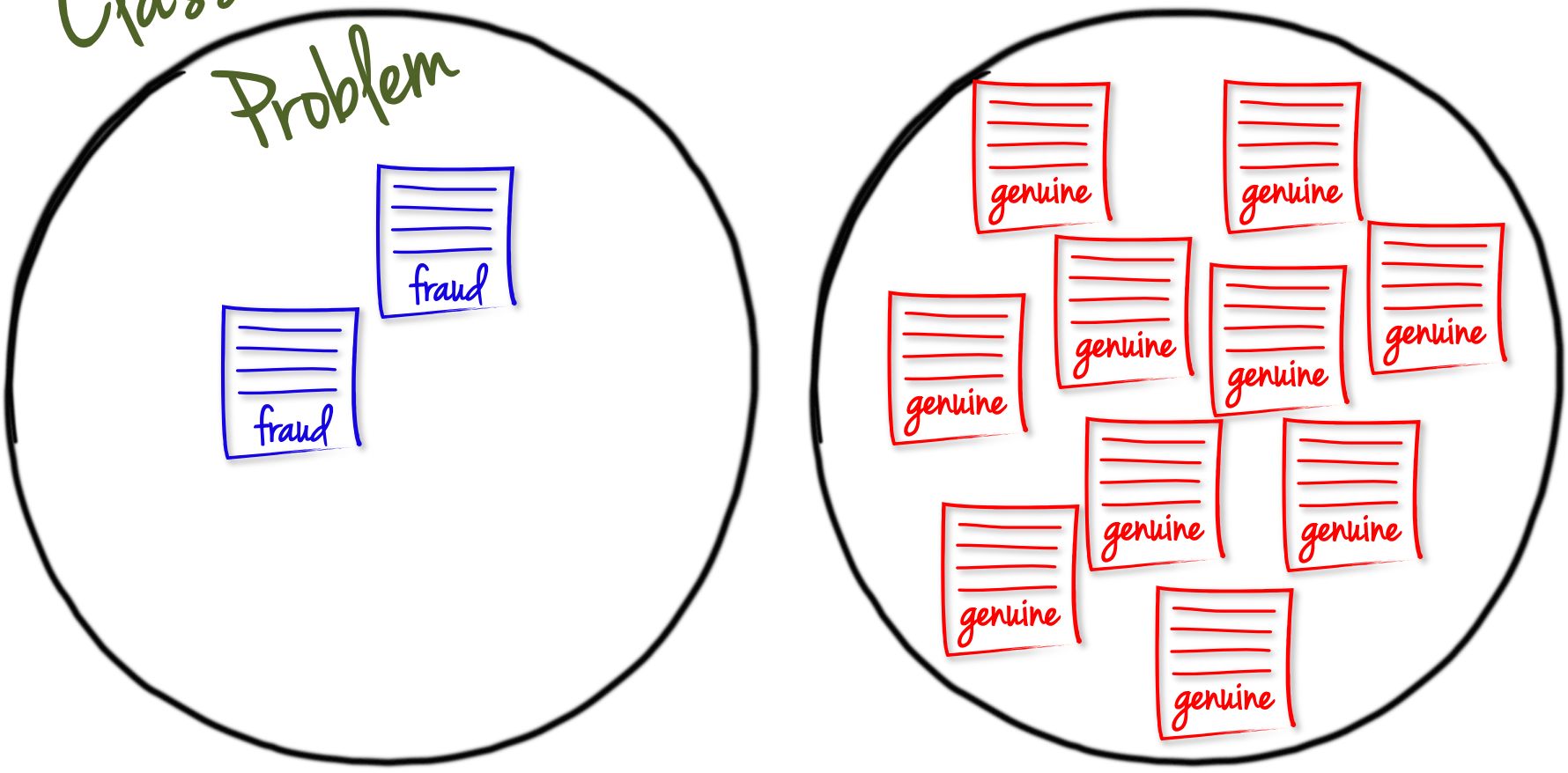


Training Set




Performance
Measure?
Linear Kernel
Multi
earning
Classification
timization
based
Regression
Features
Approach
Adaptive
Representations
IL 2013
Atlanta

Class Imbalance Problem



Correctly predicted positive examples

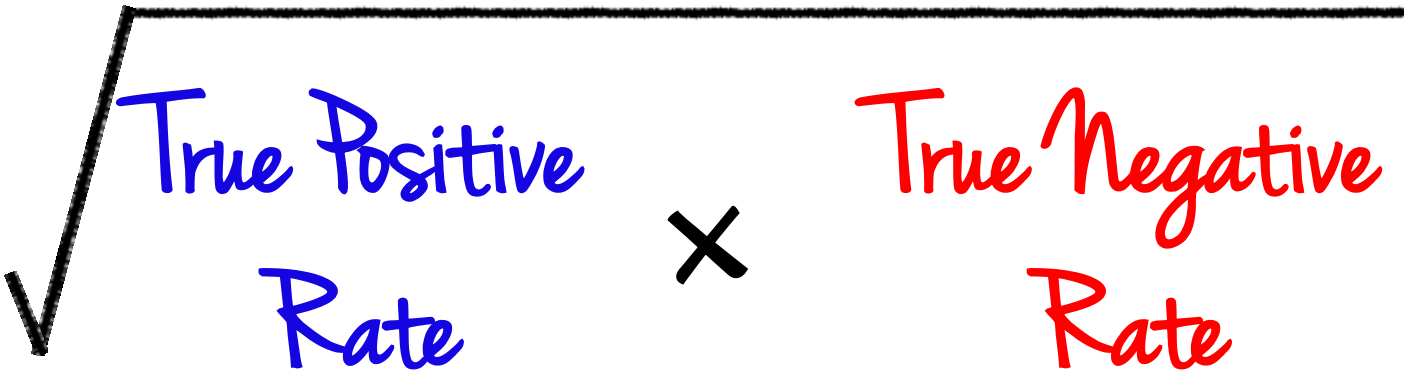
Correctly predicted positive examples

Correctly predicted negative examples

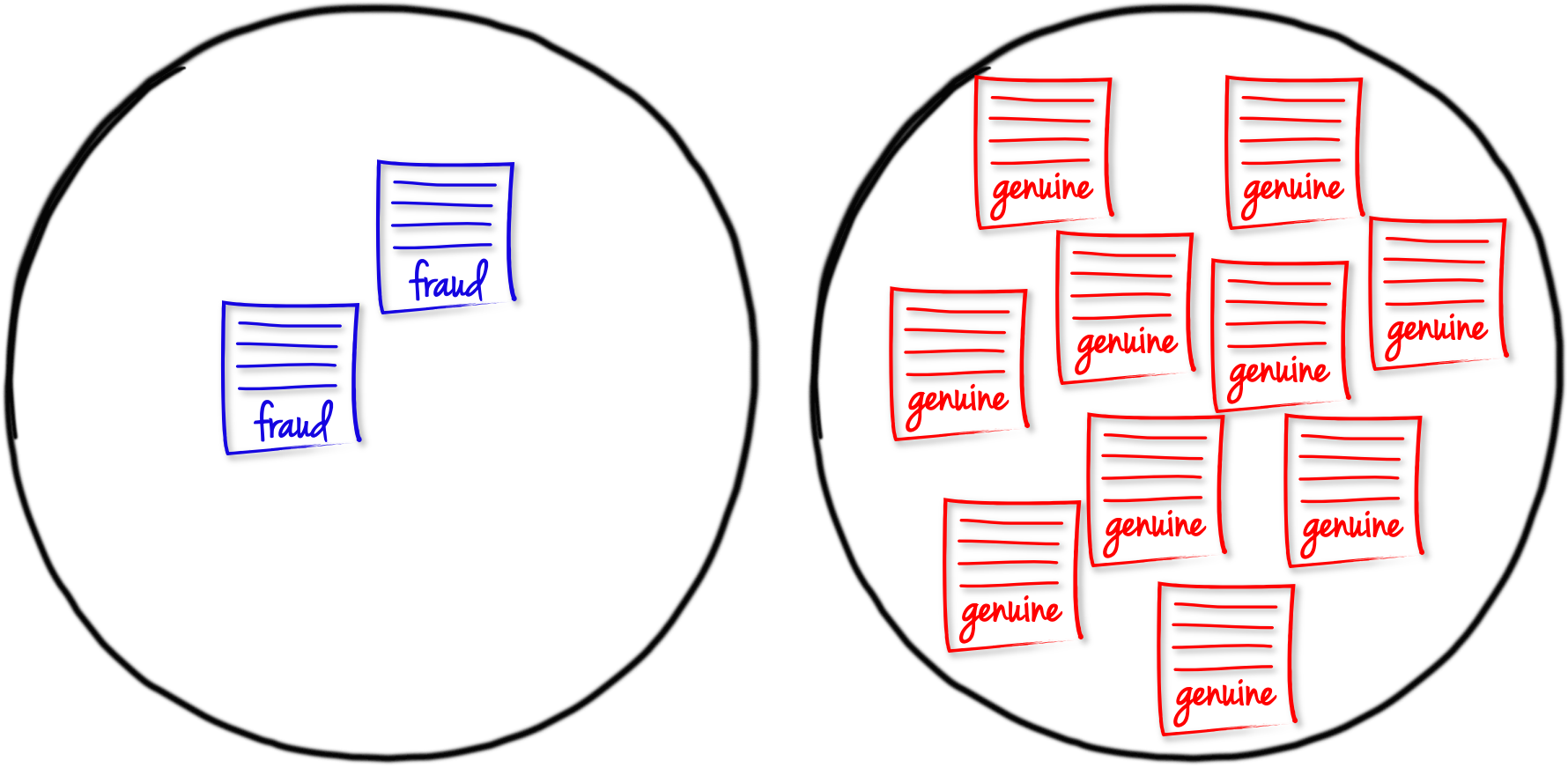
True Positive Rate

True Negative Rate

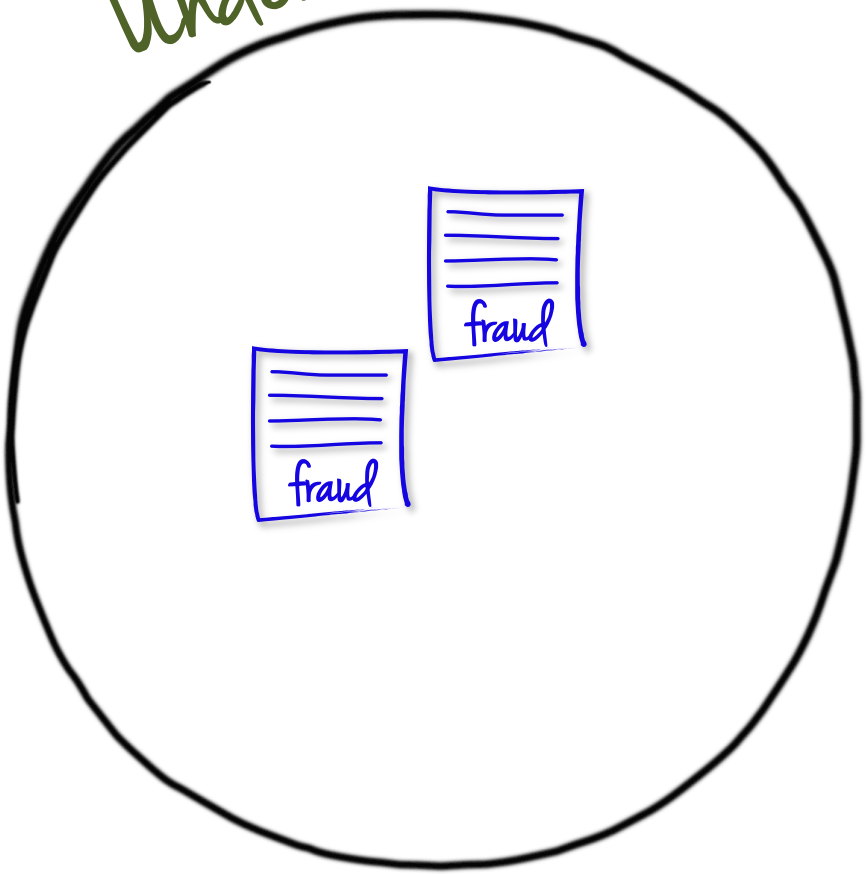
$$\frac{\text{True Positive Rate} + \text{True Negative Rate}}{2}$$

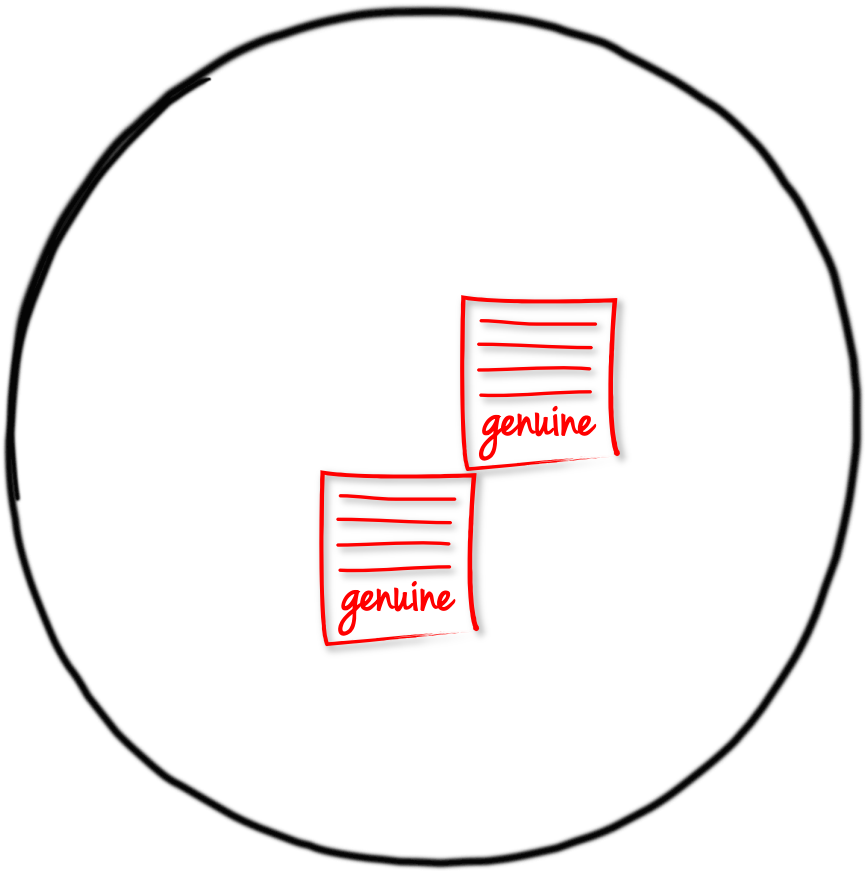

True Positive Rate
×
True Negative Rate

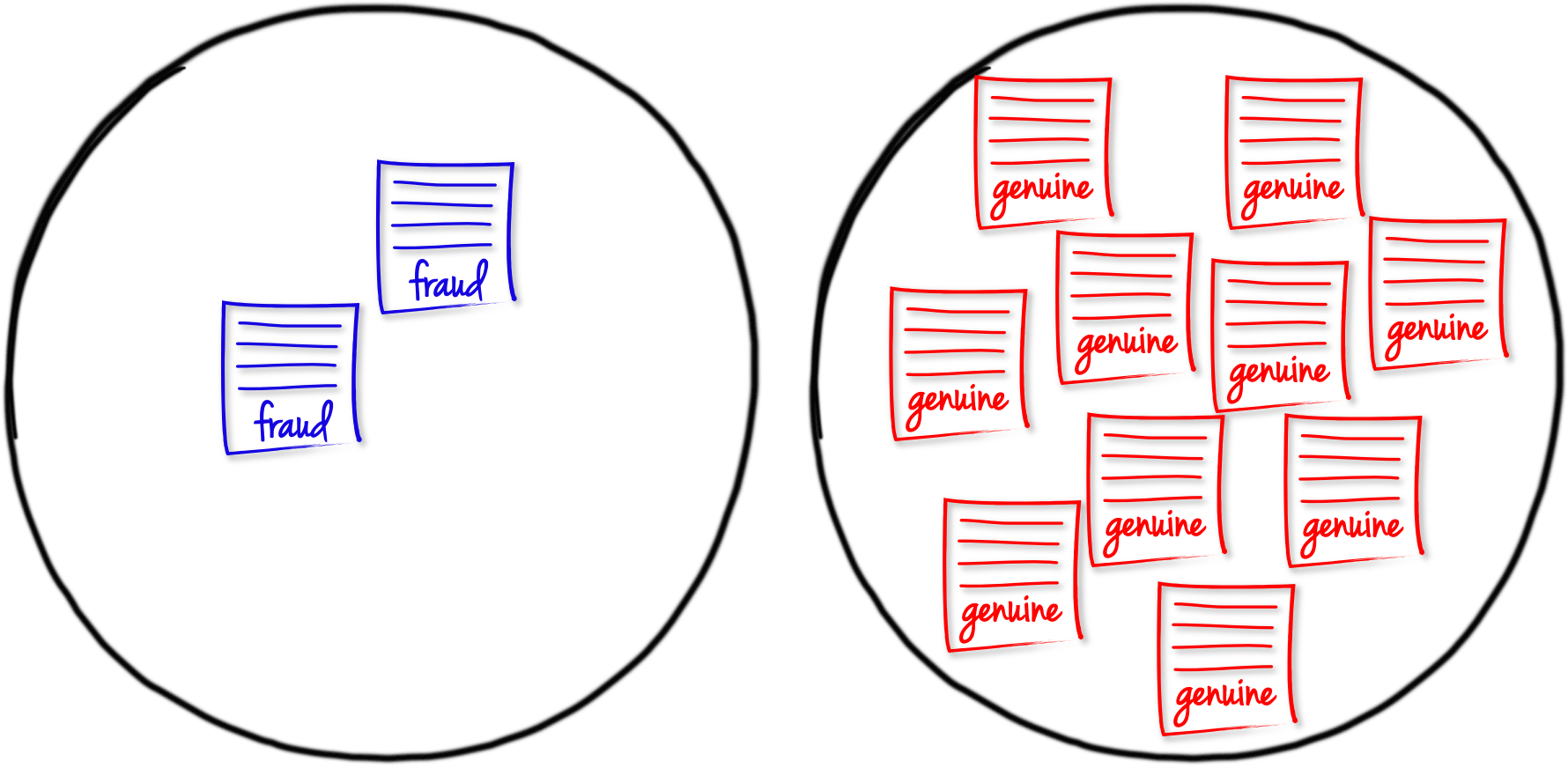

Algorithm?
Linear Kernel
Multi
earning
Efficient
Approach
Features
Classification
timization
based
Regression
Adaptive
Representations
L 2013
Atlanta


fraud
fraud
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine

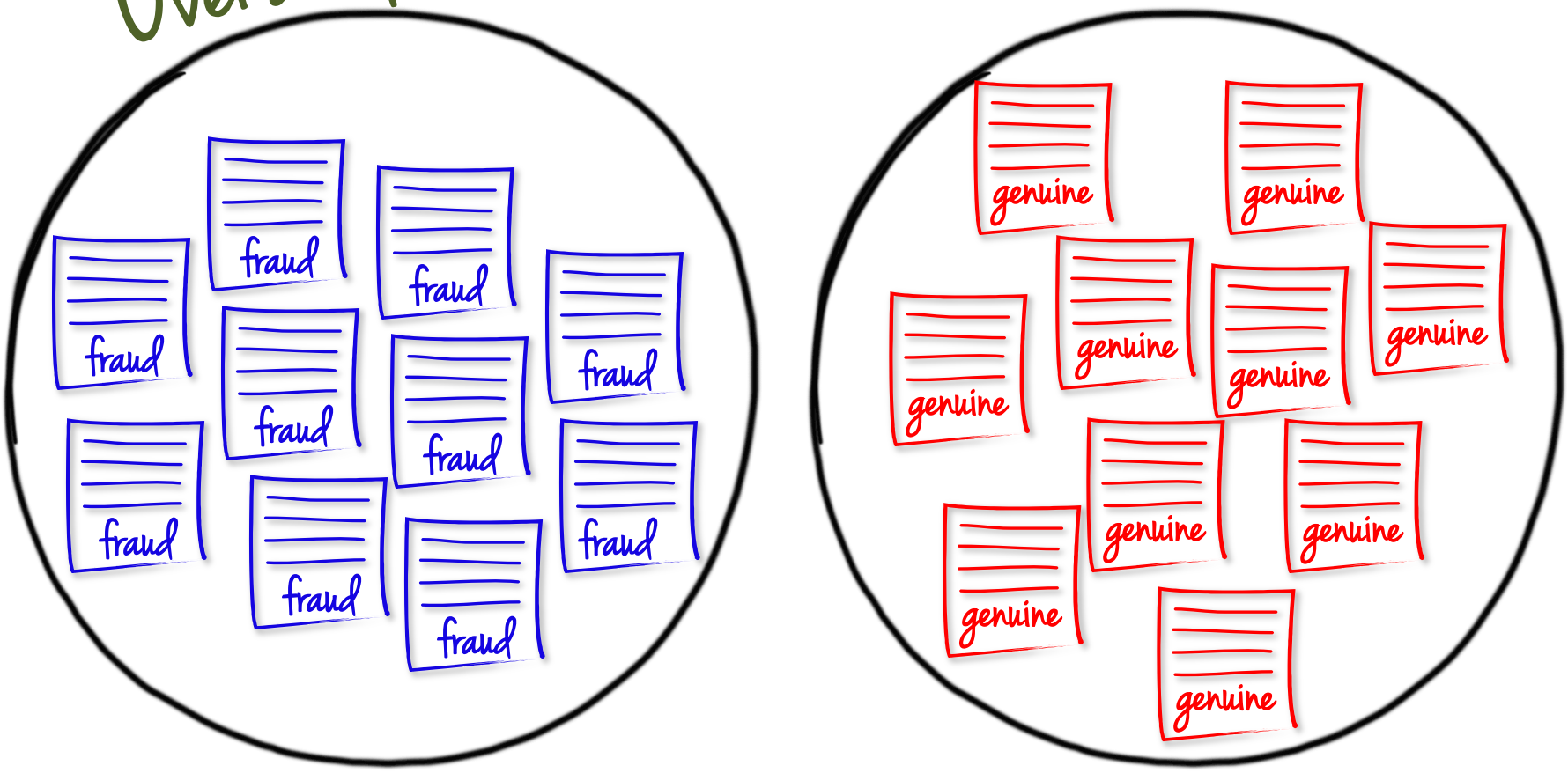
Undersampling

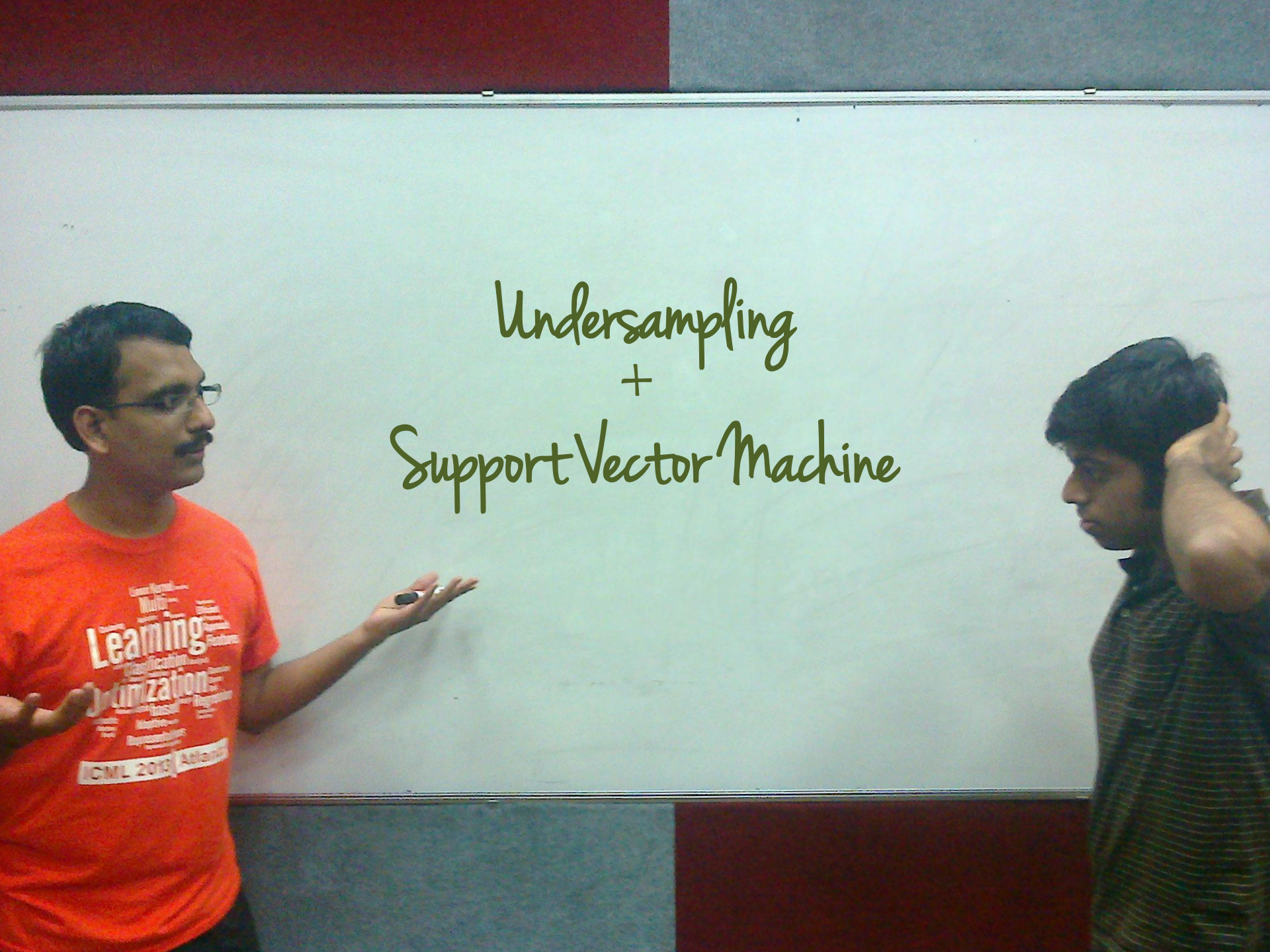


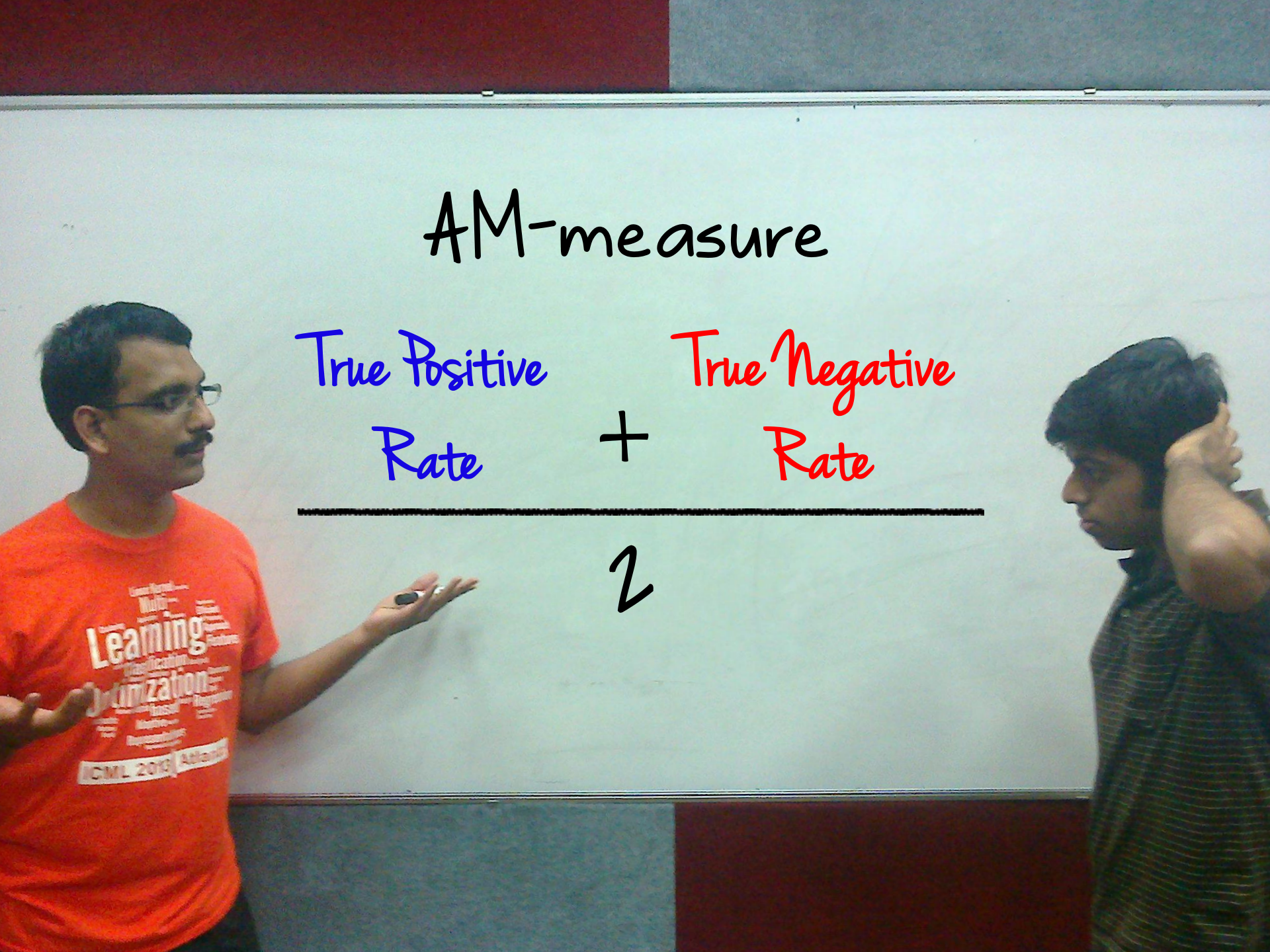



fraud
fraud
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine

Oversampling




Undersampling
+
Support Vector Machine
Learning
ICML 2013


AM-measure
True Positive Rate
+
True Negative Rate
2

[Age | Injury Type | # of Days in Hospital | ... | ... | Amount Claimed]

Undersampling + Support Vector Machine

[19 Age | 12 Injury Type | 21 # of Days in Hospital | ... | ... | 17 Amount Claimed]

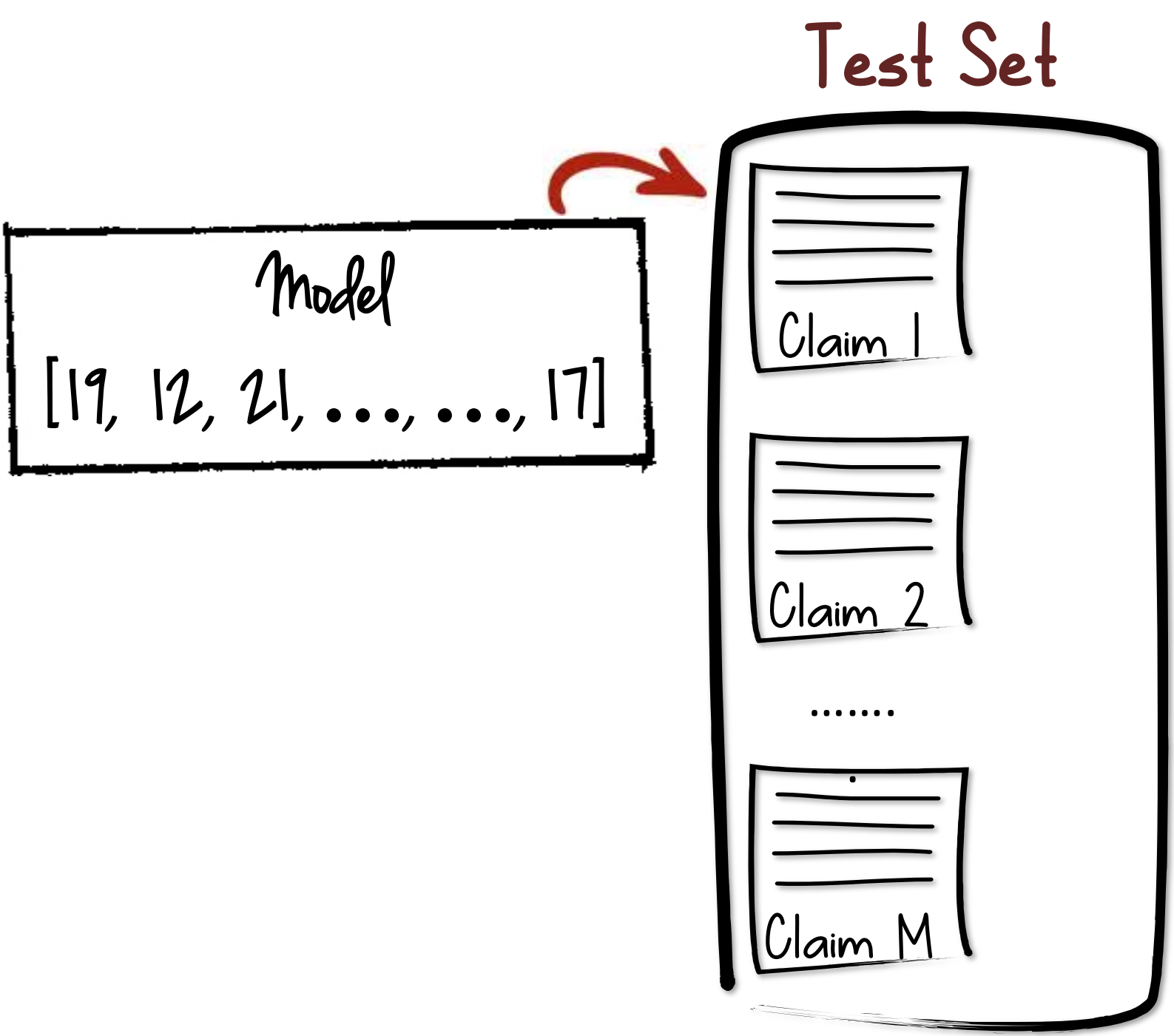
Undersampling
+
Support Vector Machine

[19 Age | 12 Injury Type | 21 # of Days in Hospital | ... | ... | 17 Amount Claimed]

Model

[19, 12, 21, ..., ..., 17]

94% AM-measure on training set


Test Set
Model
[19, 12, 21, ..., ..., 17]
Claim 1
Claim 2
.....
Claim M

Test Set

Model

[19, 12, 21, ..., ..., 17]

Claim 1 — Fraud

Claim 2 — Genuine

.....

Claim M — Genuine

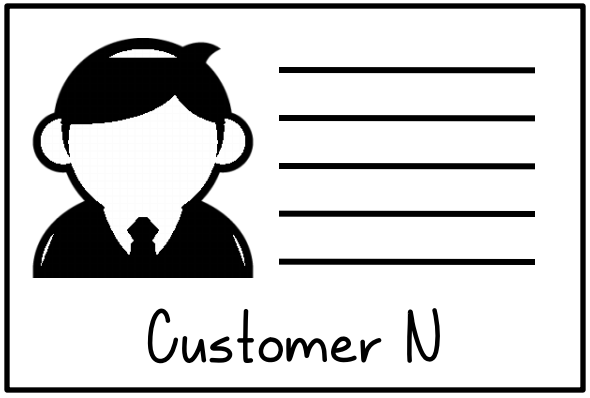
91% AM measure


Customer
Churn

Churn

Non-Churn

.....

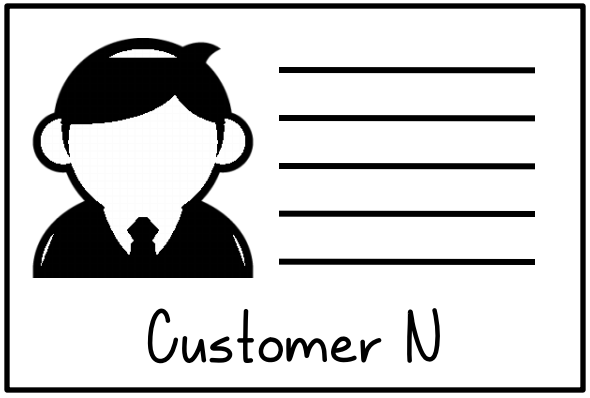


Churn

Churn

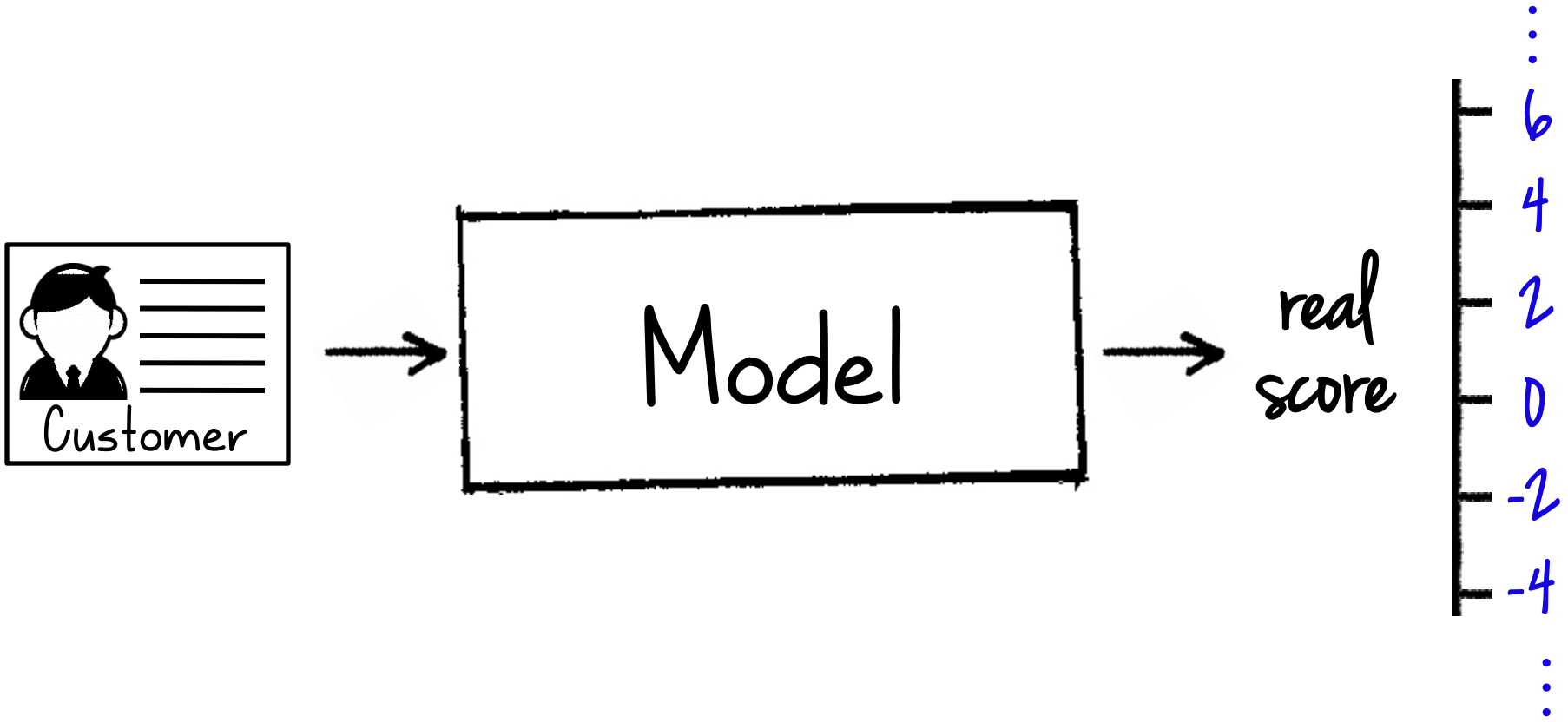
Non-Churn

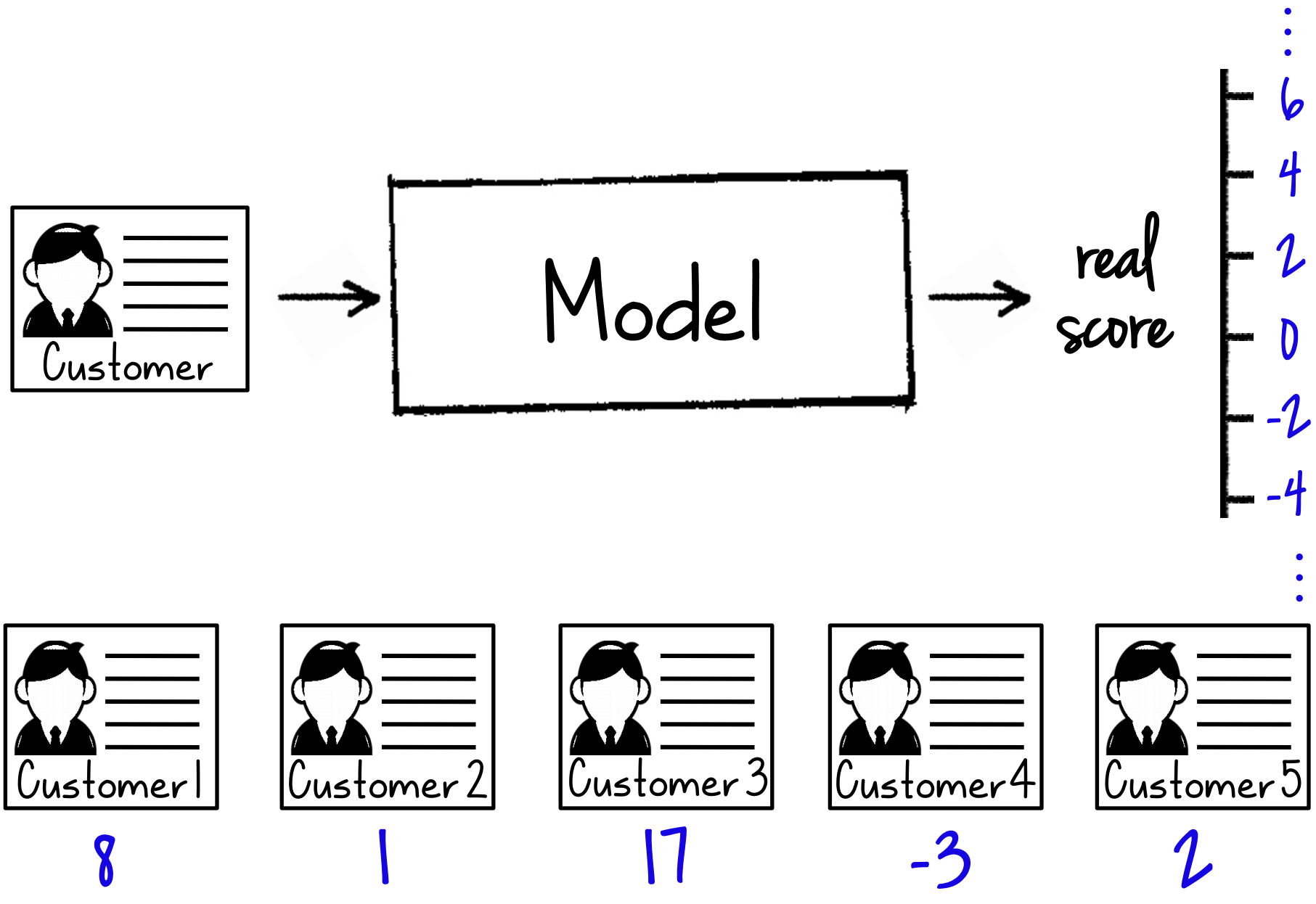
.....

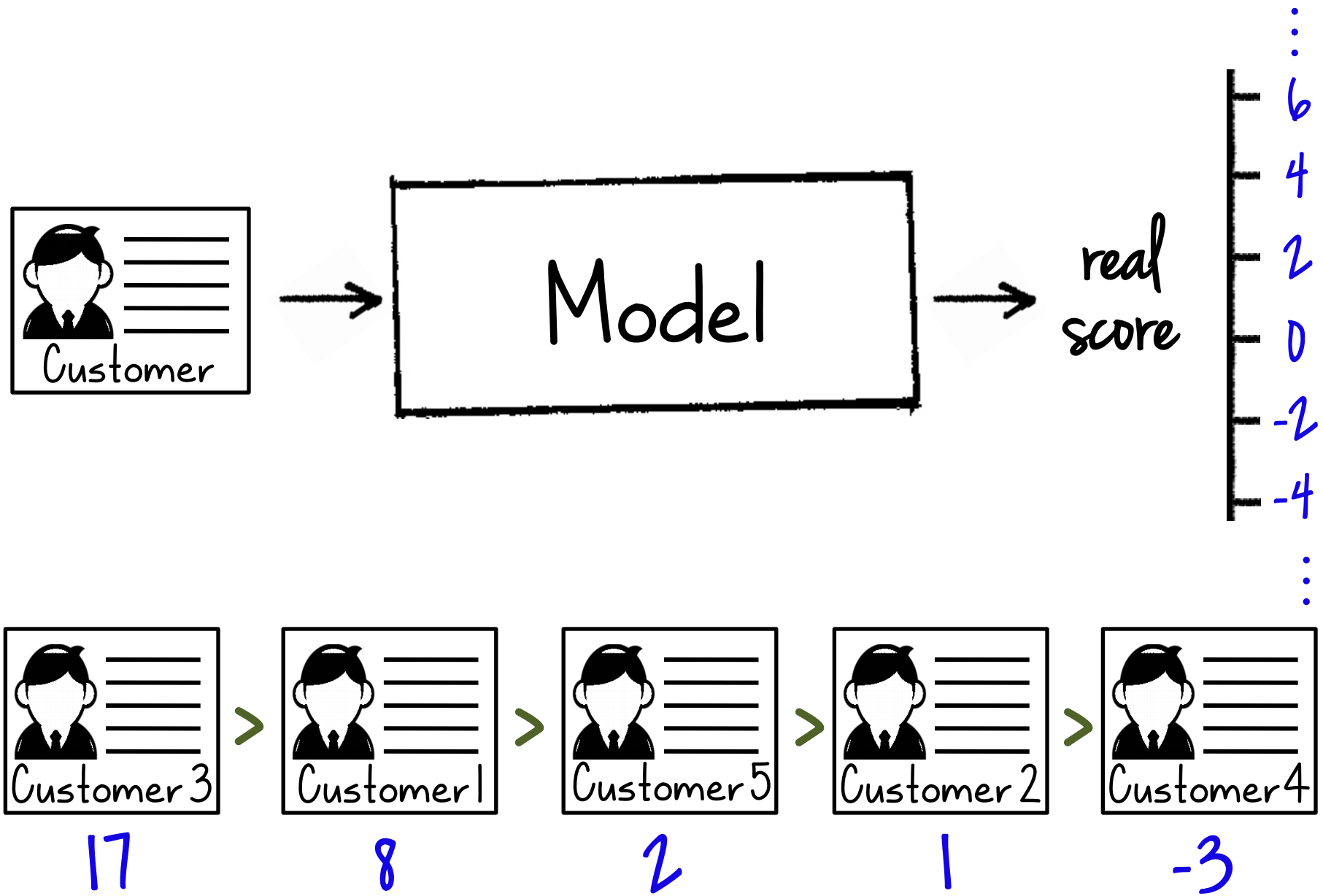


Churn

.....


Customer
Model
real score
6
4
2
0
-2
-4


Customer
Model
real score
6
4
2
0
-2
-4
Customer 1
Customer 2
Customer 3
Customer 4
Customer 5
8
1
17
-3
2


Customer
Model
real score
6
4
2
0
-2
-4
Customer 3
Customer 1
Customer 5
Customer 2
Customer 4
>
>
>
>
17
8
2
1
-3

Performance Measure?

Customer → Model → real score

6
4
2
0
-2
-4

Customer 3 > Customer 1 > Customer 5 > Customer 2 > Customer 4

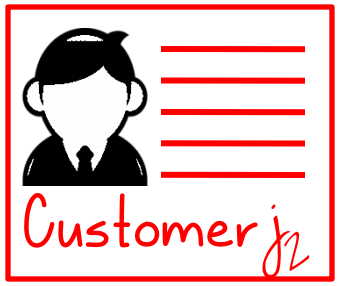
17 > 8 > 2 > 1 > -3


~~Binary Classification~~
Bipartite Ranking

>

?

>



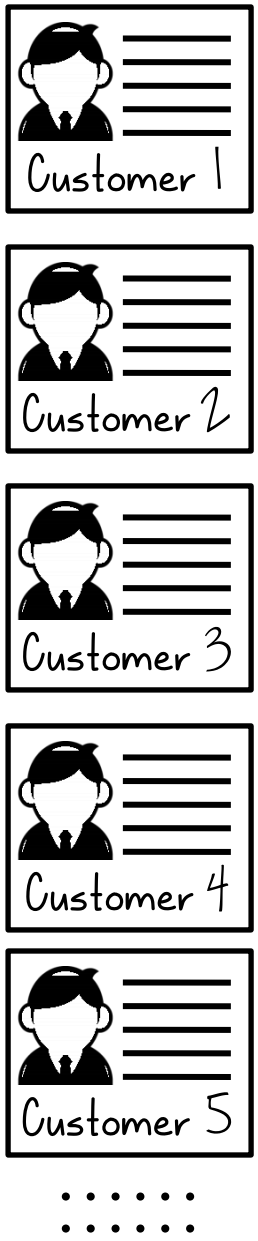
?

Pair-wise Accuracy

.....

>

?


Customer 1
Customer 2
Customer 3
Customer 4
Customer 5

Positives in top K positions

Precision@K

Positives in top K positions

Average Precision

RankSVM

RankNet

RankBoost

SVMMAP

SVMPerf

...


~~Binary Classification~~
Bipartite Ranking
E.g. RankSVM

[Location | Avg. #. of calls per day | Avg. duration of call | ... | ... | Mobile Tarrif]

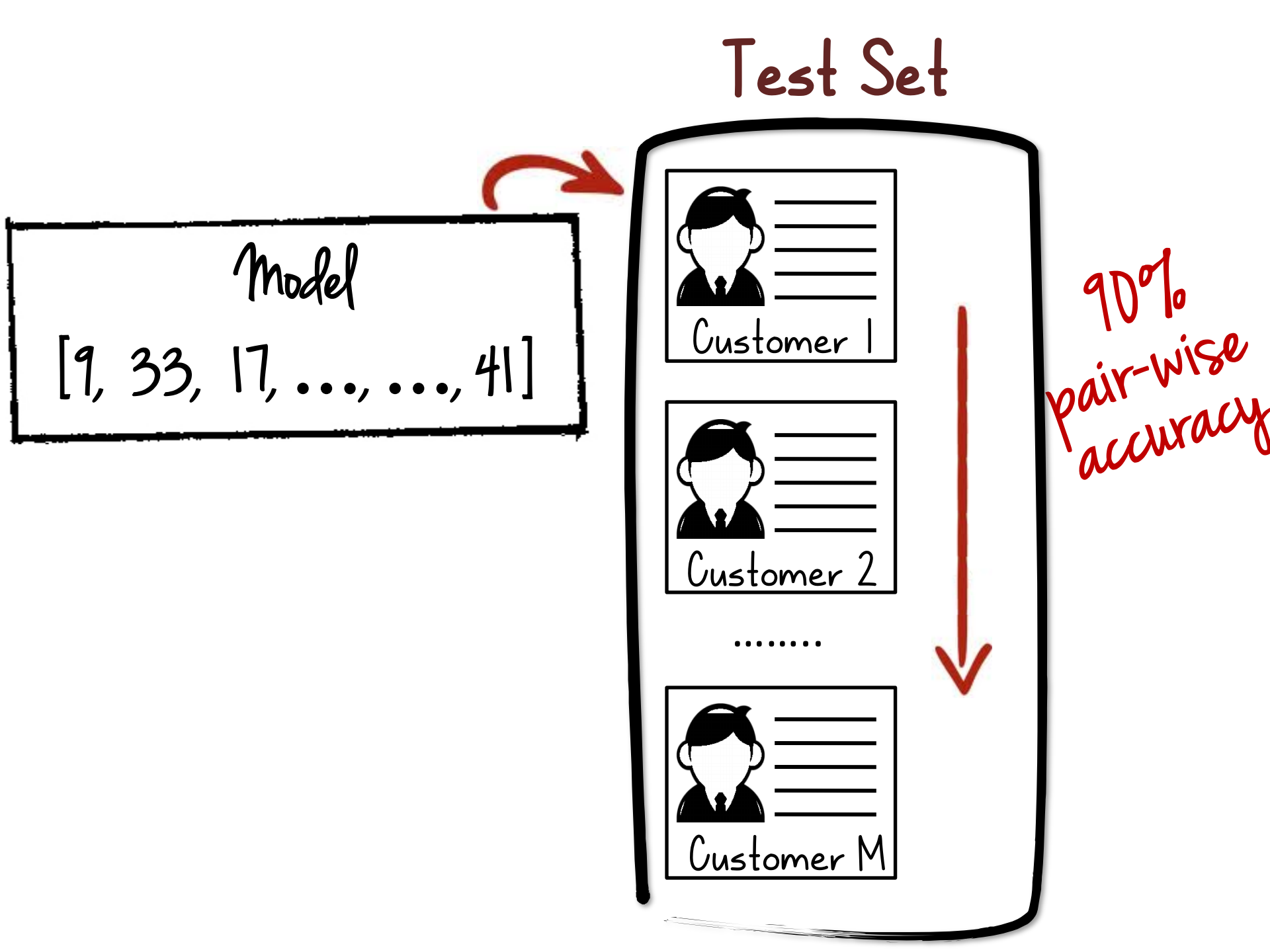
RankSVM

[9 Location | 33 Avg. #. of calls per day | 17 Avg. duration of call | ... | ... | 41 Mobile Tarrif]

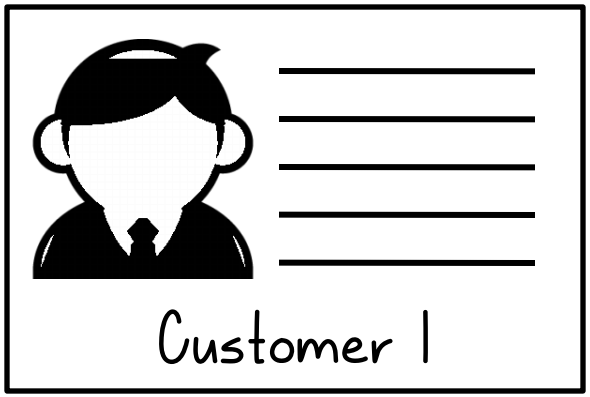
Model

[9, 33, 17, ..., ..., 41]

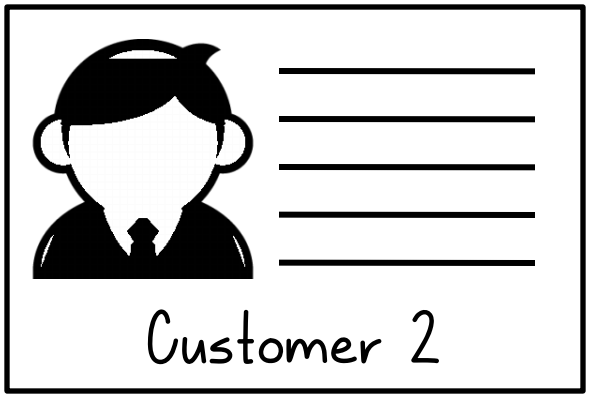
93% pair-wise accuracy on training set


Test Set
Model
[9, 33, 17, ..., ..., 41]
Customer 1
Customer 2
.....
Customer M
90%
pair-wise
accuracy


Customer
Churn



Churn



Non-Churn

.....

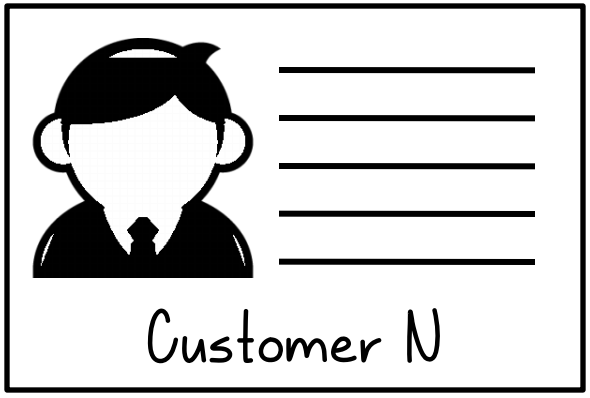
Churn

.....

Churn

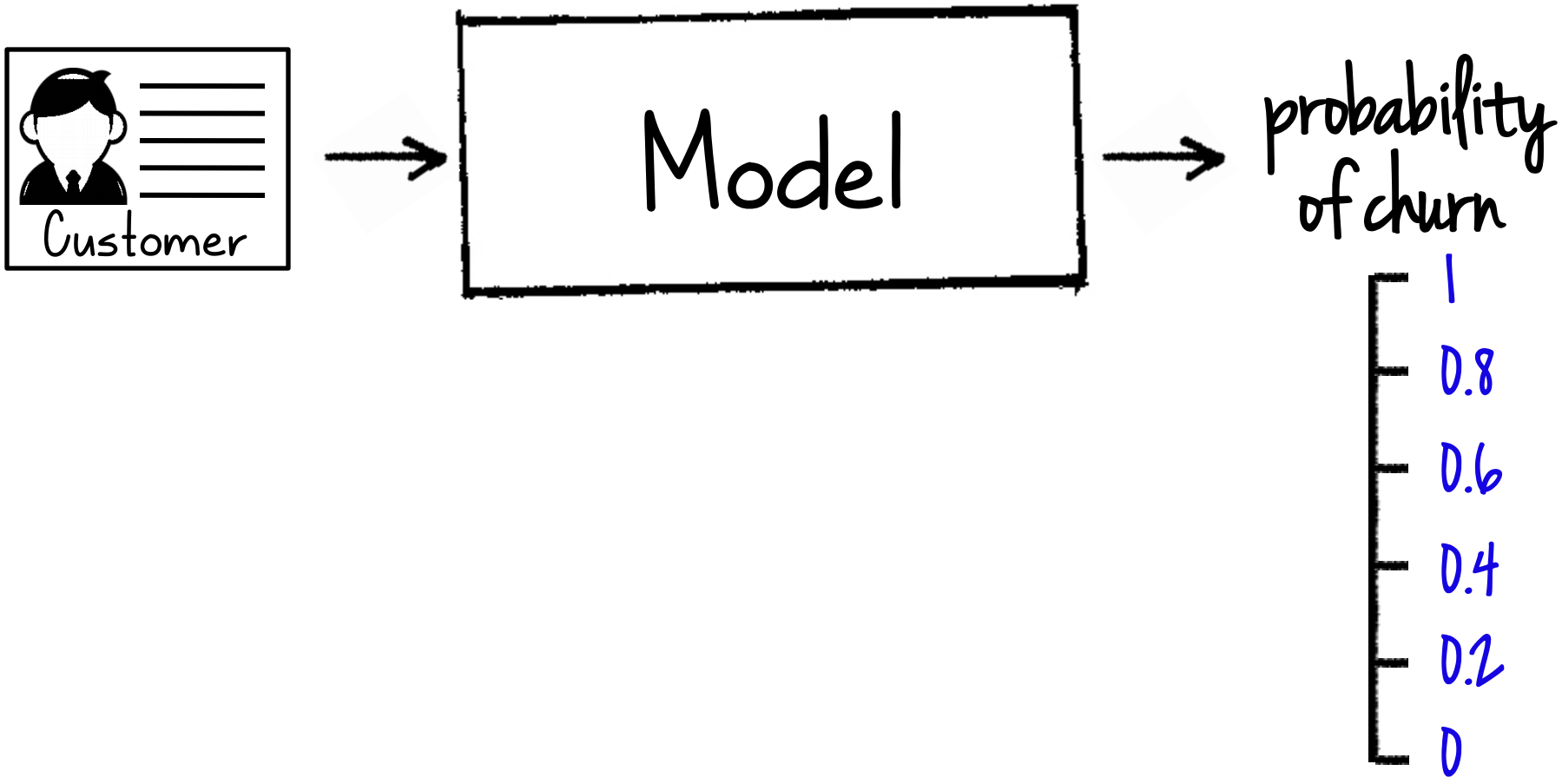
Non-Churn

.....



Churn

What is the probability of churn?


Customer
Model
probability of churn
1
0.8
0.6
0.4
0.2
0

Performance Measure?

Model

probability of churn

1

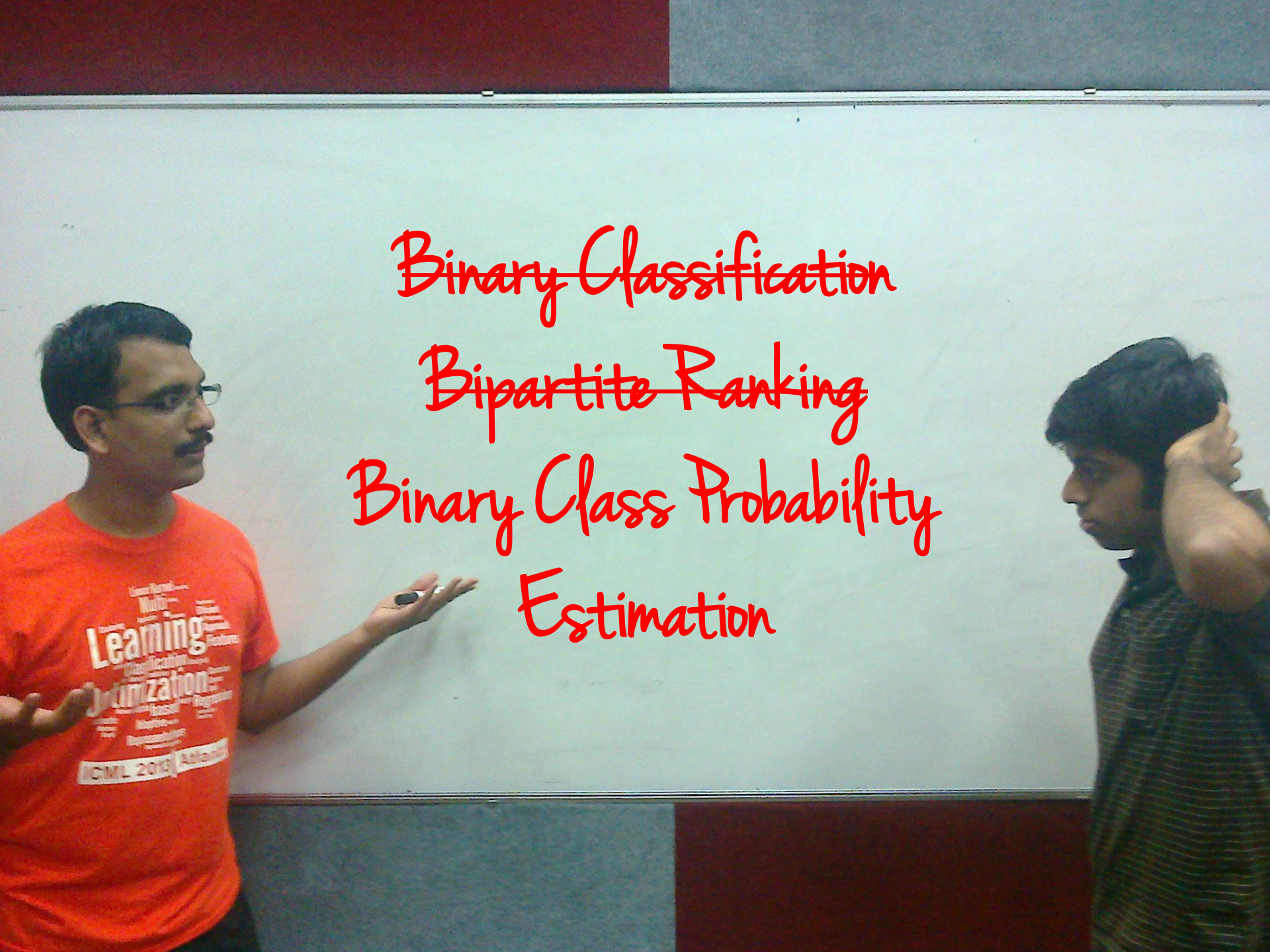
0.8

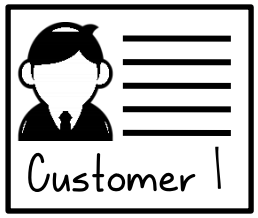
0.6

0.4

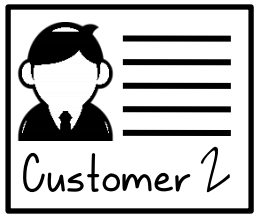
0.2

0

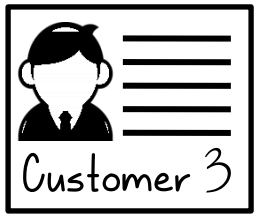

~~Binary Classification~~
~~Bipartite Ranking~~
Binary Class Probability
Estimation



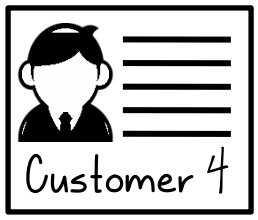
0.16



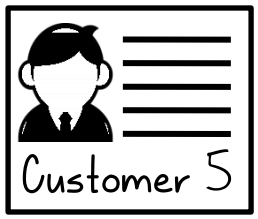
0.23



0.89



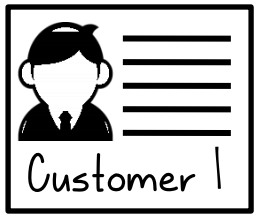
0.37



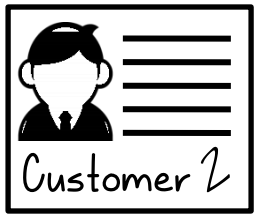
0.93

.....

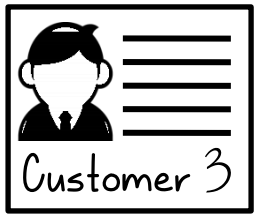
.....



0.16

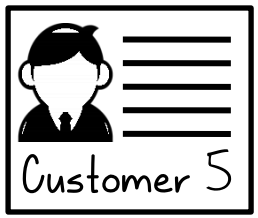


0.23



0.89

0.37



0.93

.....

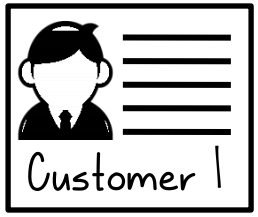
.....

Performance Measure?

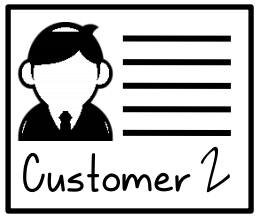
logloss

Squared Error

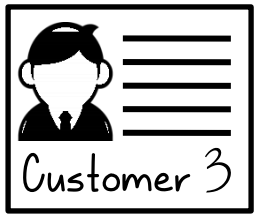
...



0.16

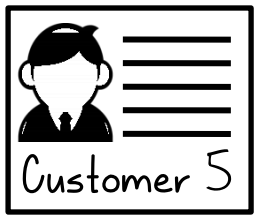


0.23



0.89

0.37



0.93

.....

.....

Performance Measure?

Logloss

Squared Error

...

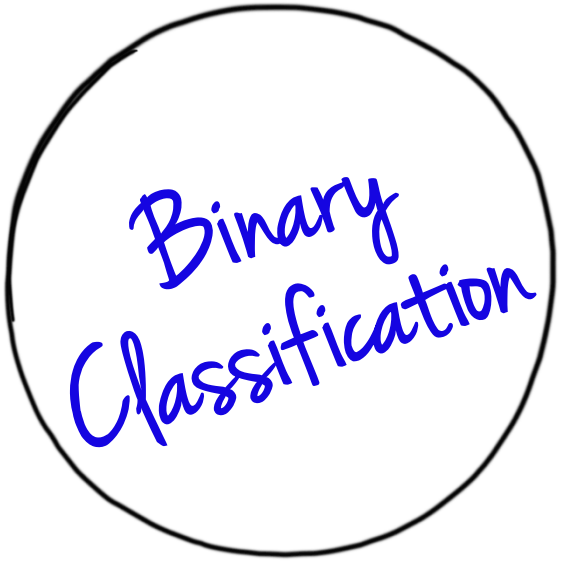
Algorithms?

Logistic Regression

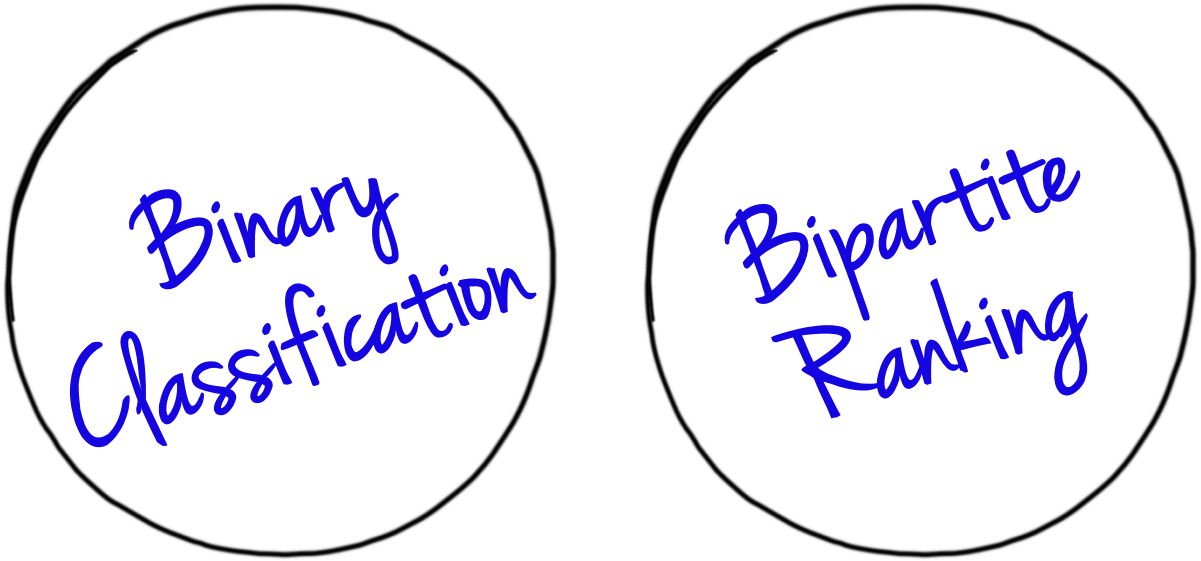
LogitBoost

...

the story so far ...



the story so far ...



the story so far ...

the story so far …

Standard Error

Precision@k

Average Precision

LogLoss

F-measure

G-mean

Pair-wise Accuracy

AM-measure

Logistic Regression

RankSVM

Support Vector Machine

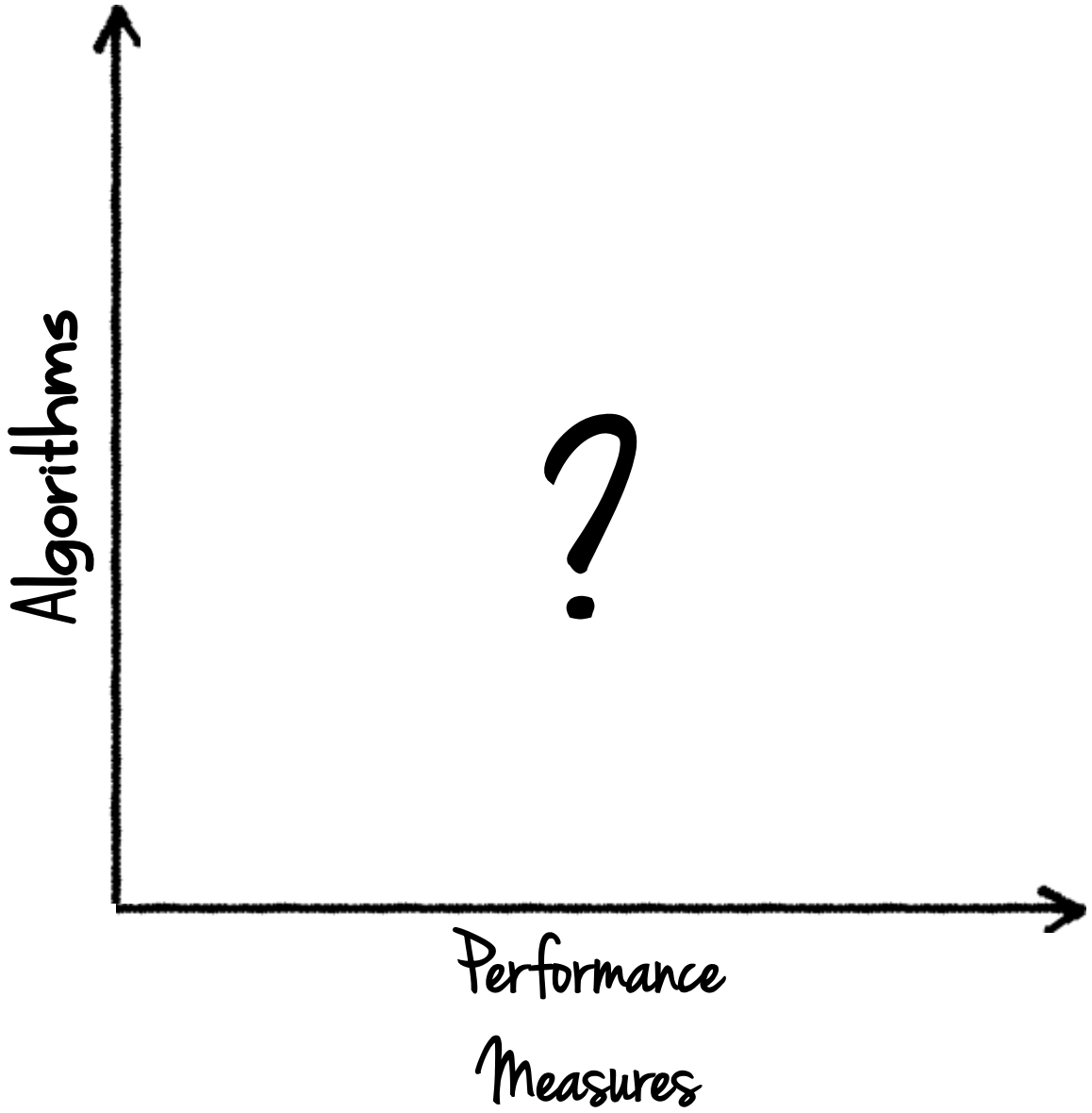
Neural Network

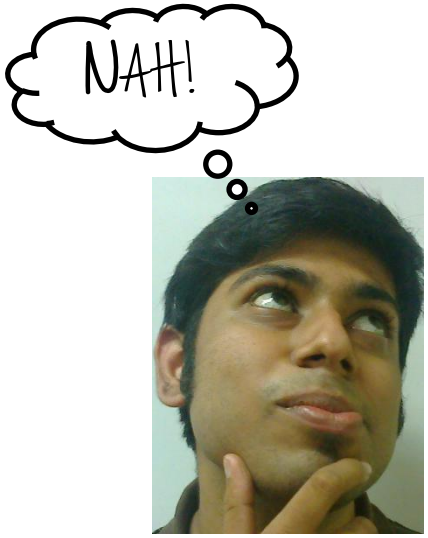
Decision Tree

RankBoost

AdaBoost

RankNet


Algorithms
?
Performance
Measures



Bipartite Ranking

Binary Classification

?

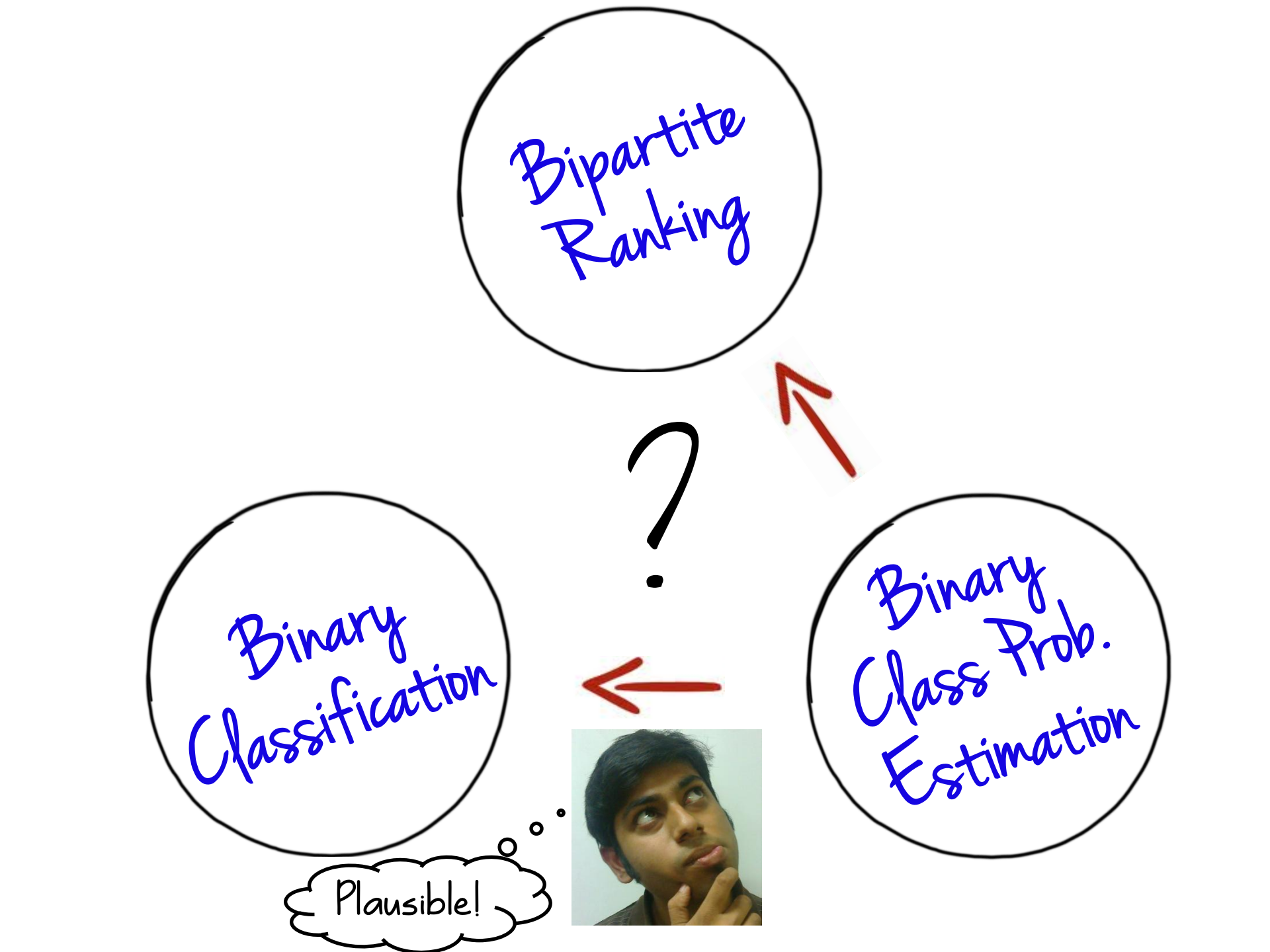
Binary Class Prob. Estimation

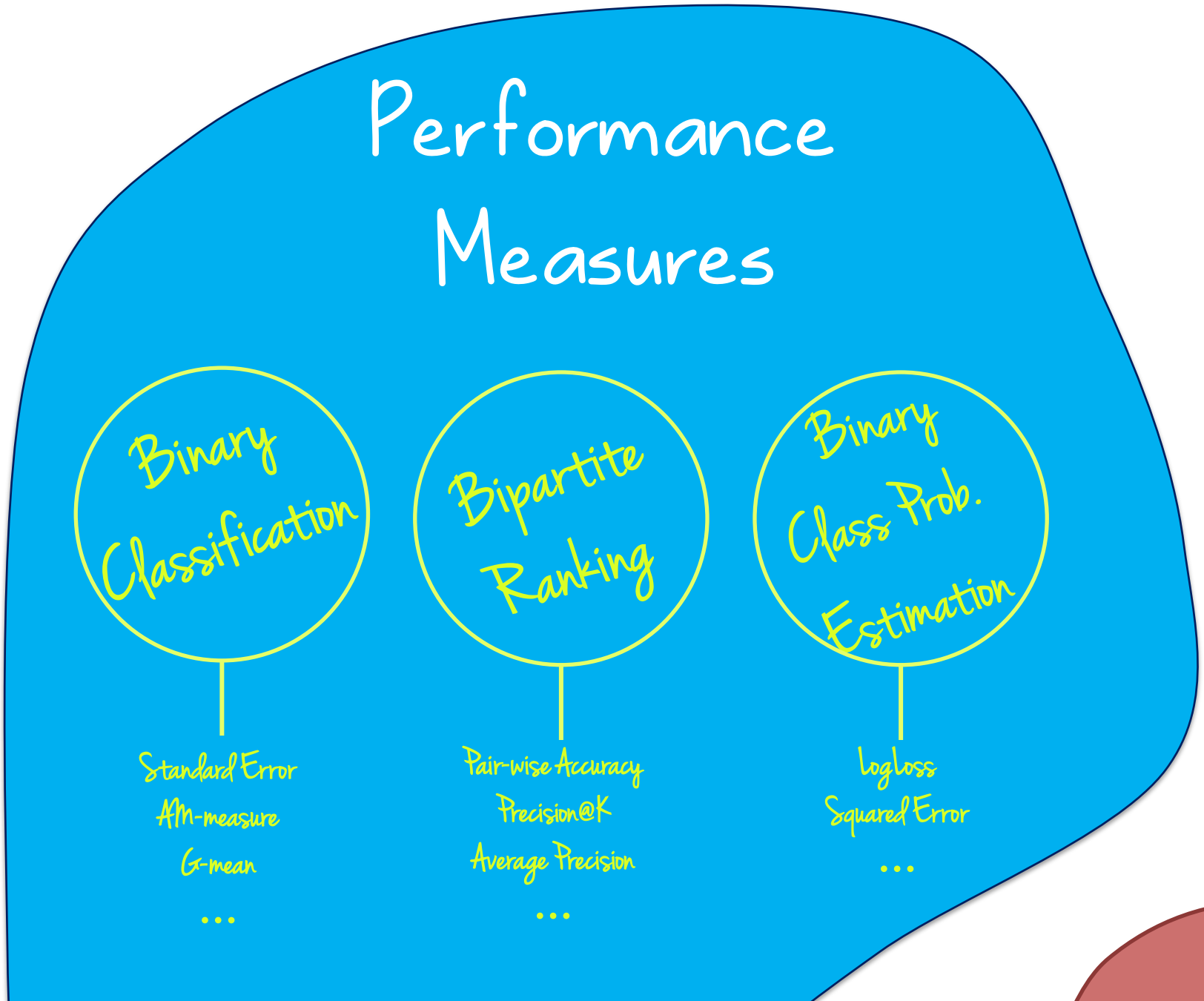
Bipartite Ranking

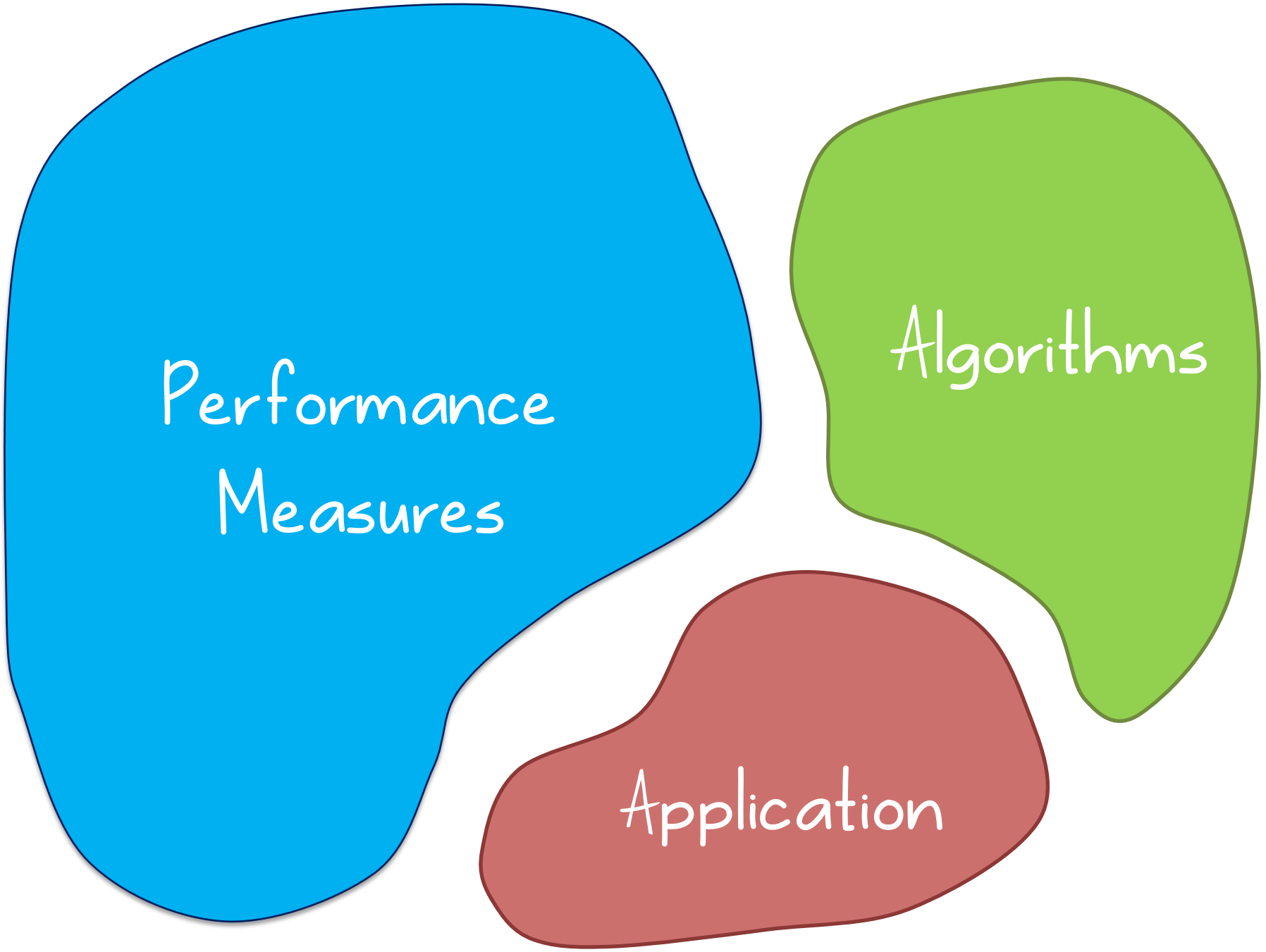
?

Binary Classification

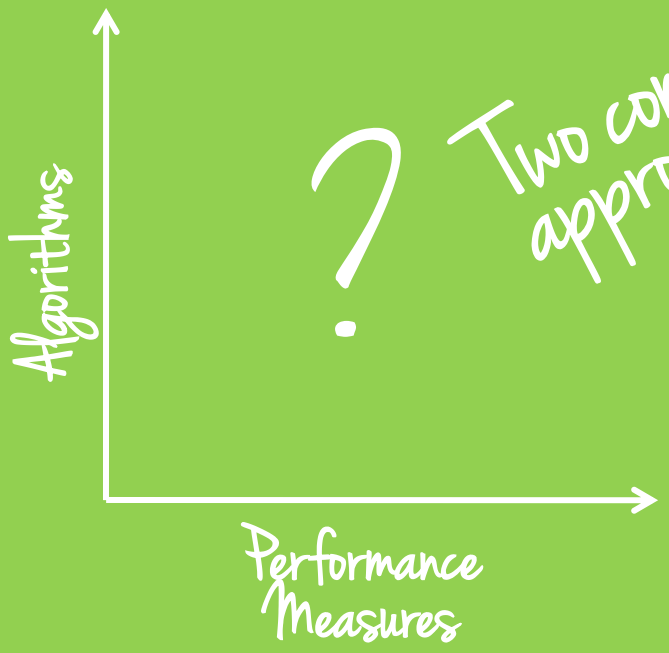
Binary Class Prob. Estimation


Bipartite Ranking
?
Binary Classification
Binary Class Prob. Estimation
Plausible!


Performance Measures
Binary Classification
Standard Error
AM-measure
G-mean
...
Bipartite Ranking
Pair-wise Accuracy
Precision@K
Average Precision
...
Binary Class Prob. Estimation
Logloss
Squared Error
...


Performance Measures
Algorithms
Application

Algorithms

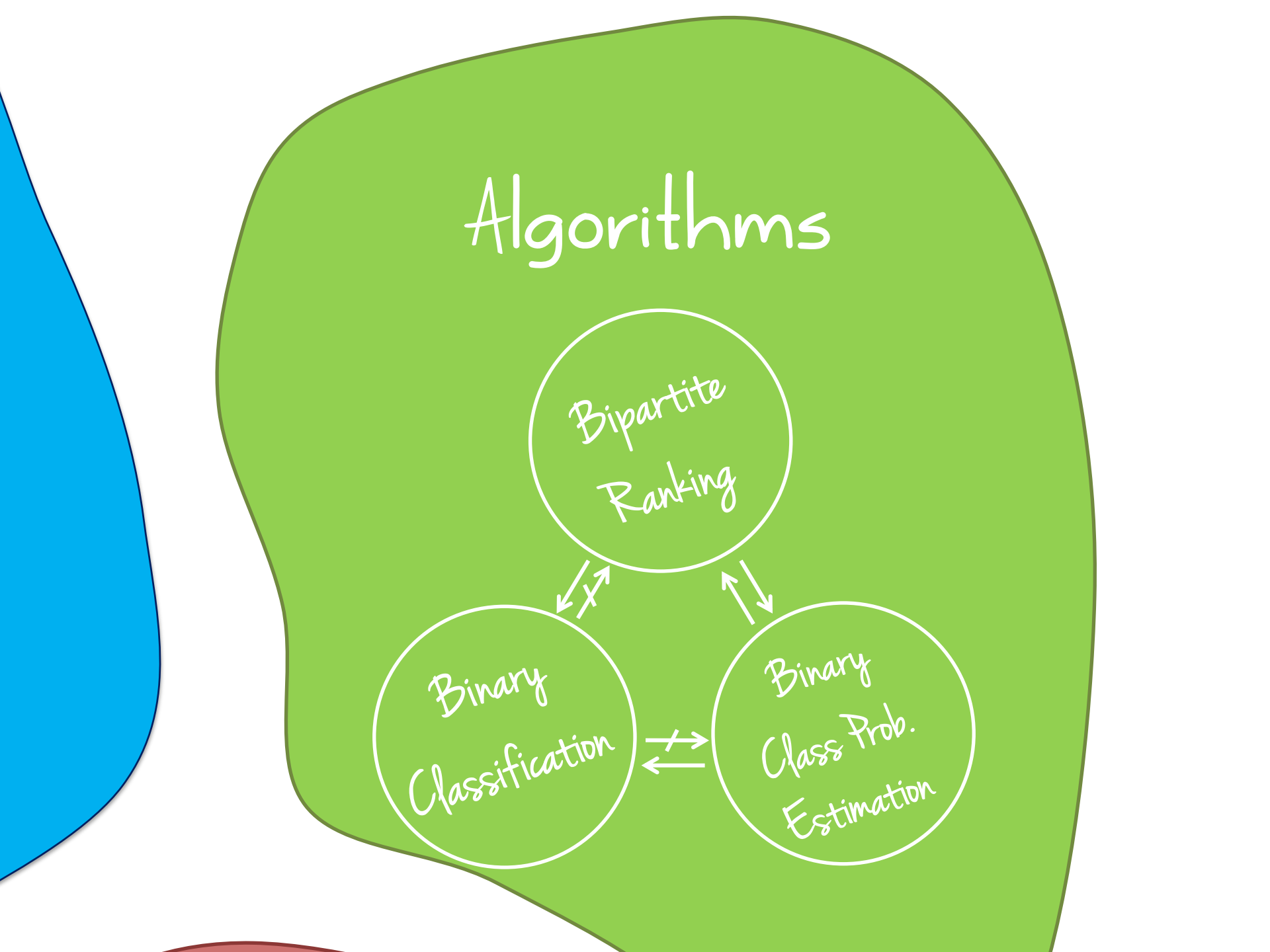


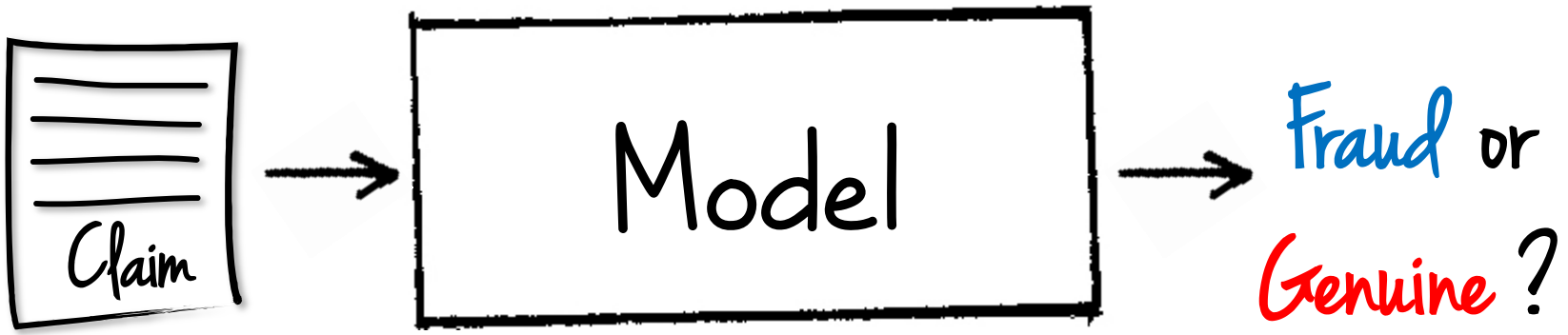
?

Two common approaches

Algorithms

A Case Study


Algorithms
Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation

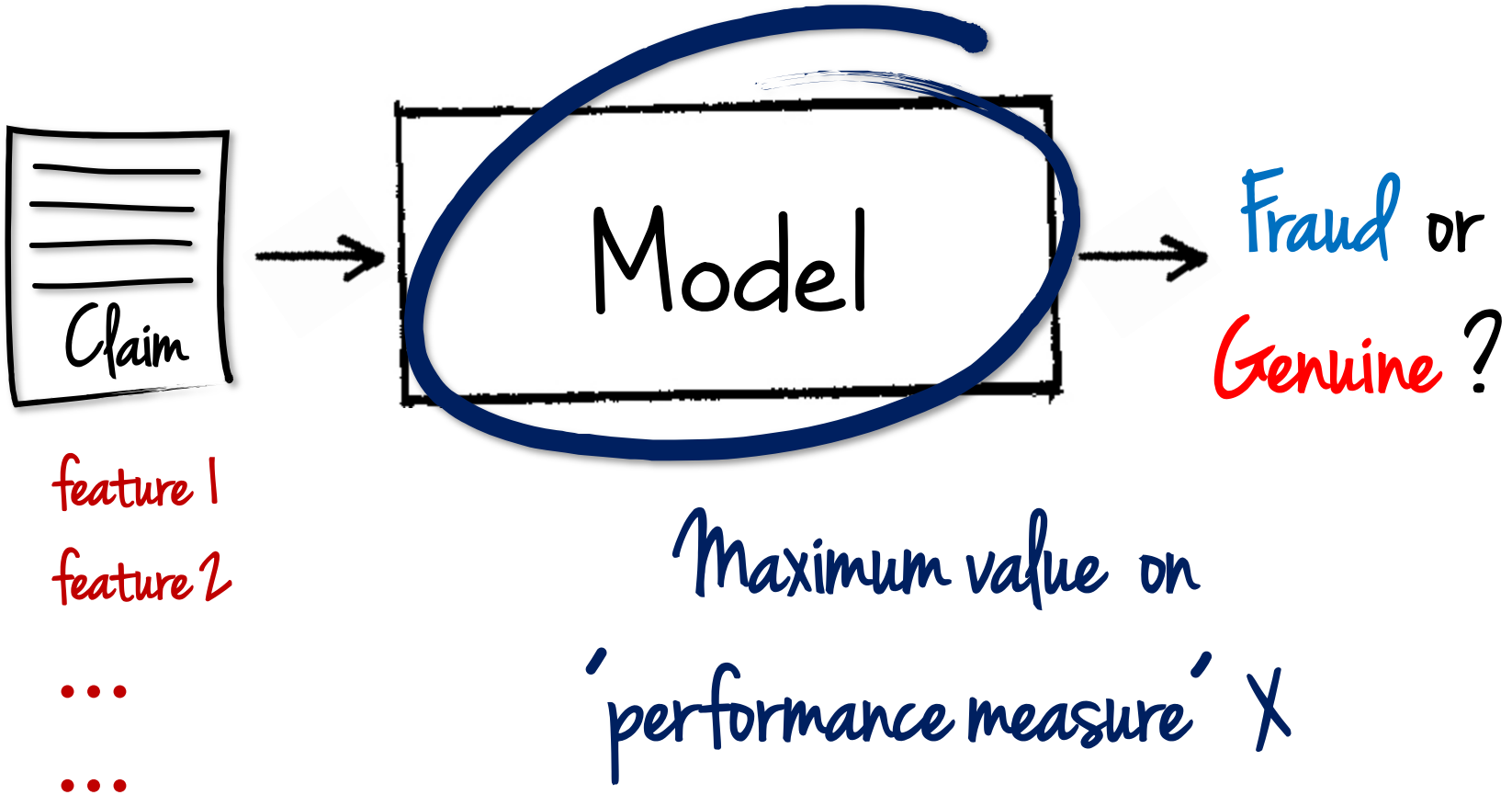


feature 1

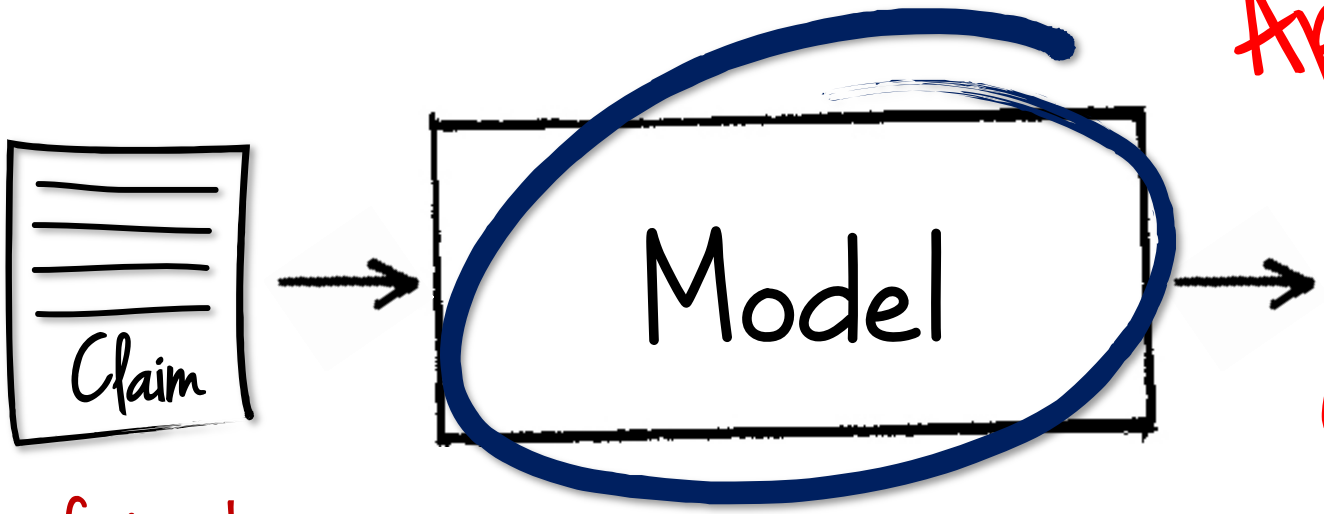
feature 2

...

...


Claim
Model
Fraud or
Genuine ?
feature 1
feature 2
...
...
Maximum value on
'performance measure' X

Two Common Approaches



feature 1

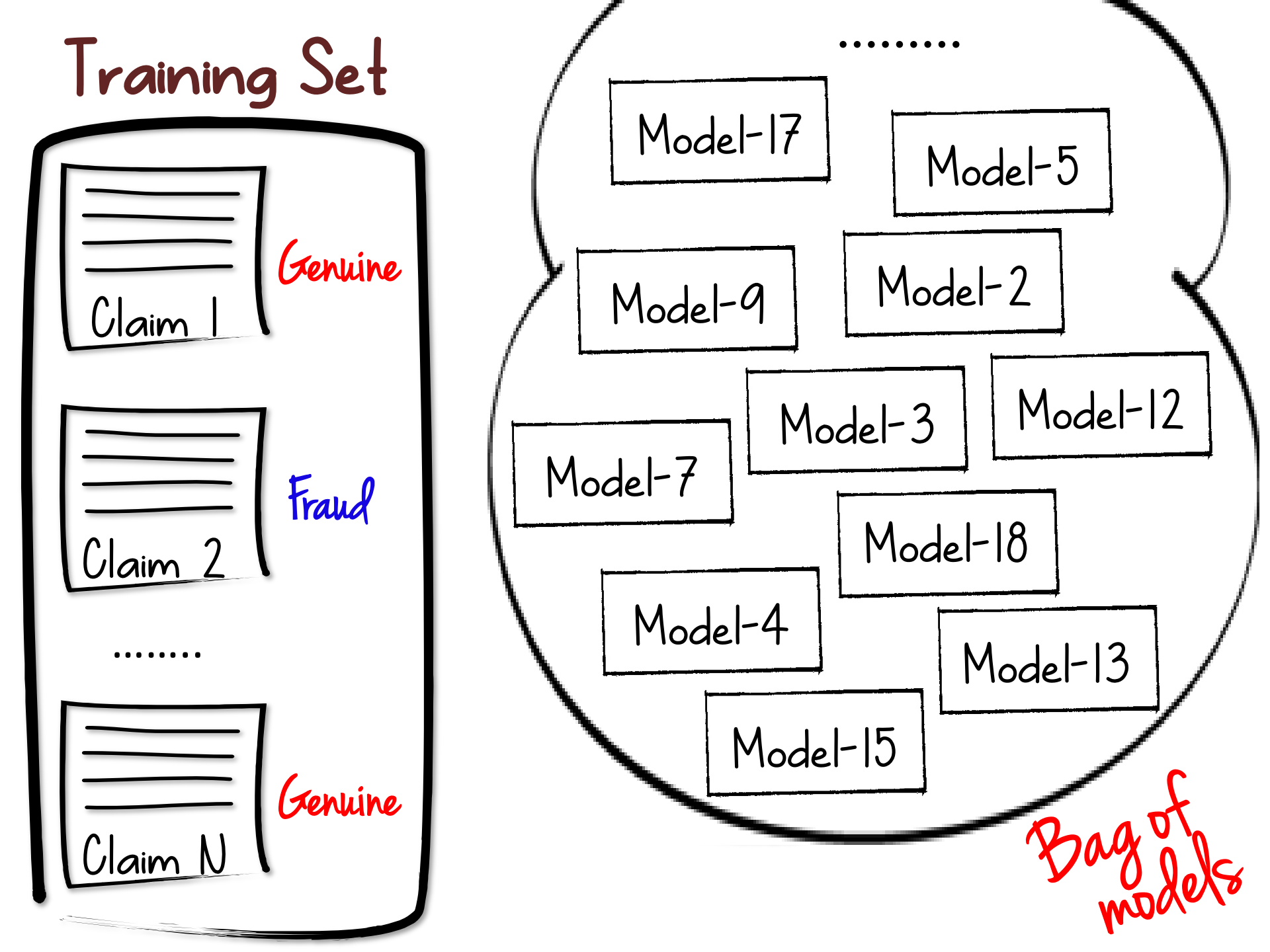
feature 2

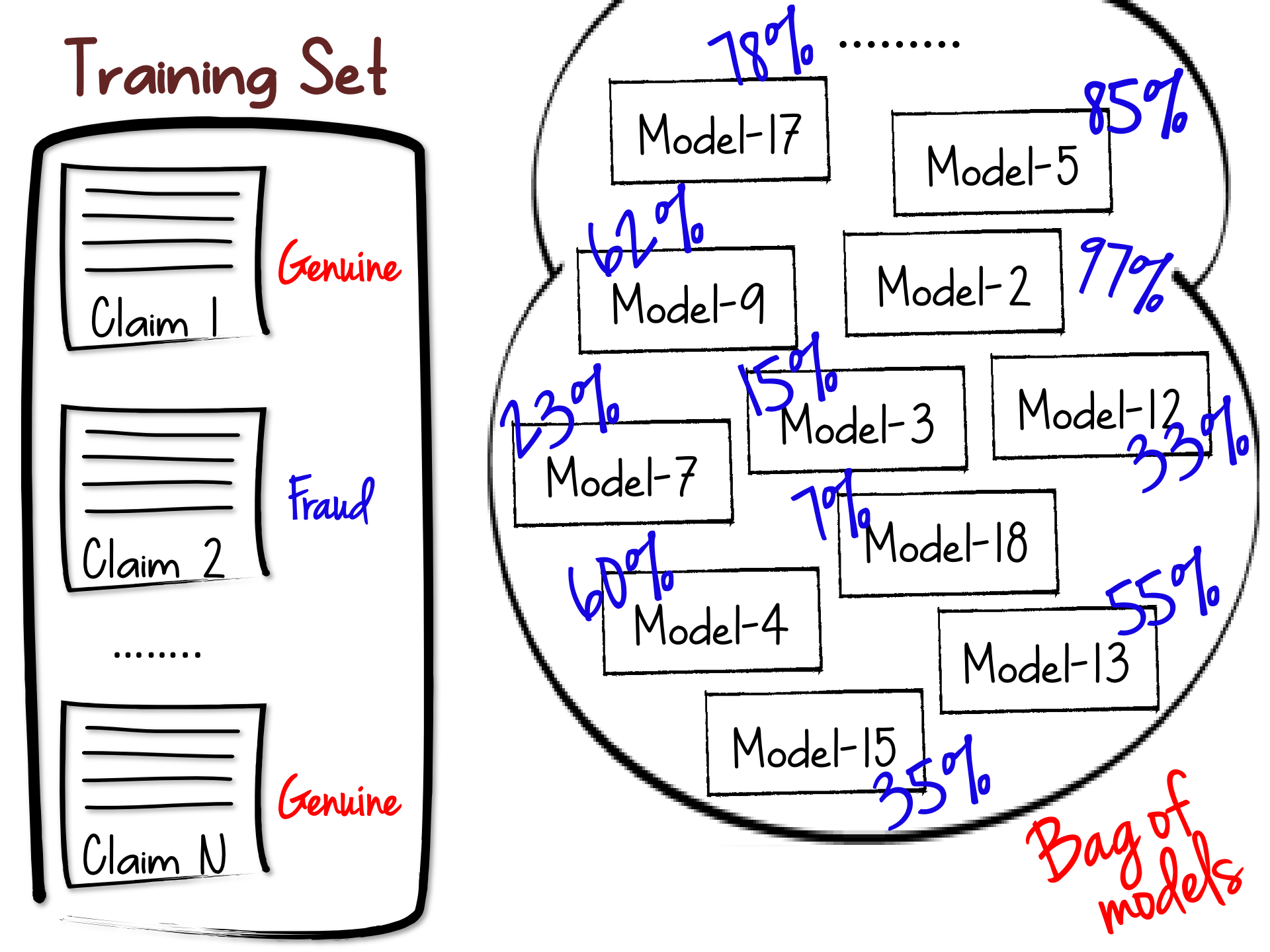
...

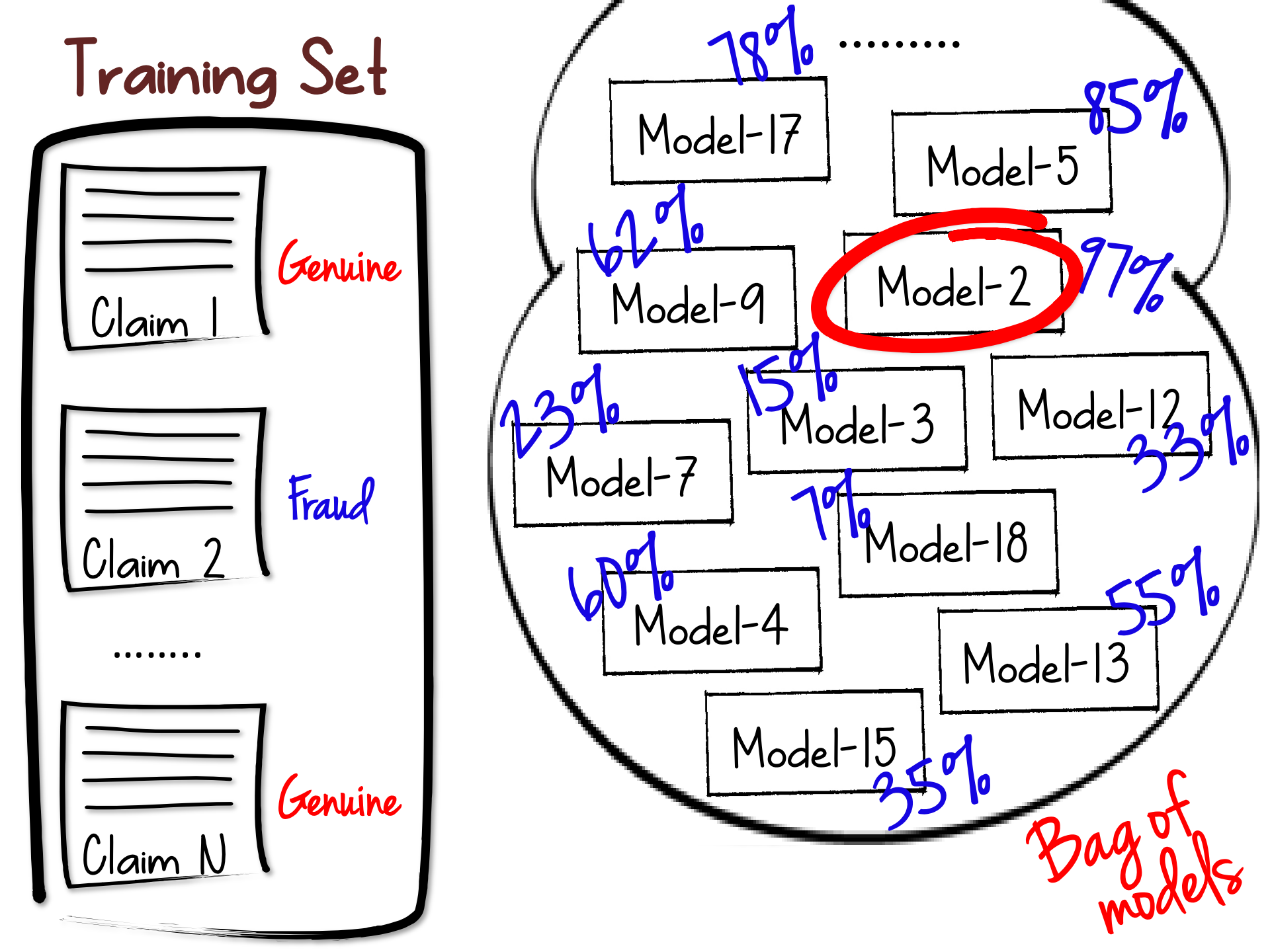
...

Maximum value on 'performance measure' X

Approach 1

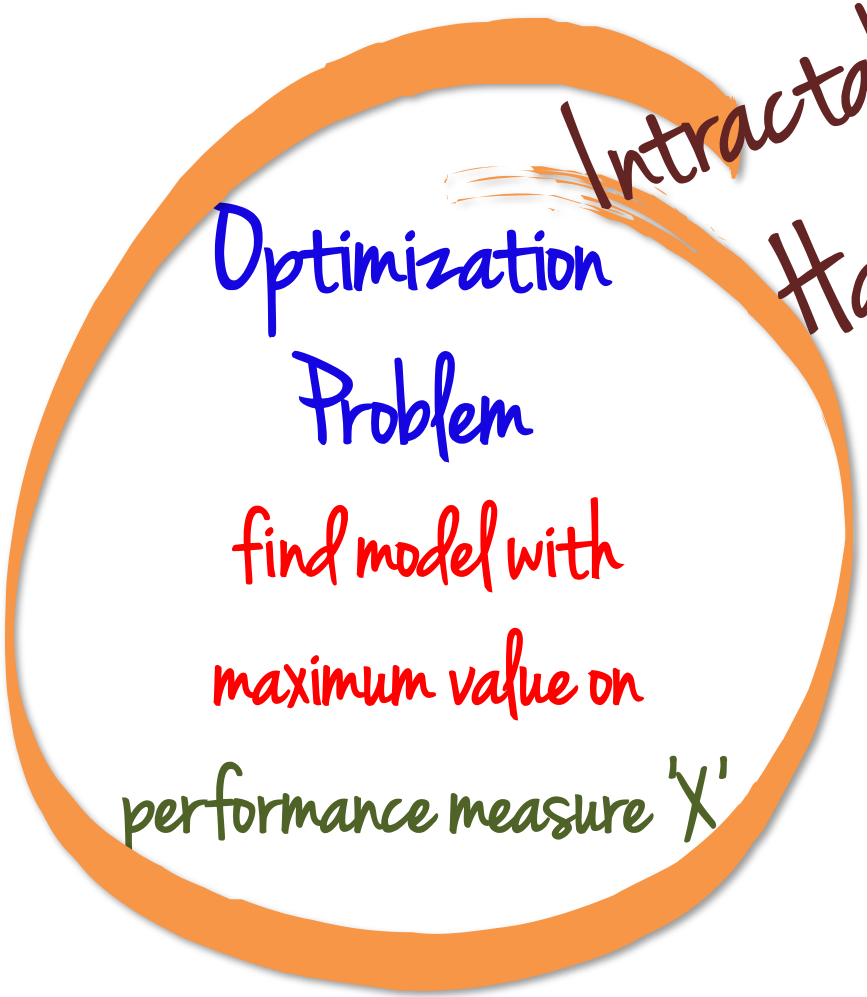

Training Set
Claim 1
Genuine
Claim 2
Fraud
.....
Claim N
Genuine
.....
Model-17
Model-5
Model-9
Model-2
Model-3
Model-12
Model-7
Model-18
Model-4
Model-13
Model-15
Bag of models


Training Set
Claim 1
Genuine
Claim 2
Fraud
.....
Claim N
Genuine
78%
.....
Model-17
85%
Model-5
62%
Model-9
Model-2
97%
23%
15%
Model-3
Model-12
33%
Model-7
70%
Model-18
60%
Model-4
55%
Model-13
Model-15
35%
Bag of models


Training Set
Claim 1
Genuine
Claim 2
Fraud
.....
Claim N
Genuine
78%
.....
Model-17
85%
Model-5
62%
Model-9
Model-2
97%
23%
15%
Model-3
Model-12
Model-7
33%
70%
Model-18
60%
Model-4
55%
Model-13
Model-15
35%
Bag of models

Optimization Problem

find model with maximum value on performance measure 'X'



Proxy
Optimization
Problem
(often possible*)

Approach I

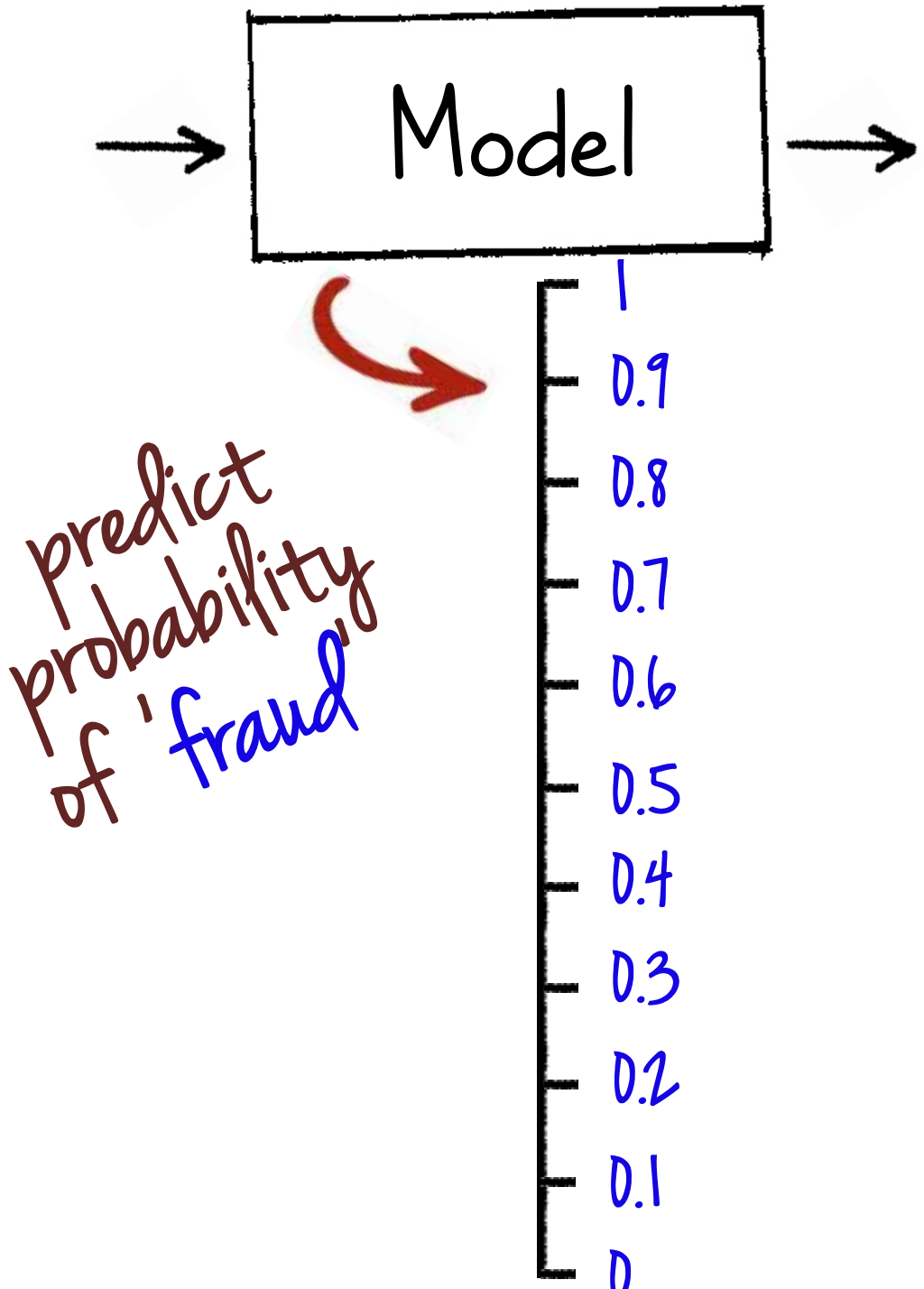
for different performance measures

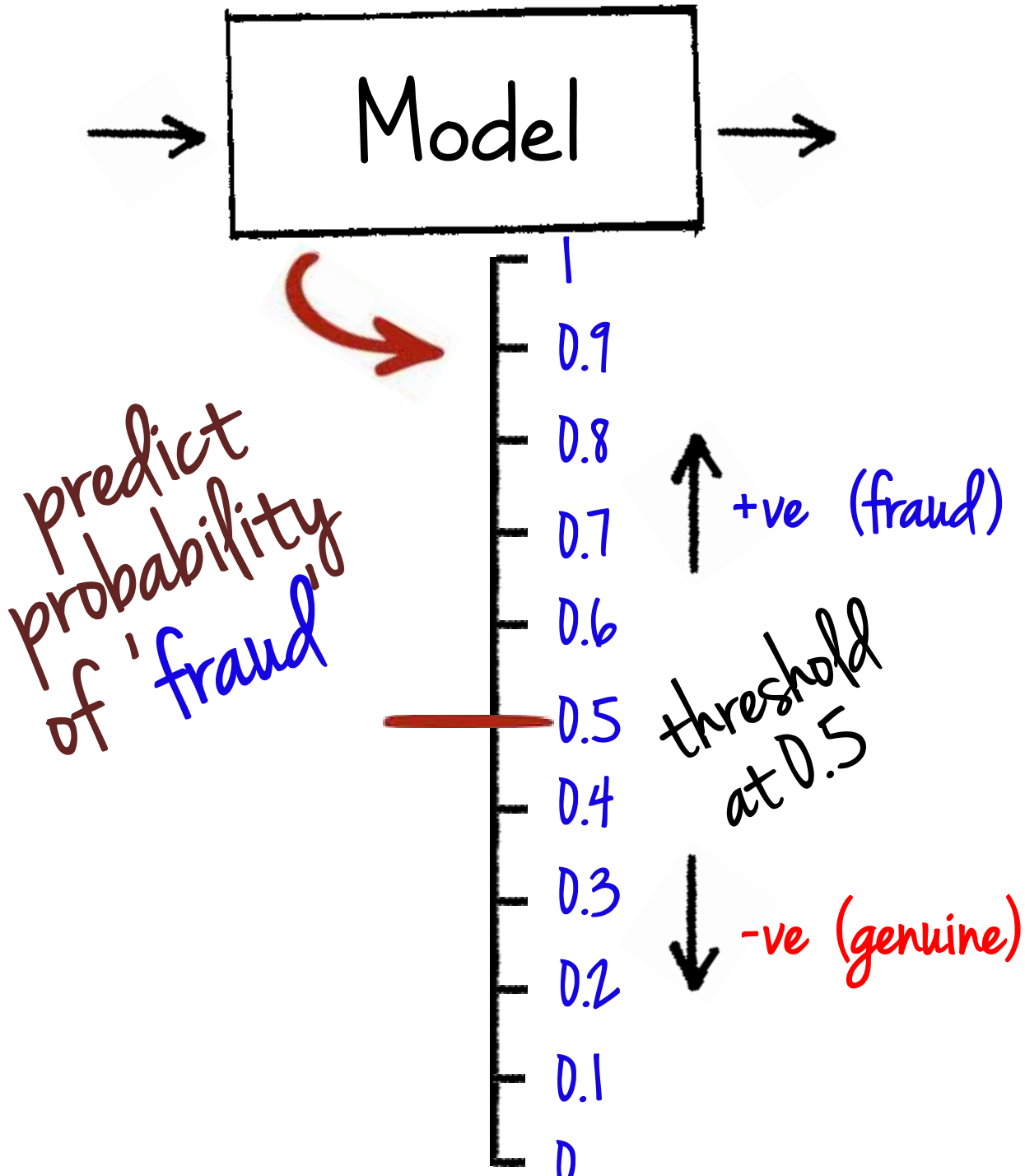
Performance Measure	Run-time	References
F-measure PBREP Precision@K	$O(N^2)$	Joachims (2005);
AUC Partial AUC Pos@Top	$O(N \log N)$	Joachims (2005, 2006); Narasimhan & Agarwal (2013 a,b)
Average Precision	$O(N^2)$	Yue et al. (2007)
Discounted Cumulative Gain (DCG) *	$O(Nk)$	Chakrabarti et al. (2008)
Mean Reciprocal Rank (MRR) *	$O(N \log N + k^2)$	Chakrabarti et al. (2008)

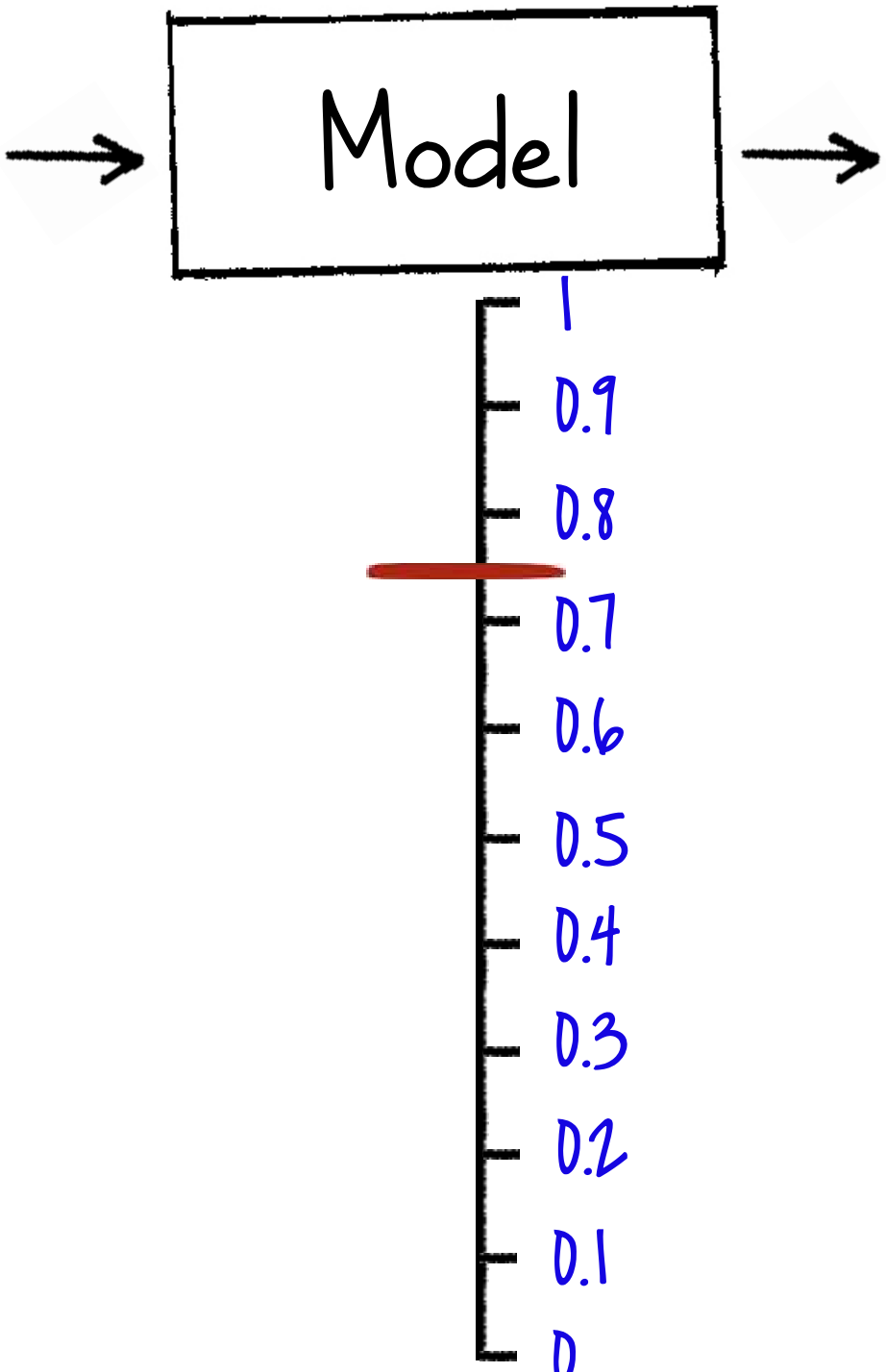
* performance measure truncated to top 'k' positions

Approach II

(Plug-in)


Model
predict probability of 'fraud'
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

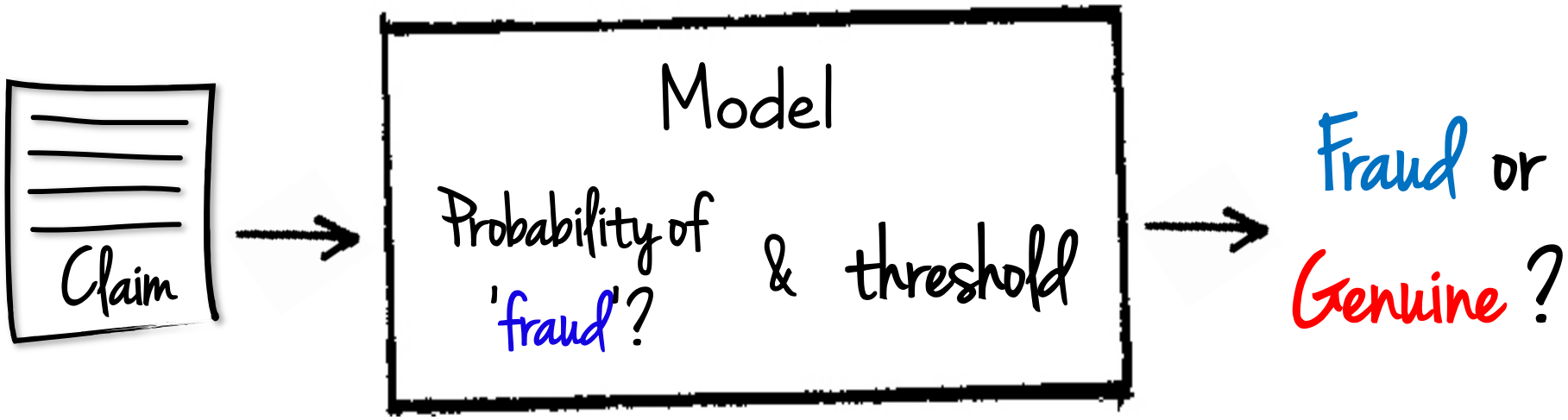

Model
predict probability of 'fraud'
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
+ve (fraud)
threshold at 0.5
-ve (genuine)

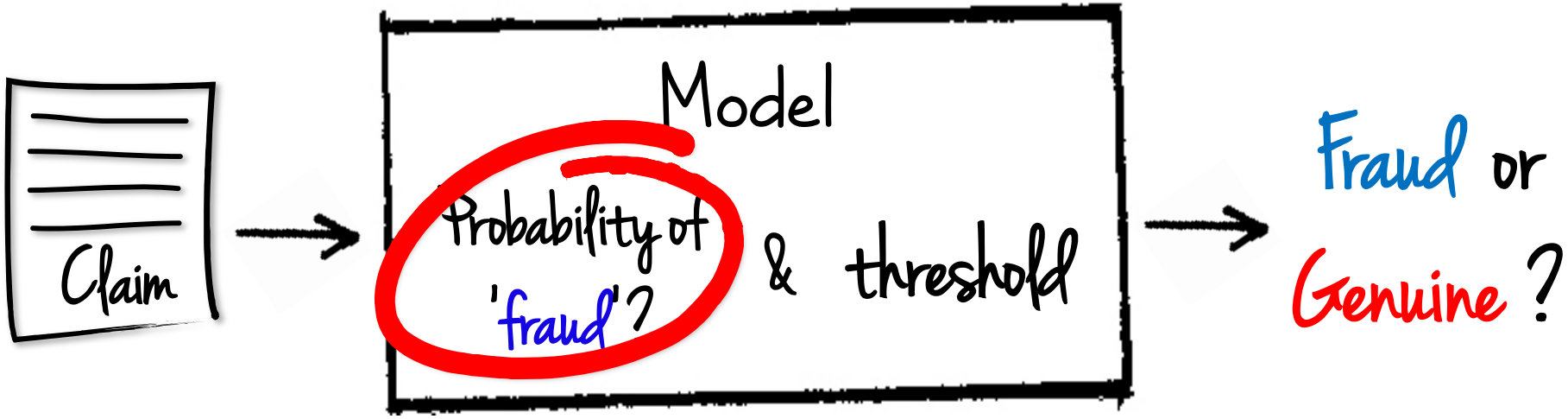


Choose threshold

that maximizes

'performance measure' X


Claim
Model
Probability of 'fraud'?
&
threshold
Fraud or Genuine?

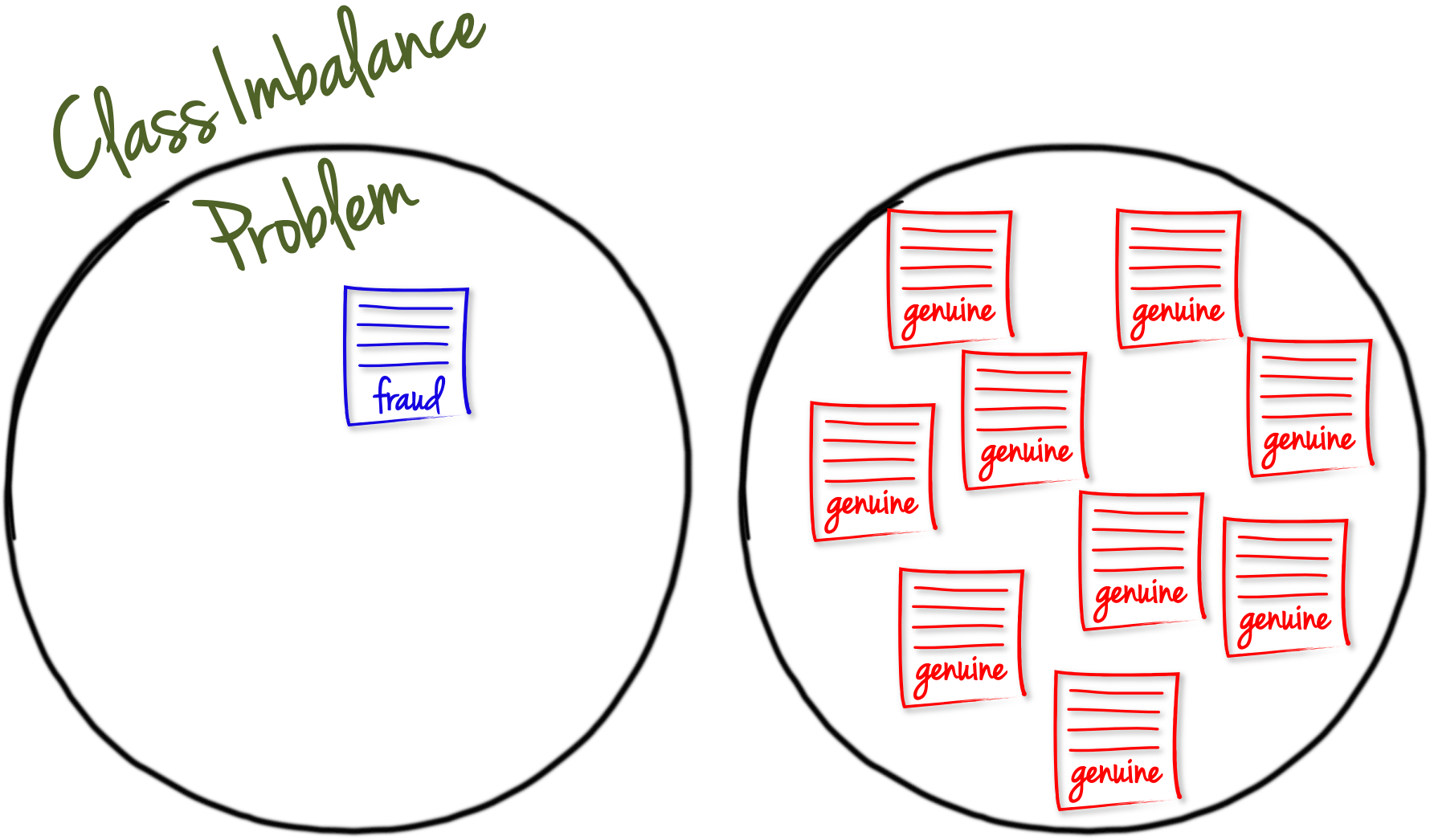

Claim
Model
Probability of 'fraud'?
& threshold
Fraud or Genuine?

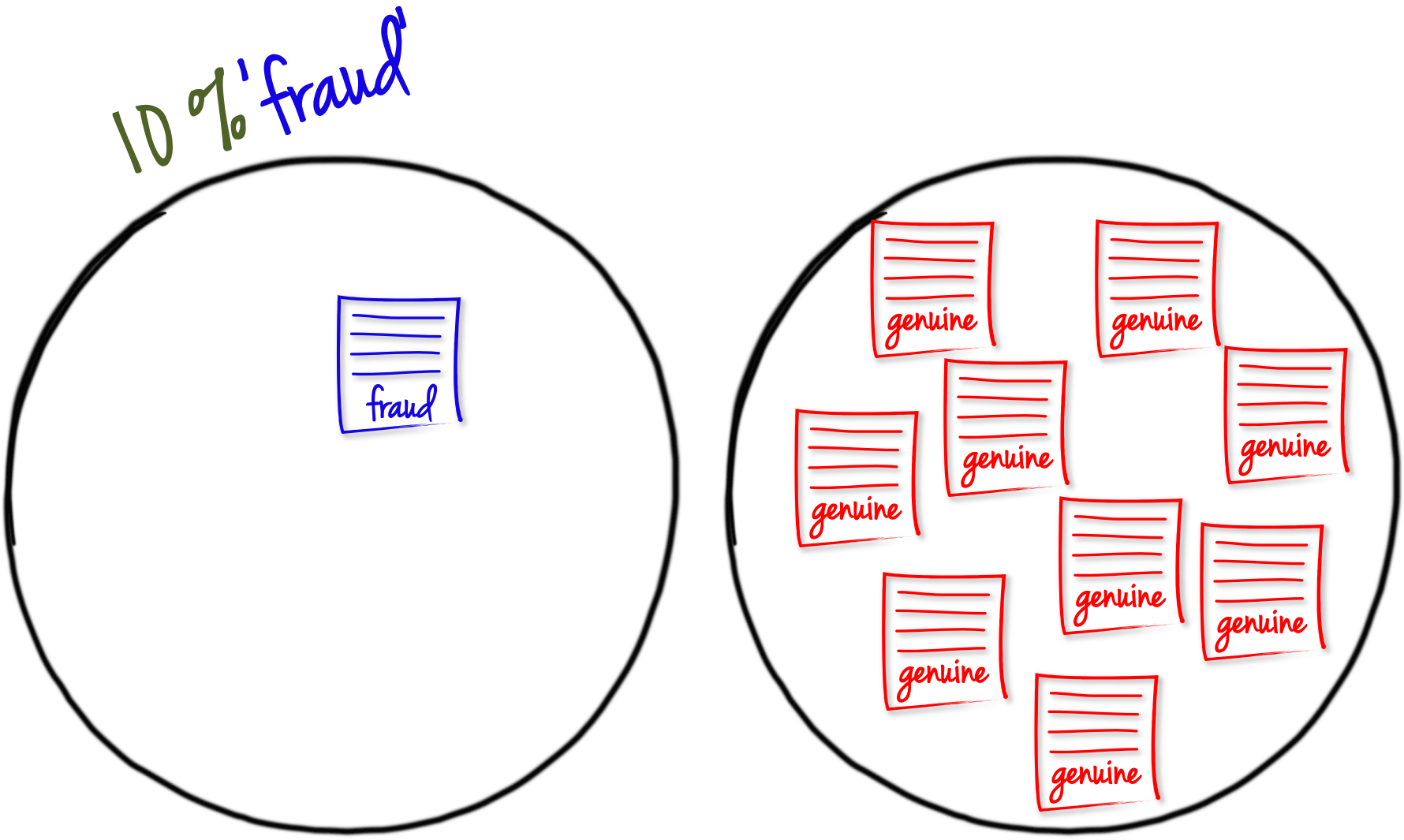
Example

AM-measure

$$\frac{\text{True Positive Rate} + \text{True Negative Rate}}{2}$$

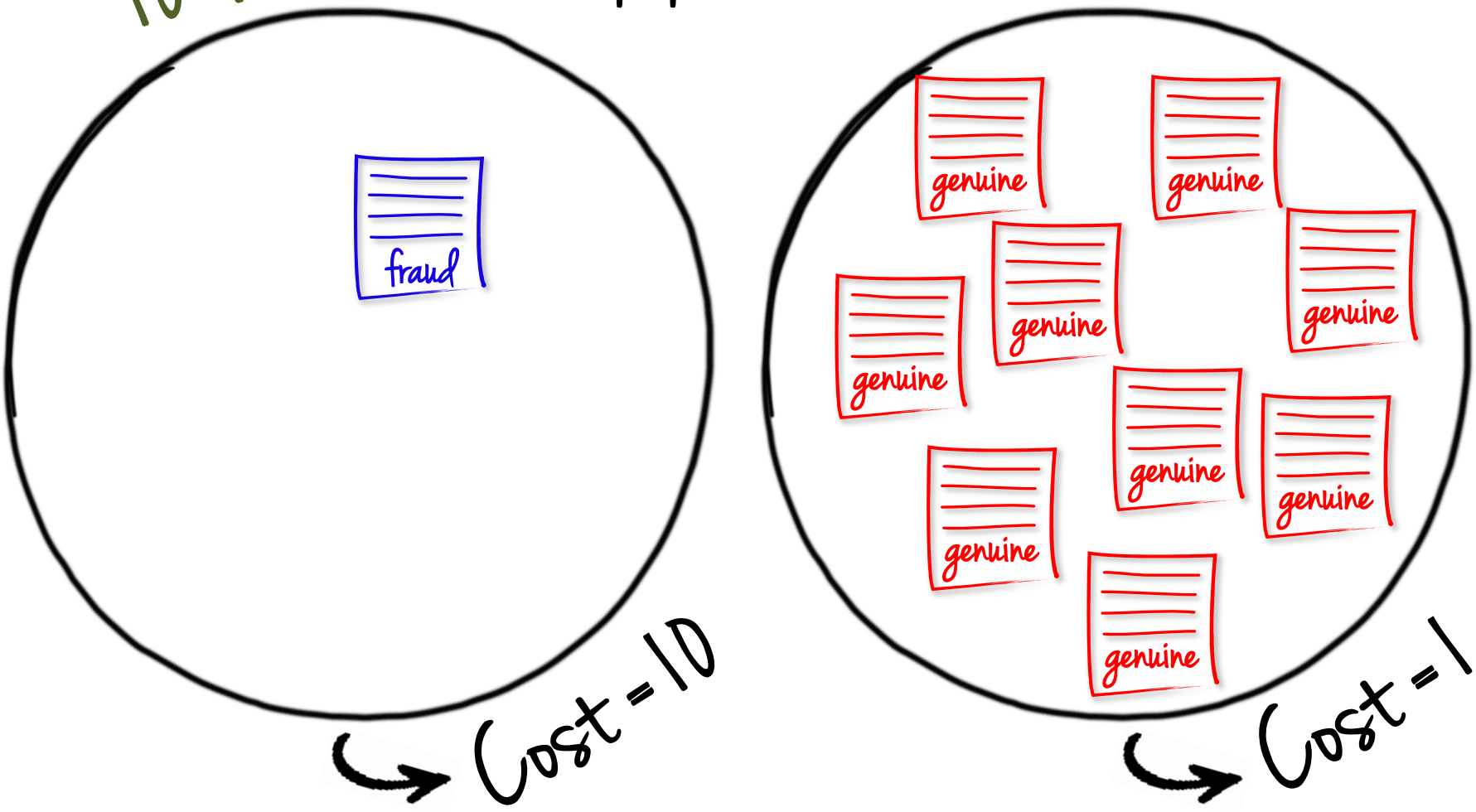
Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. "*On the statistical consistency of algorithms for binary classification under class imbalance* ", ICML 2013.


Class Imbalance Problem
fraud
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine


10 % fraud
fraud
genuine
genuine
genuine
genuine
genuine
genuine
genuine
genuine

Approach 1

10 % 'fraud'



Approach 1

10 % 'fraud'

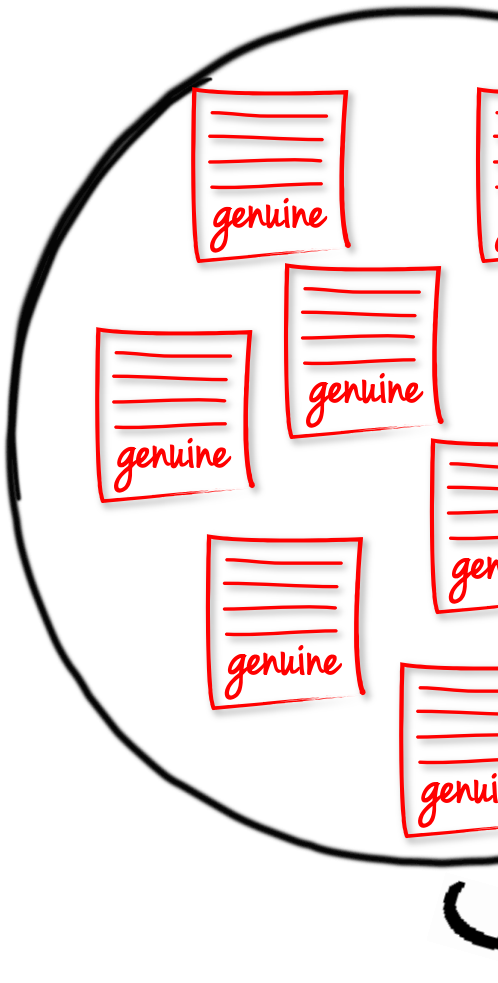
Balanced Logistic Regression

Balanced SVM

Balanced AdaBoost

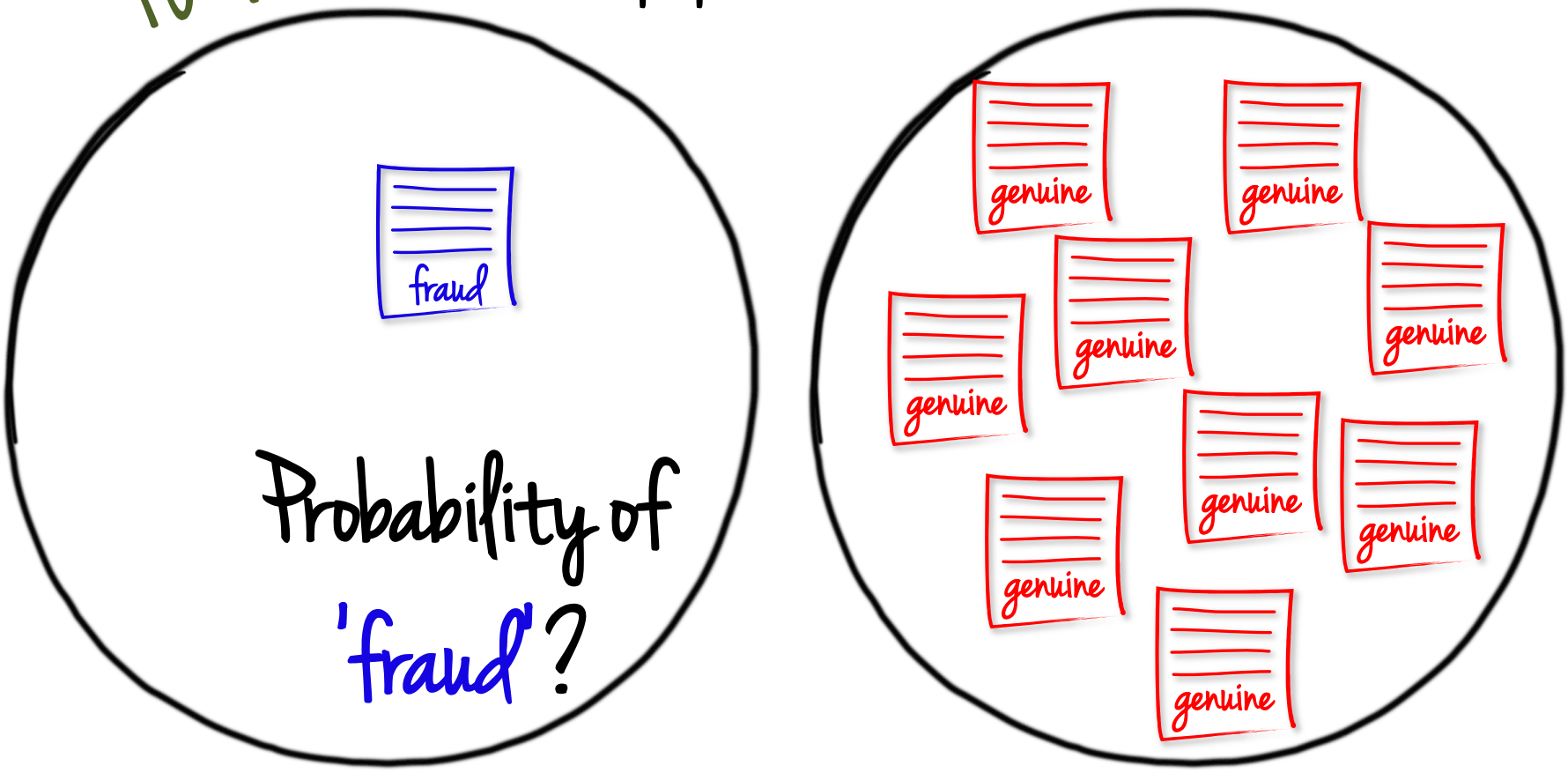
Balanced Decision Trees

...



Approach II

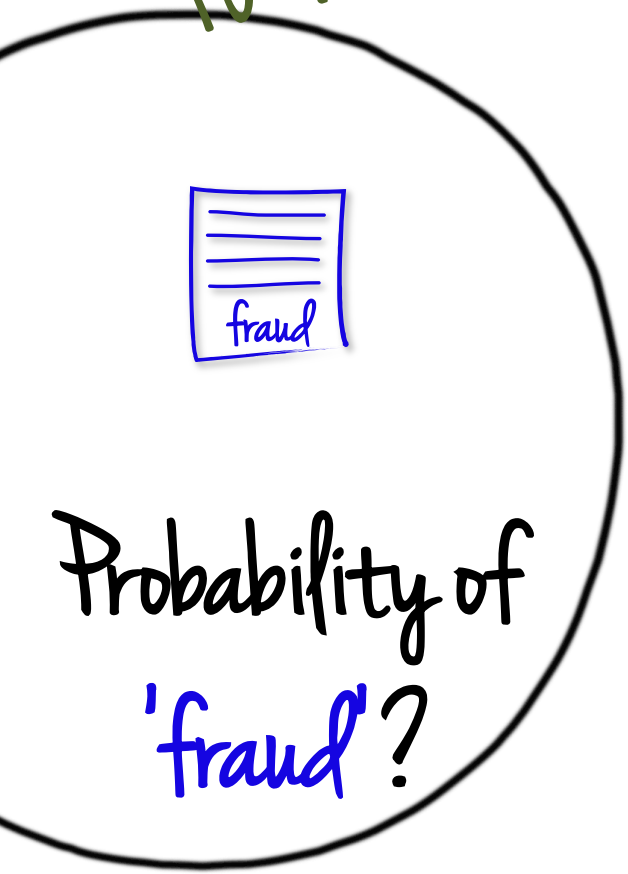
10 % 'fraud'

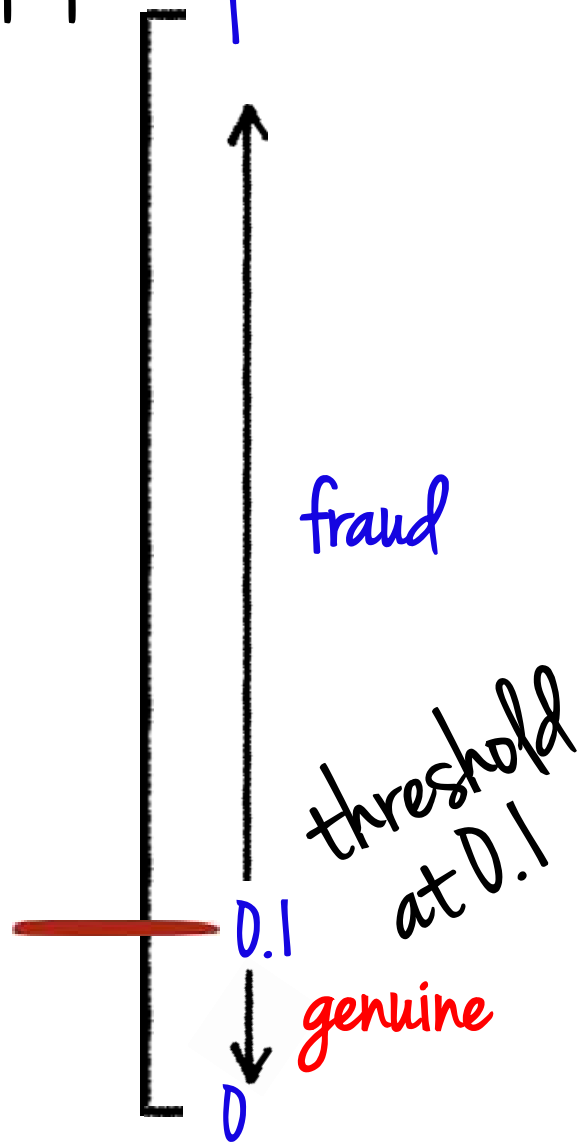


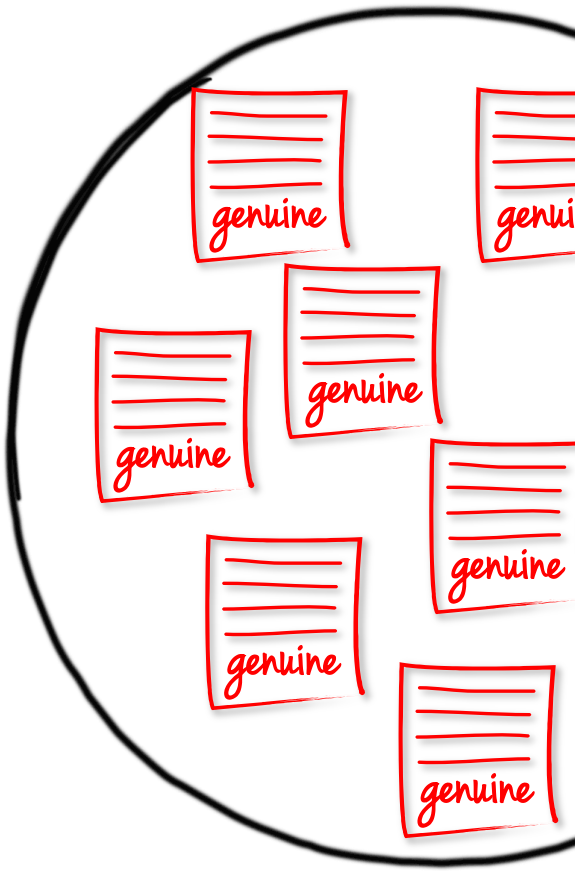
Probability of 'fraud'?

Approach II

10 % 'fraud'







Experiments

AM-measure

$$\frac{\text{True Positive Rate} + \text{True Negative Rate}}{2}$$

Experiments

Thyroid Diagnosis

~2% positive

	AM-measure
Standard Method	84.91
Approach 1	**97.66**
Approach 2	96.69

Experiments

Drug Discovery

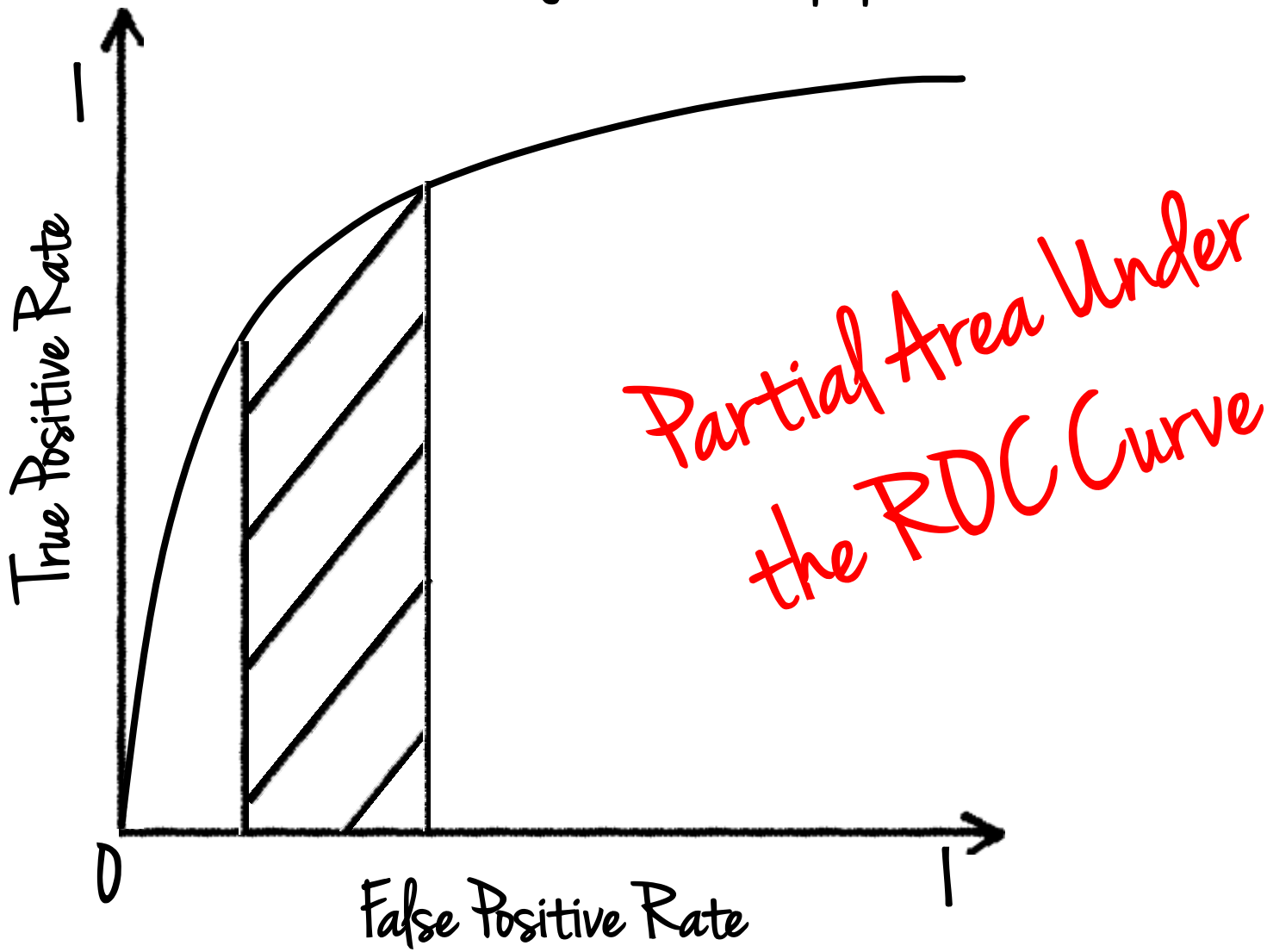
~2% positive

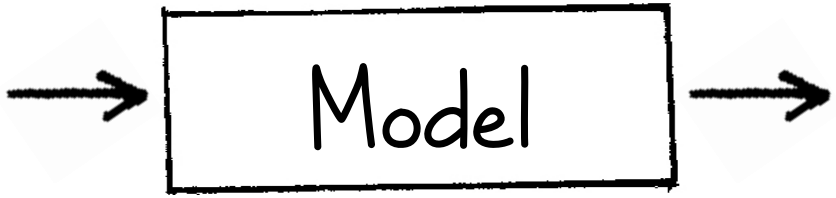
	AM-measure
Standard Method	82.74
Approach 1	89.87
Approach 2	**91.55**

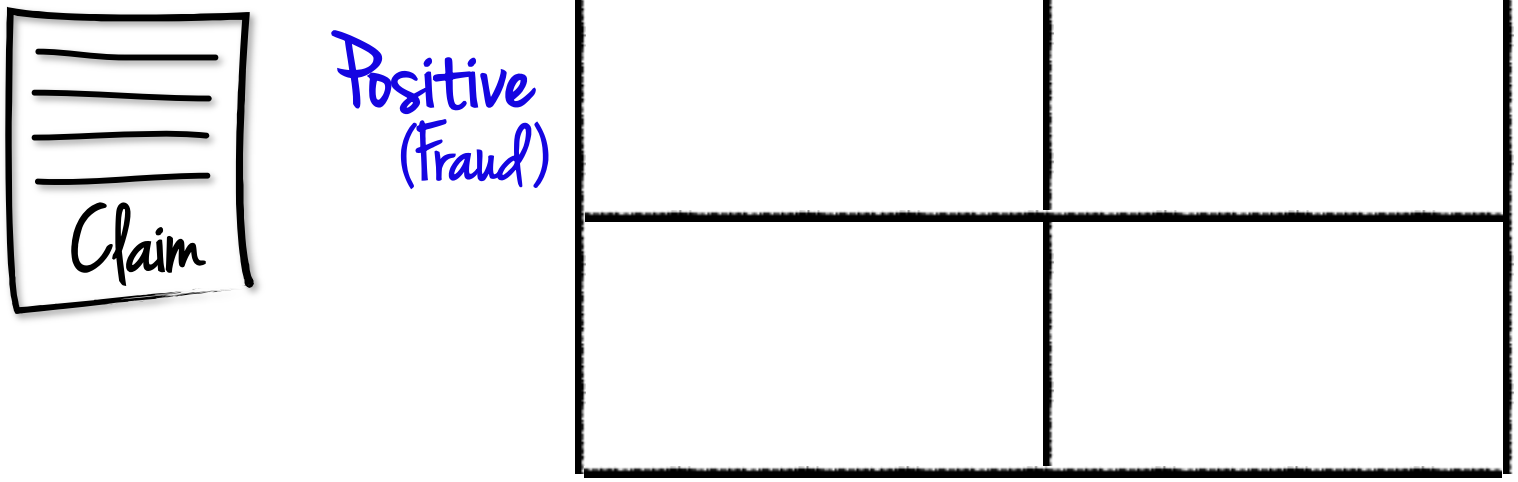
Algorithms

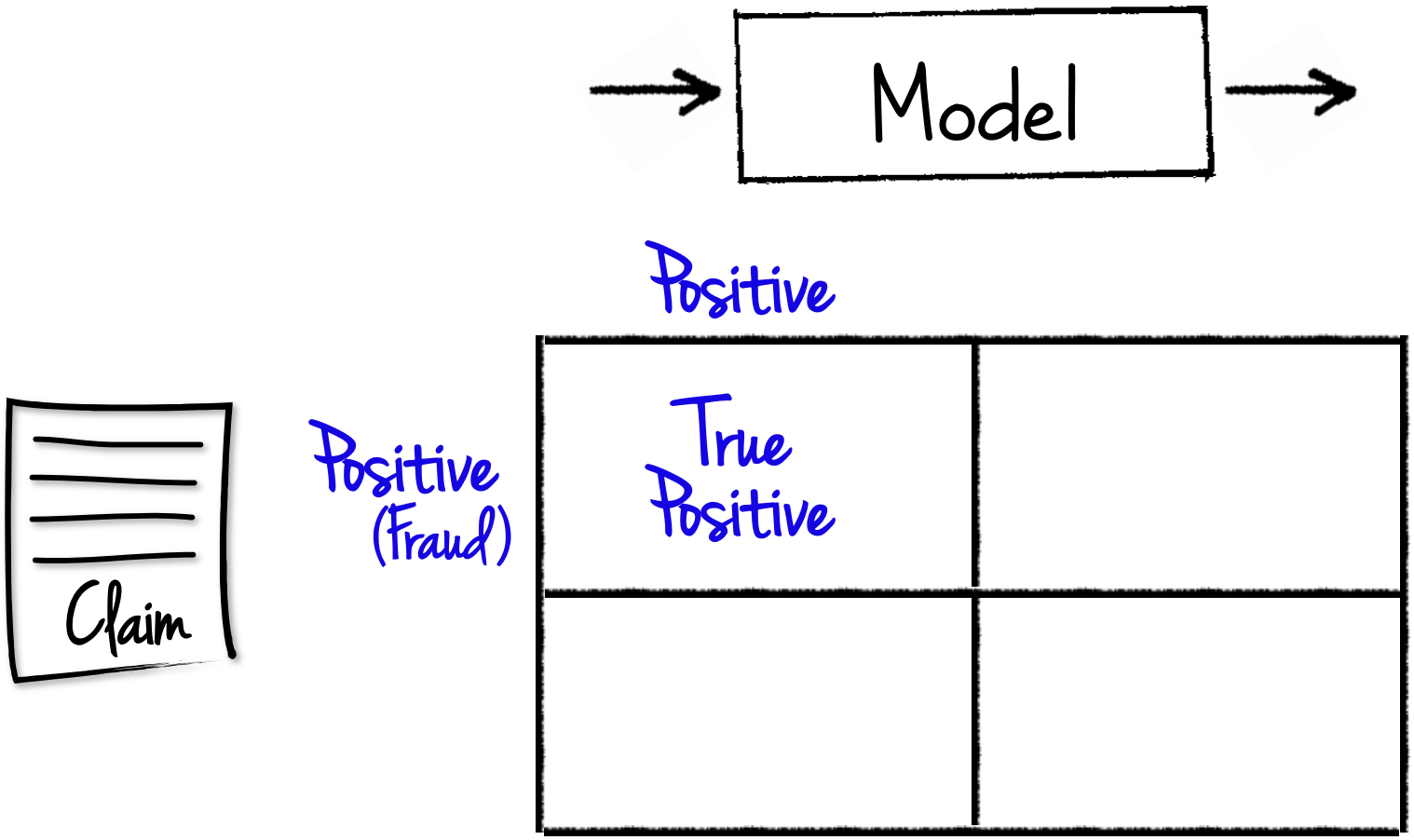
A Case Study for Approach I

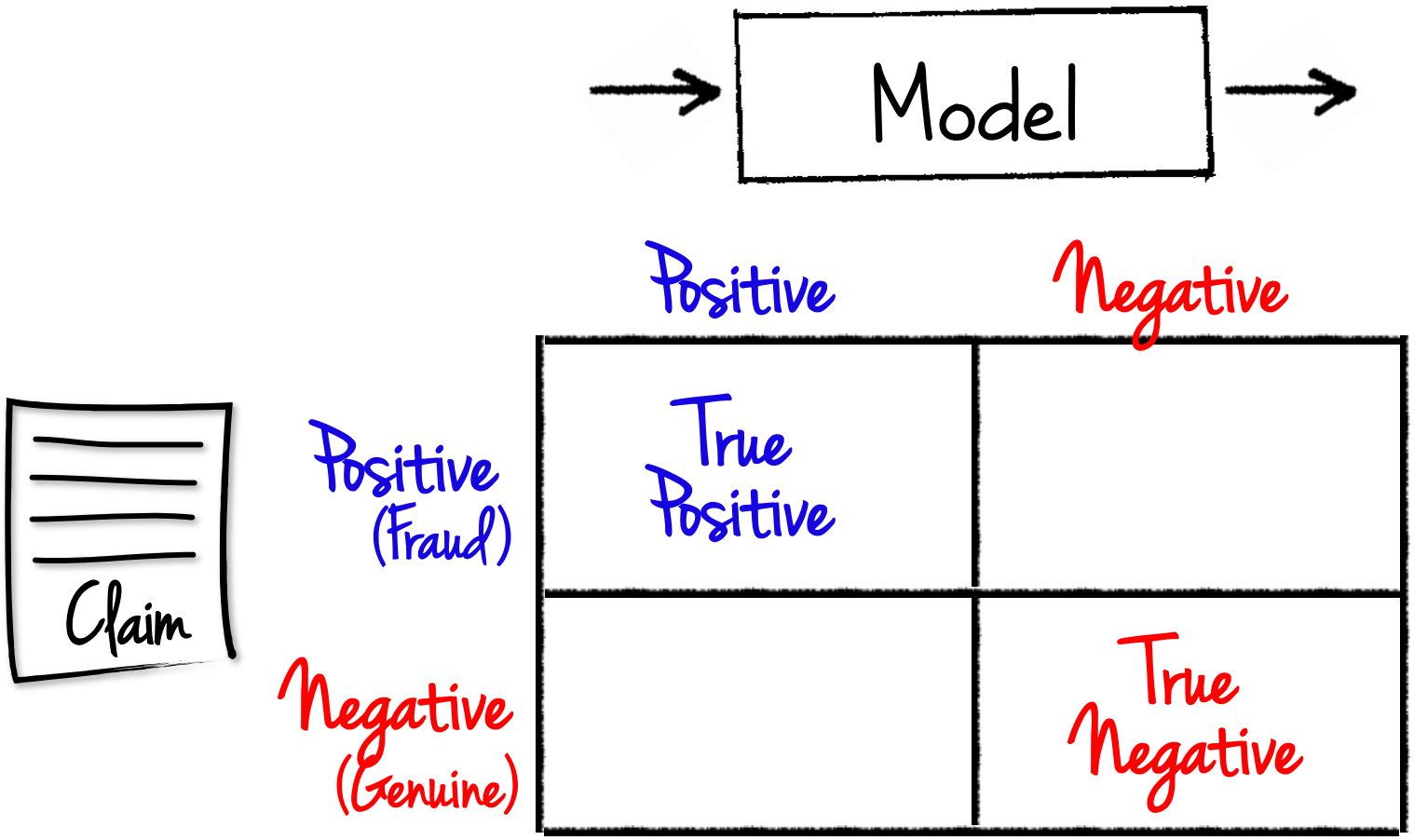
A Case Study for Approach II

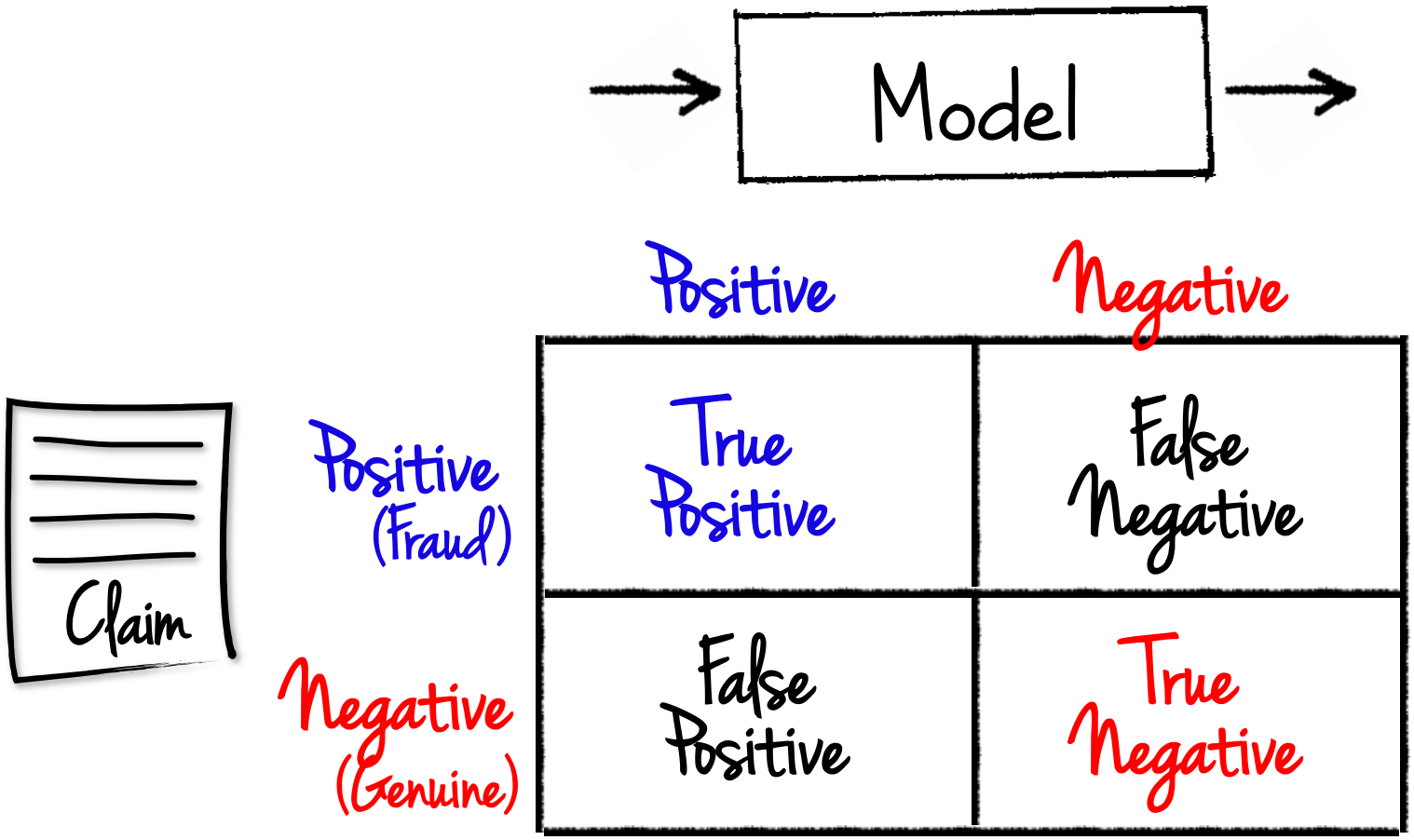


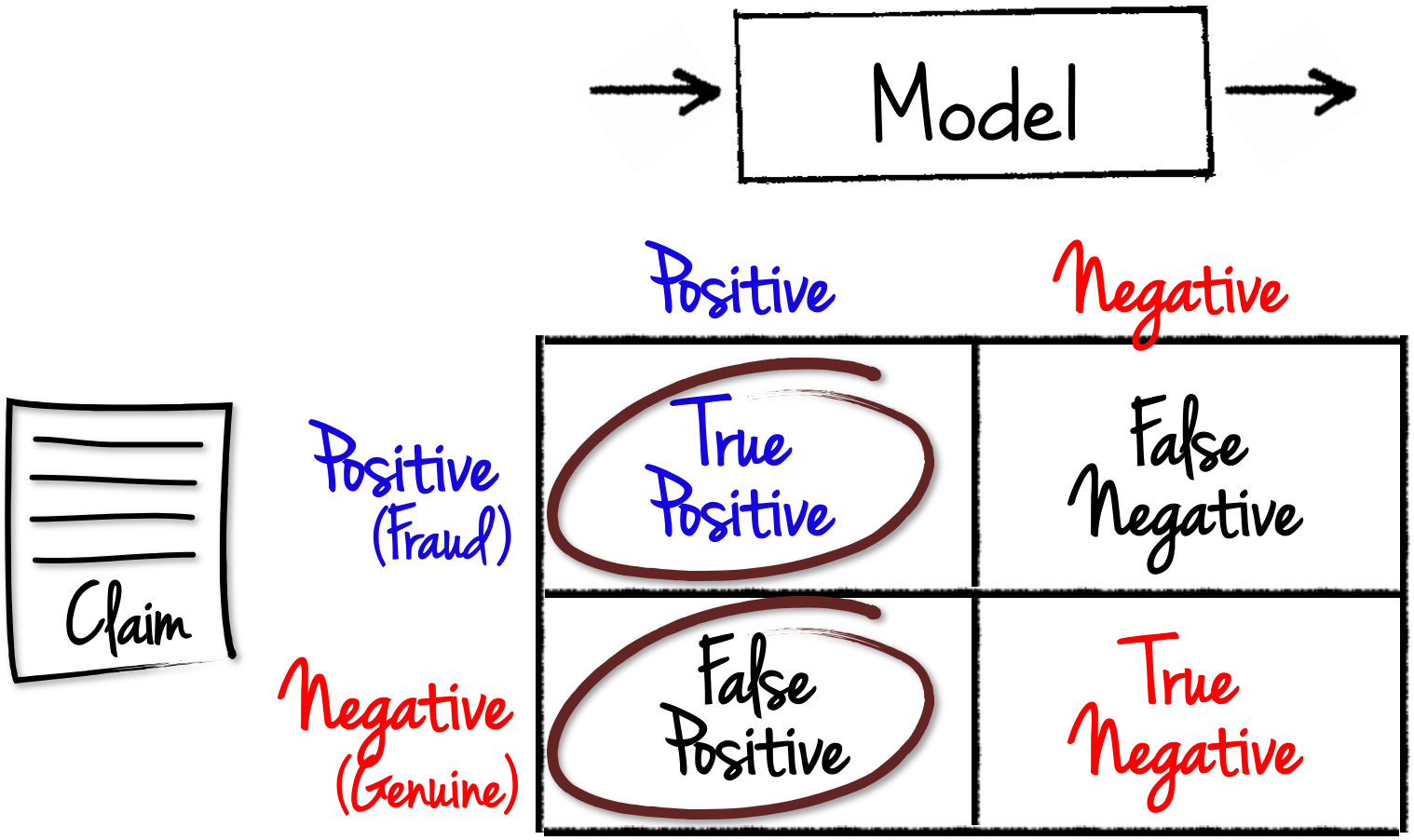

Model


Positive
Positive
(Fraud)
Claim

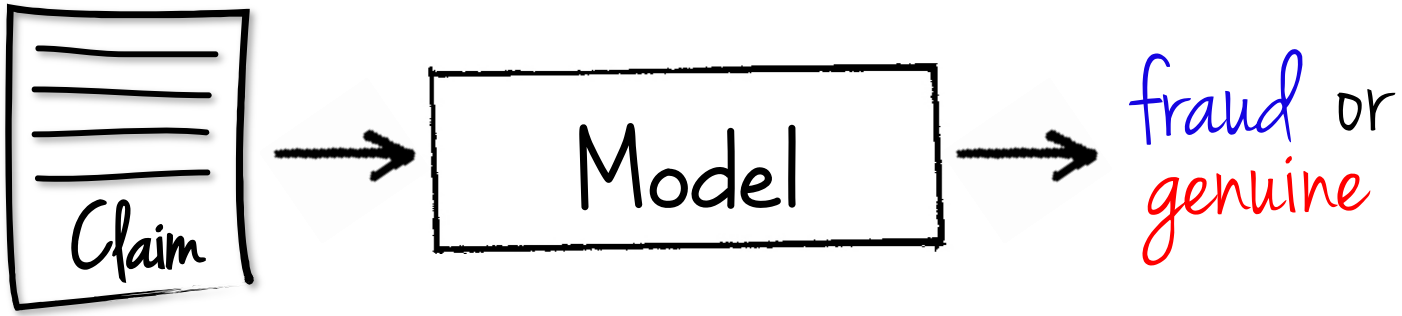

Model
Positive
Positive
(Fraud)
True
Positive
Claim


Model
Positive
Negative
Claim
Positive
(Fraud)
True
Positive
Negative
(Genuine)
True
Negative


Model
Positive
Negative
Claim
Positive
(Fraud)
Negative
(Genuine)
True
Positive
False
Negative
False
Positive
True
Negative


Model
Positive
Negative
Claim
Positive (Fraud)
Negative (Genuine)
True Positive
False Negative
False Positive
True Negative

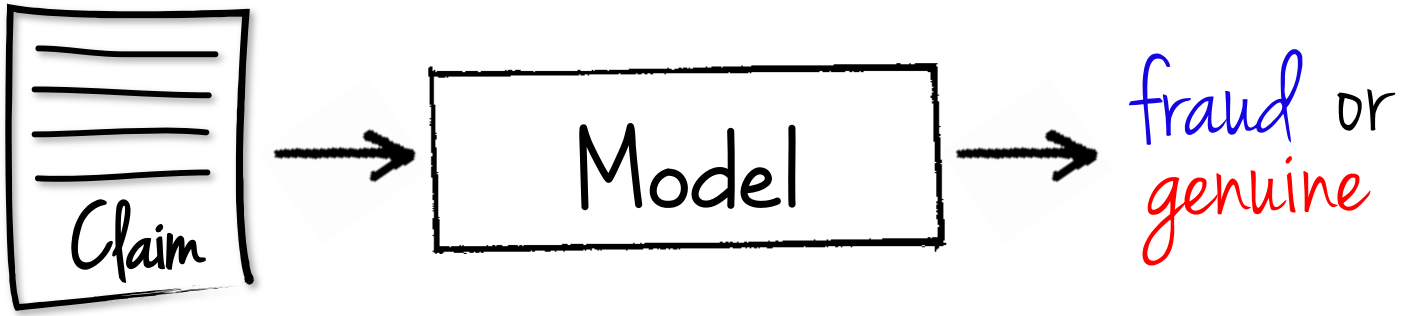
Fraud Detection



http://www.constructaquote.com

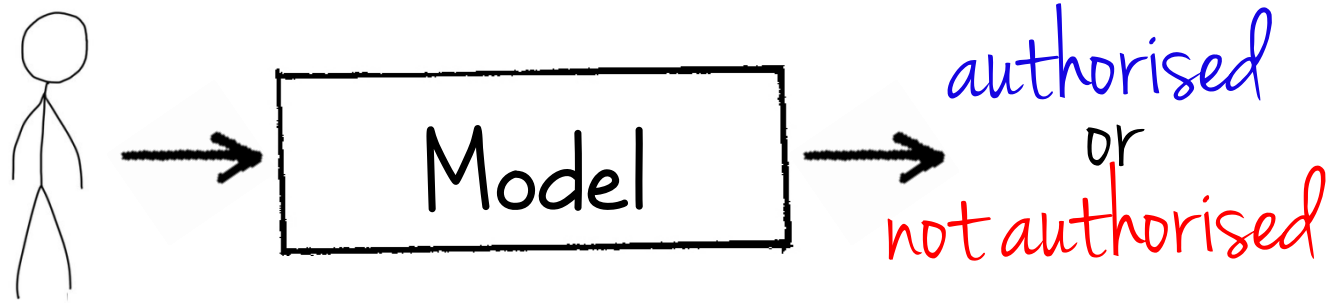
Fraud Detection

high true positive rate desirable



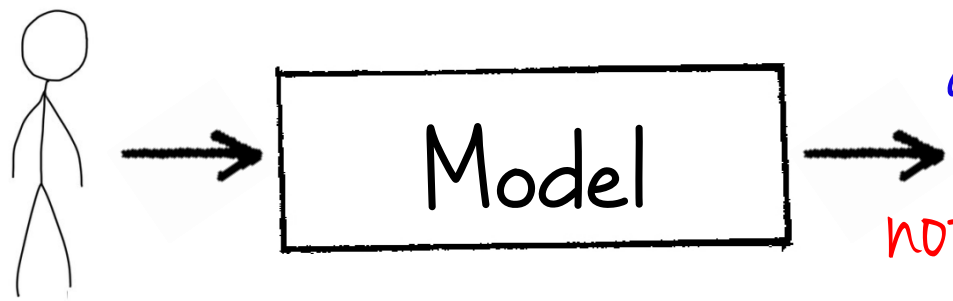
http://www.constructaquote.com

Biometric Screening



Biometric Screening

low false positive rate desirable



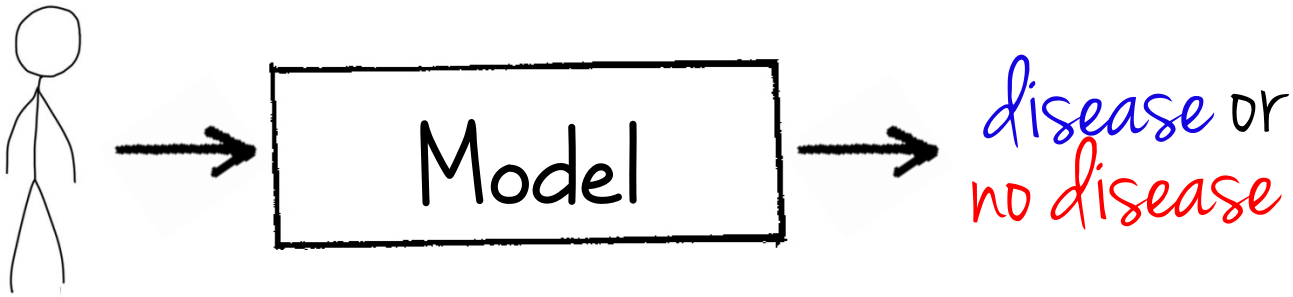
authorised
or
not authorised

Spam Filtering

low false positive rate desirable

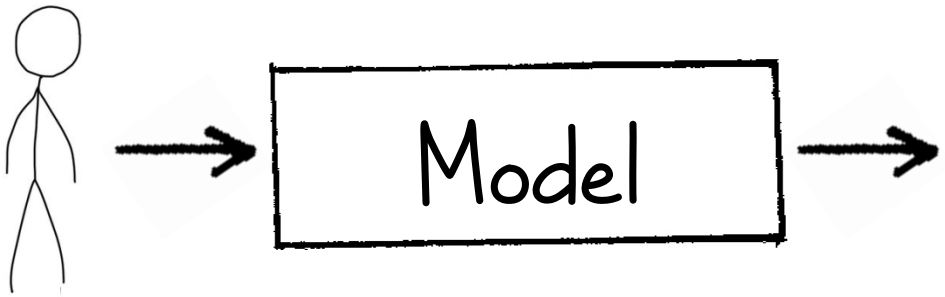
allspammedup.com

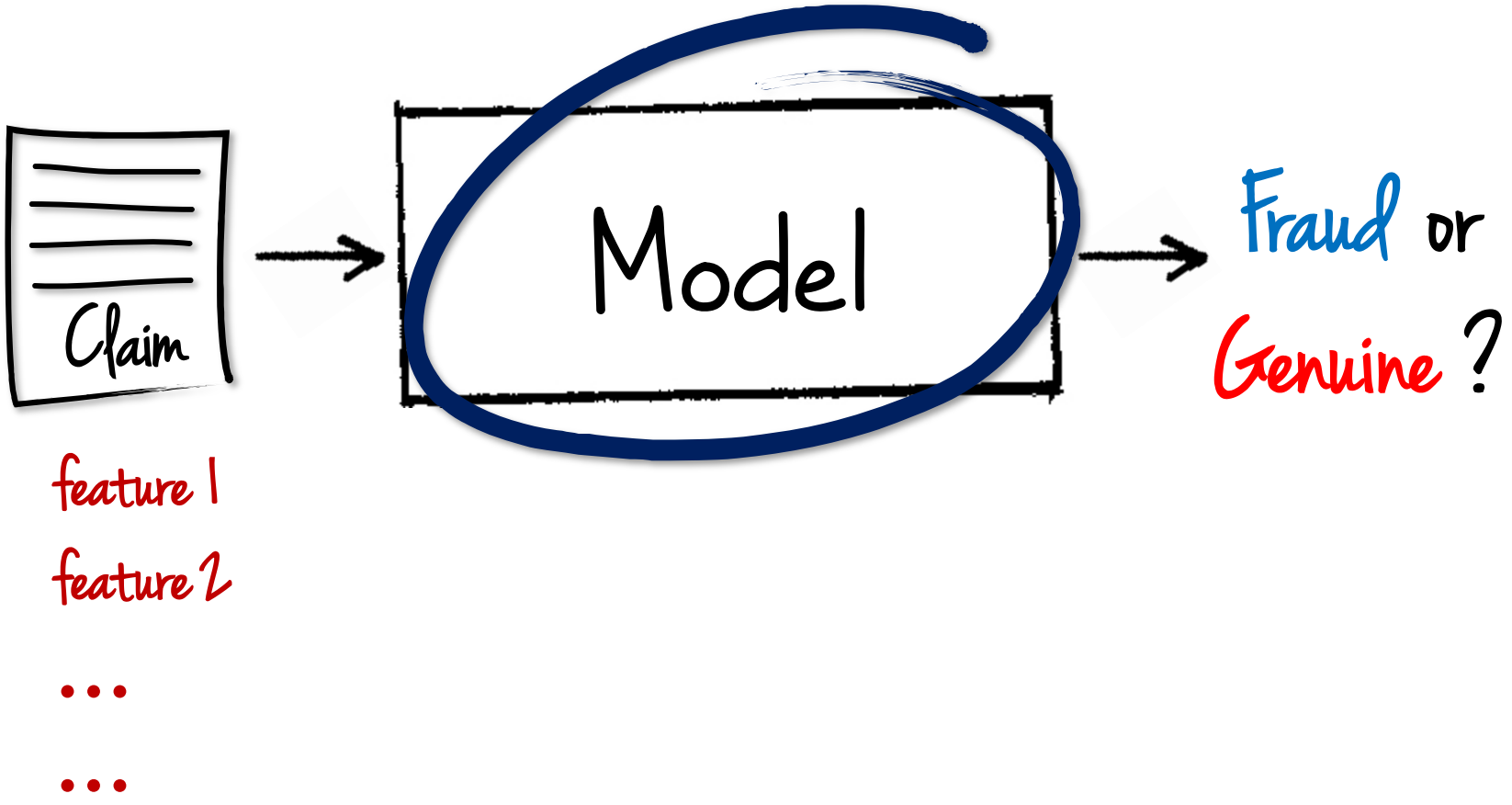
Medical Diagnosis

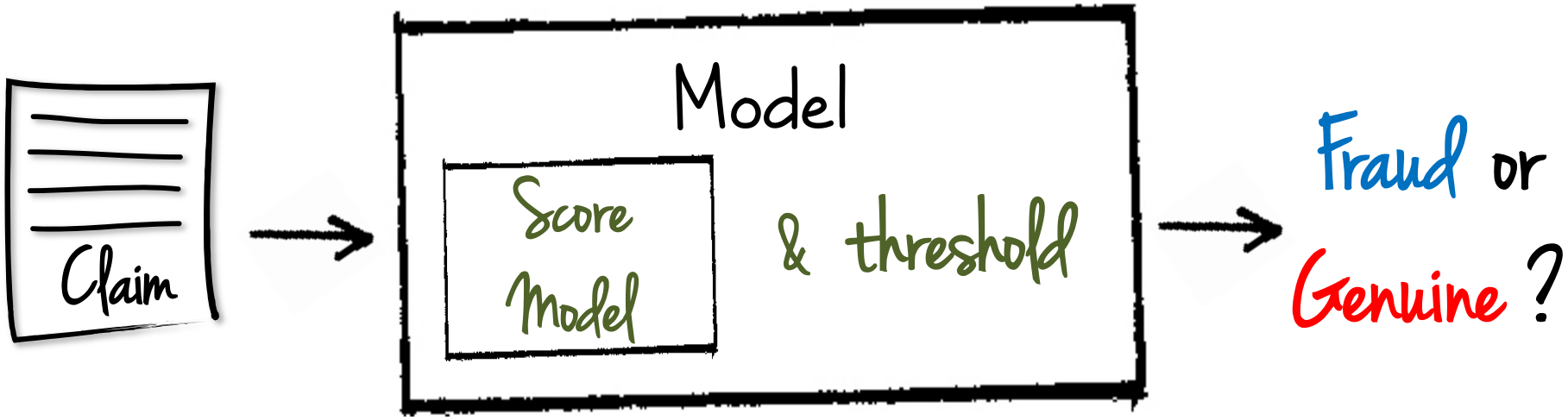


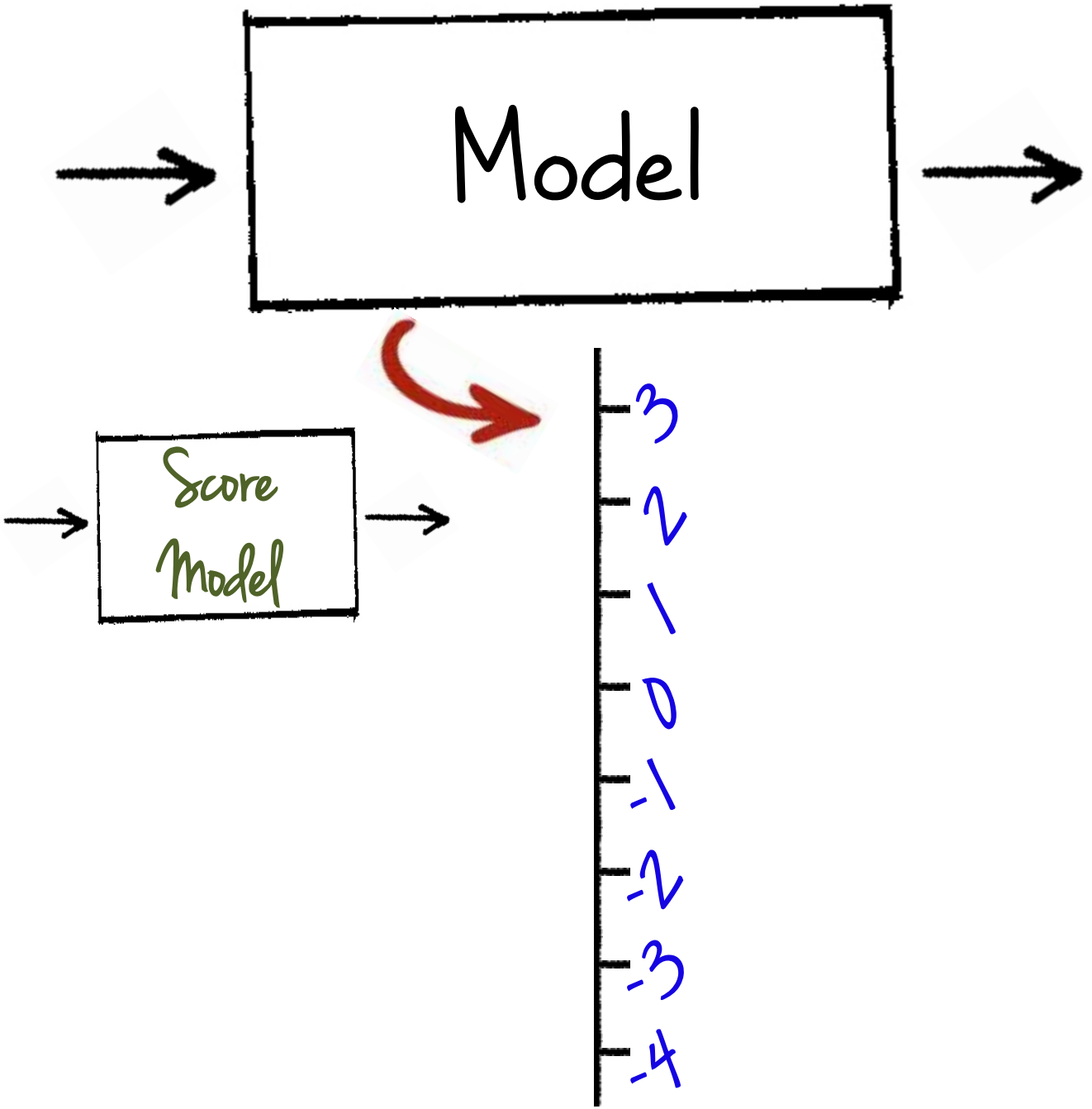
Medical Diagnosis

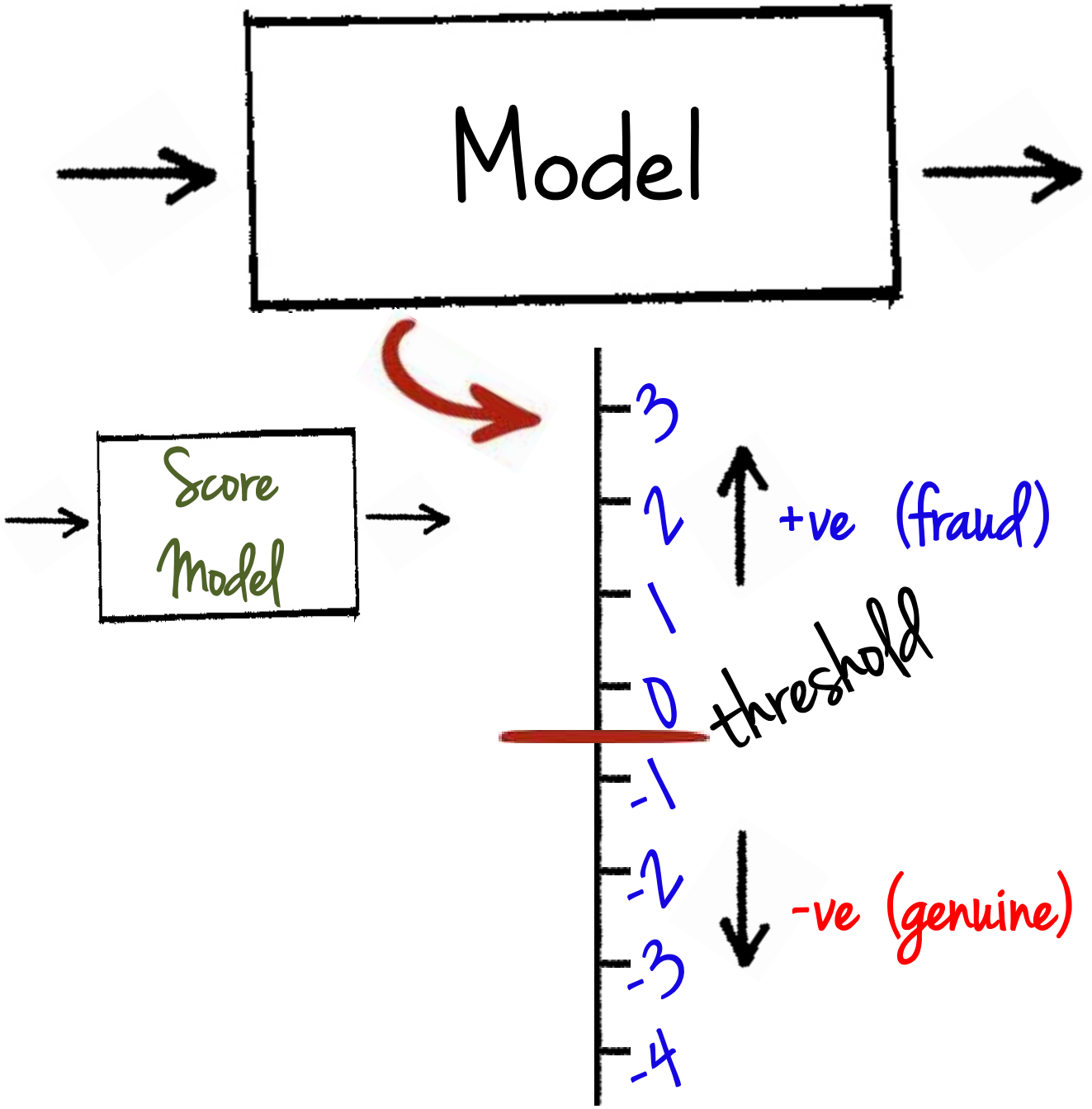
false positives within a range

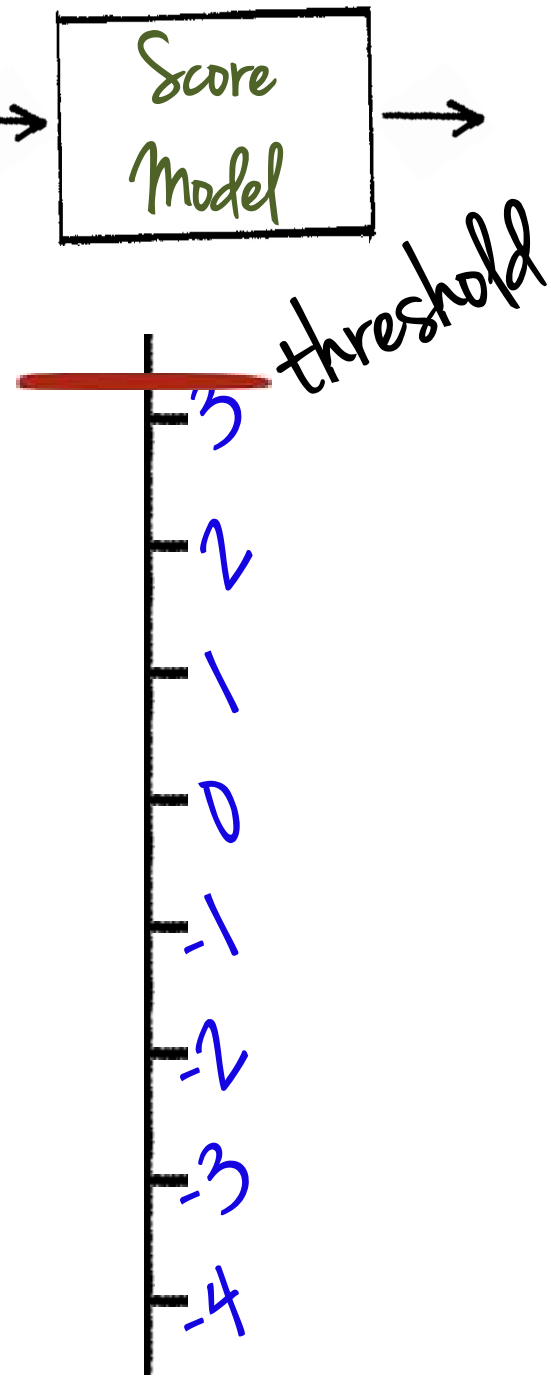


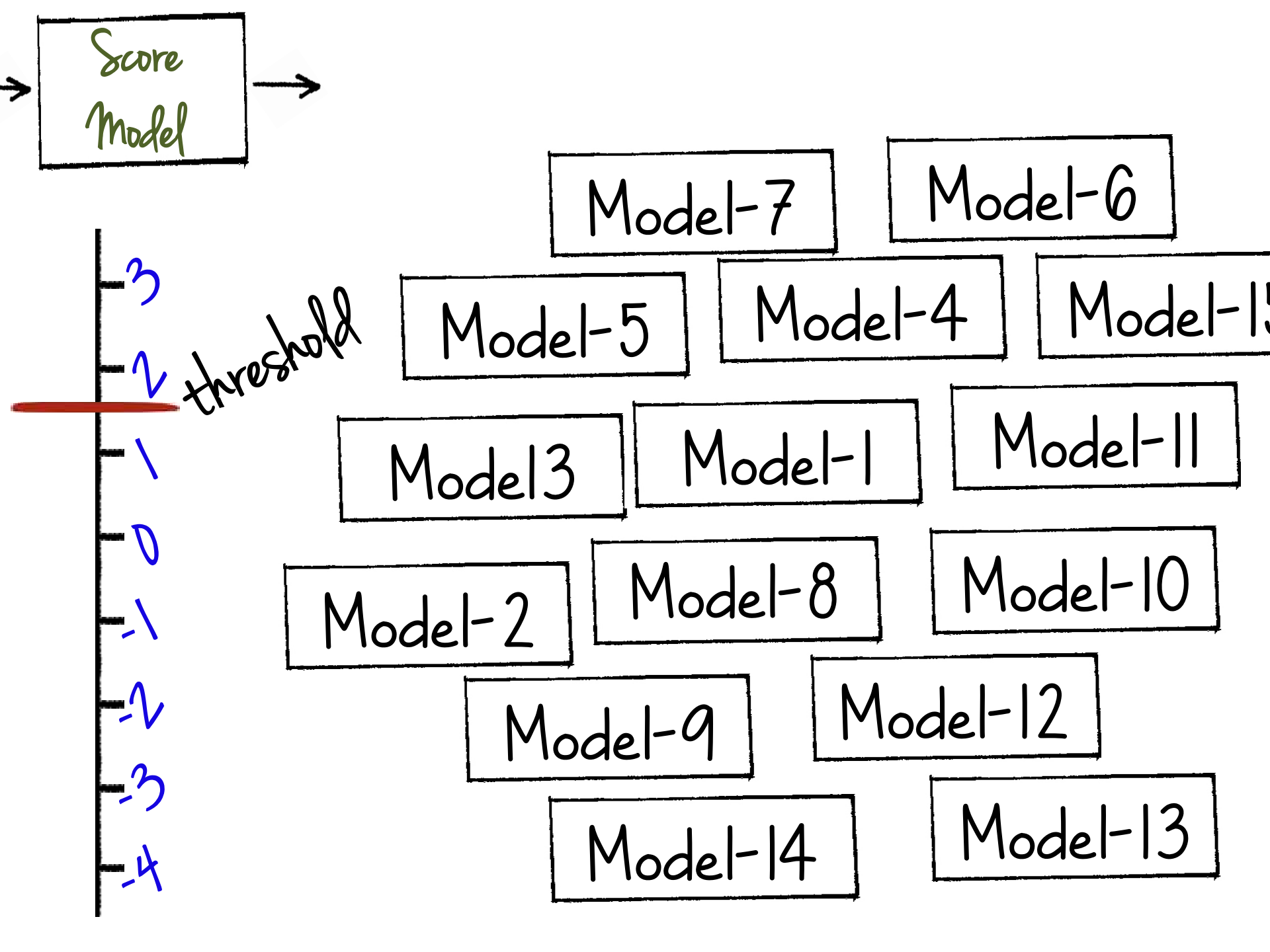

Claim
Model
Fraud or
Genuine ?
feature 1
feature 2
...
...

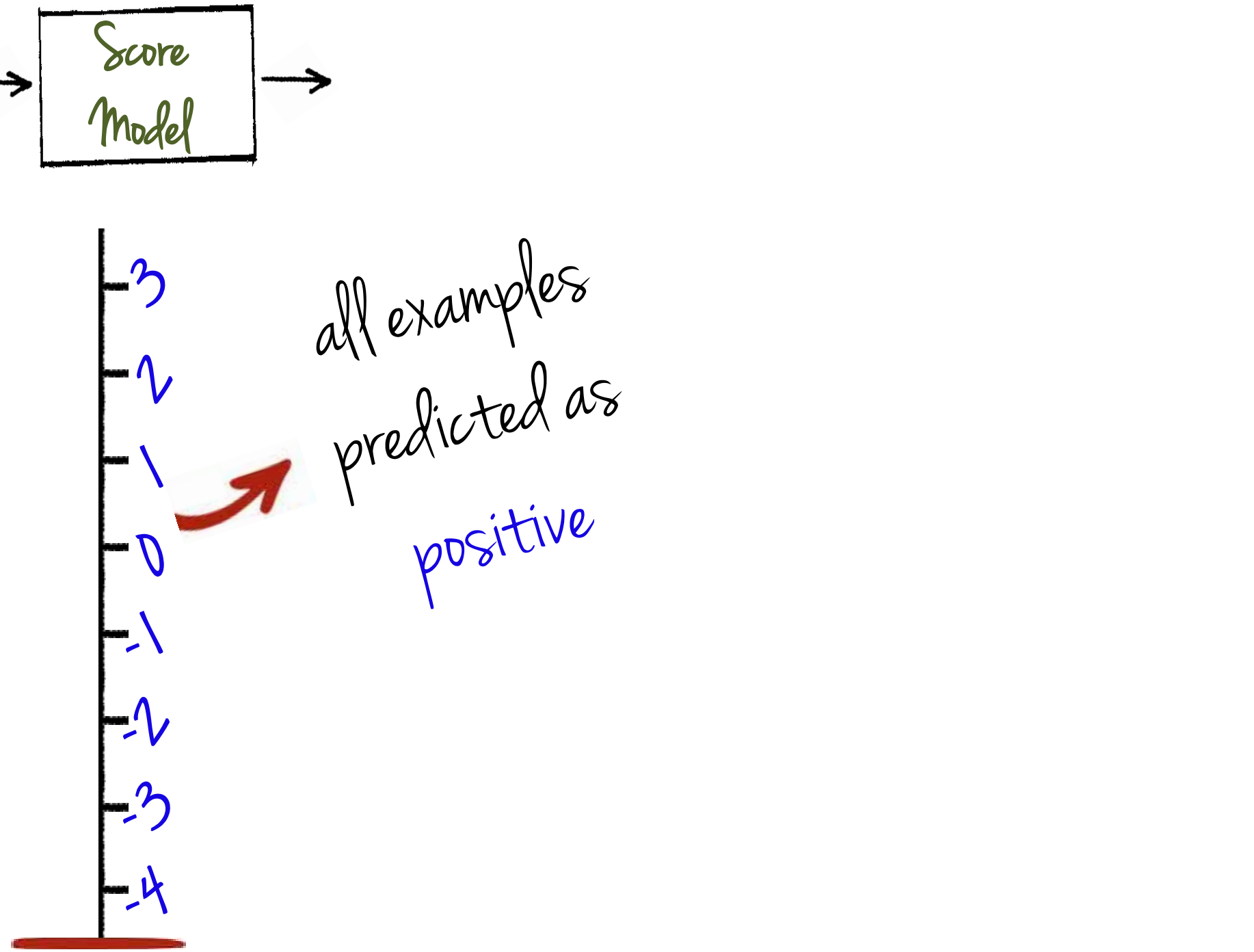

Claim
Model
Score
Model
& threshold
Fraud or
Genuine ?

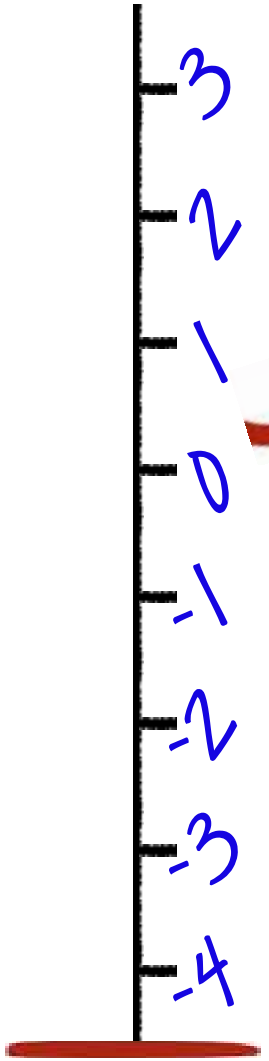

Model
Score Model
3
2
1
0
-1
-2
-3
-4


Model
Score Model
3
2
1
0
-1
-2
-3
-4
+ve (fraud)
threshold
-ve (genuine)


Score
Model
threshold
3
2
1
0
-1
-2
-3
-4

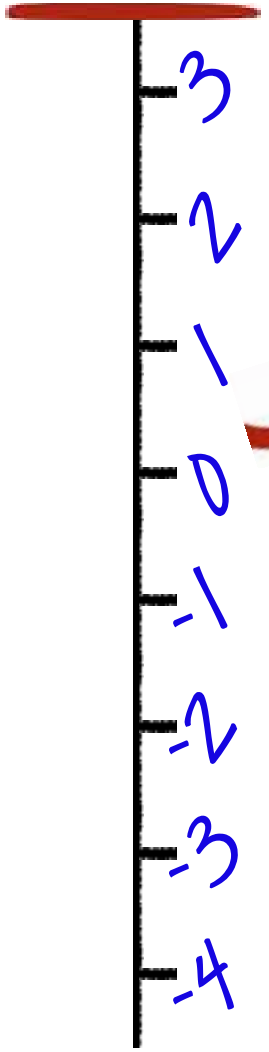

Score
Model
3
2
1
0
-1
-2
-3
-4
threshold
Model-7
Model-6
Model-5
Model-4
Model-15
Model3
Model-1
Model-11
Model-2
Model-8
Model-10
Model-9
Model-12
Model-14
Model-13


Score
Model
3
2
1
0
-1
-2
-3
-4
all examples
predicted as
positive



all examples predicted as positive

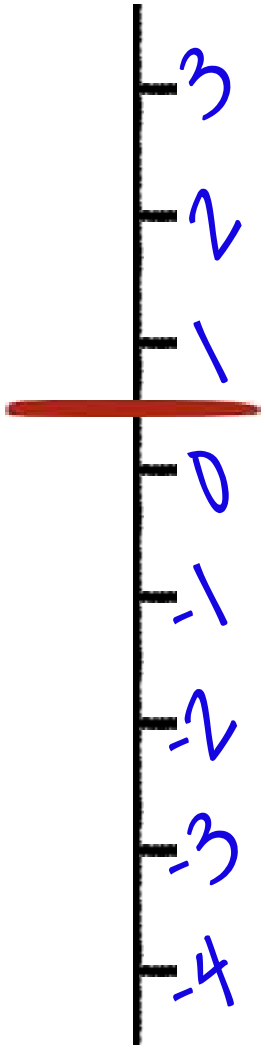
	Positive	Negative
Positive (Fraud)	Maximum True Positive Rate	
Negative (Genuine)	Maximum False Positive Rate	



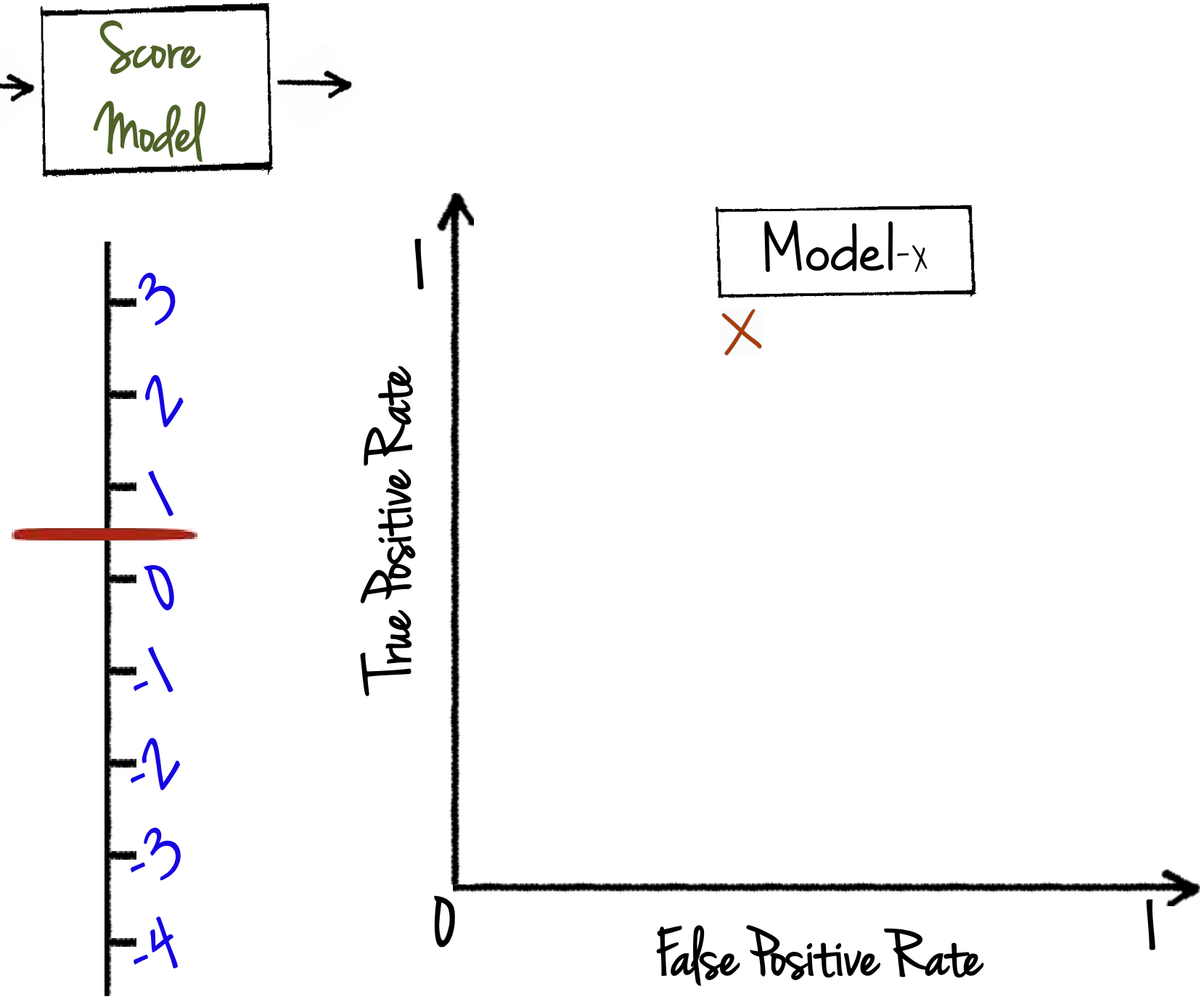
all examples predicted as

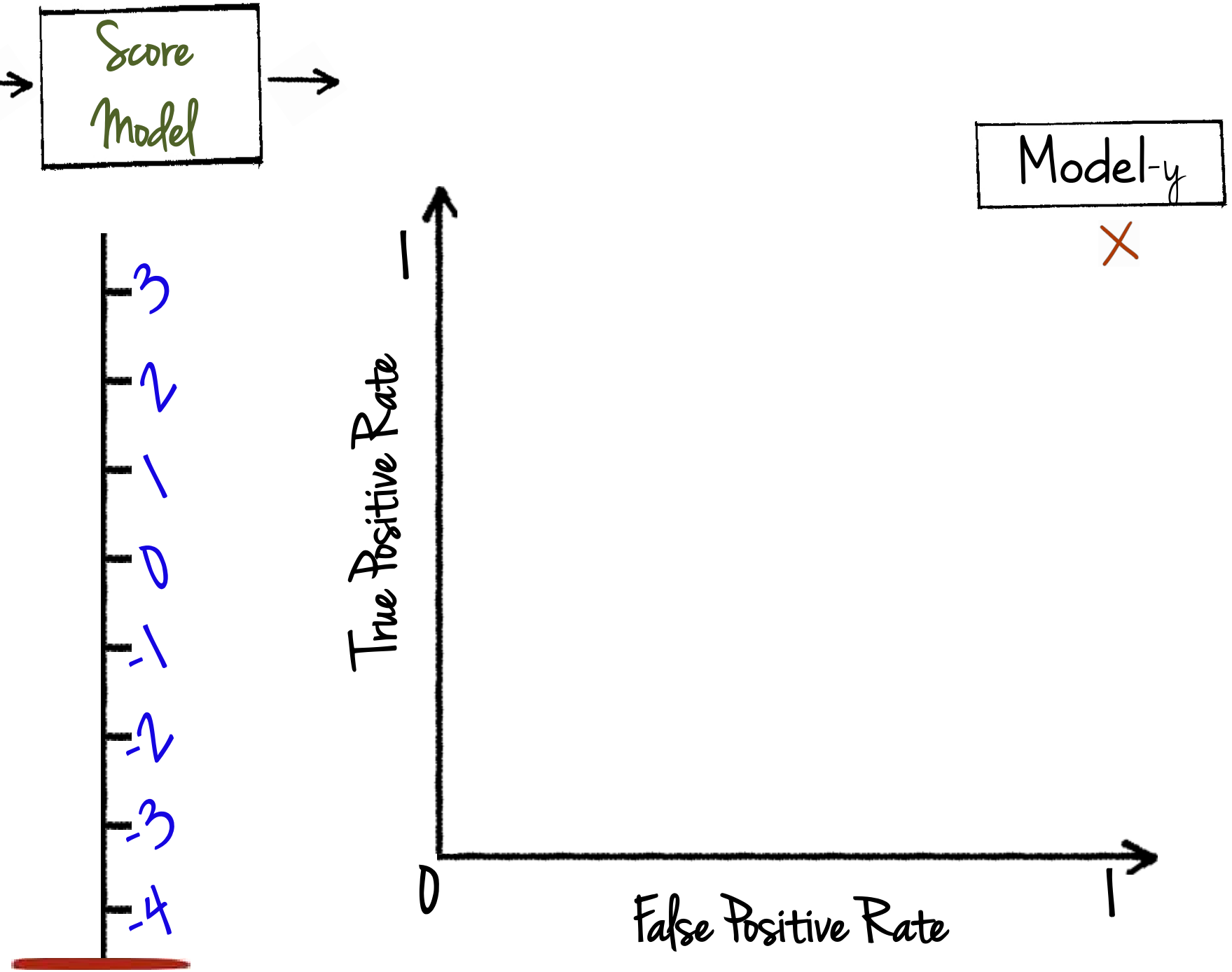
negative

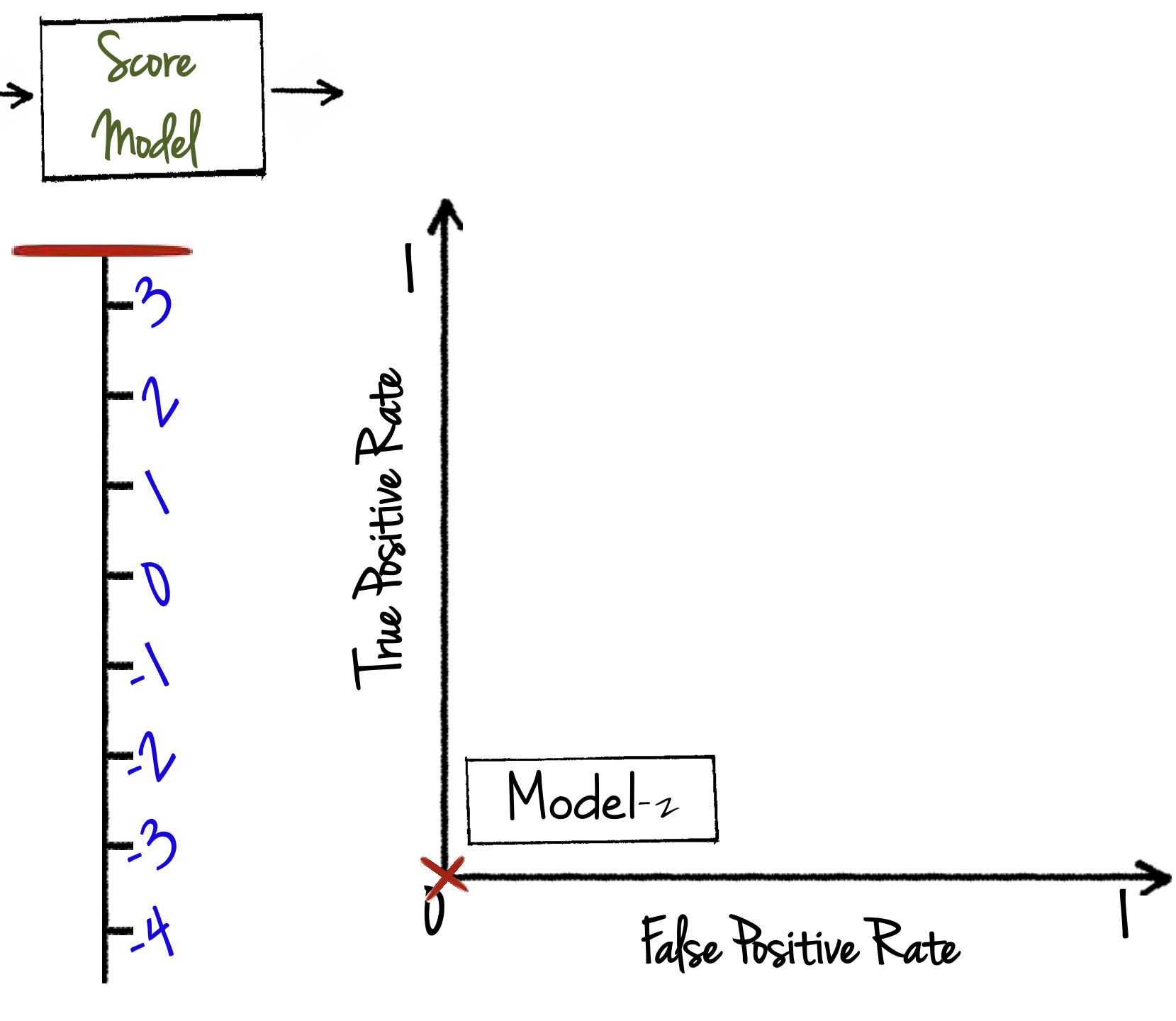
	Positive	Negative
Positive (Fraud)	Zero True Positive Rate	
Negative (Genuine)	Zero False Positive Rate	

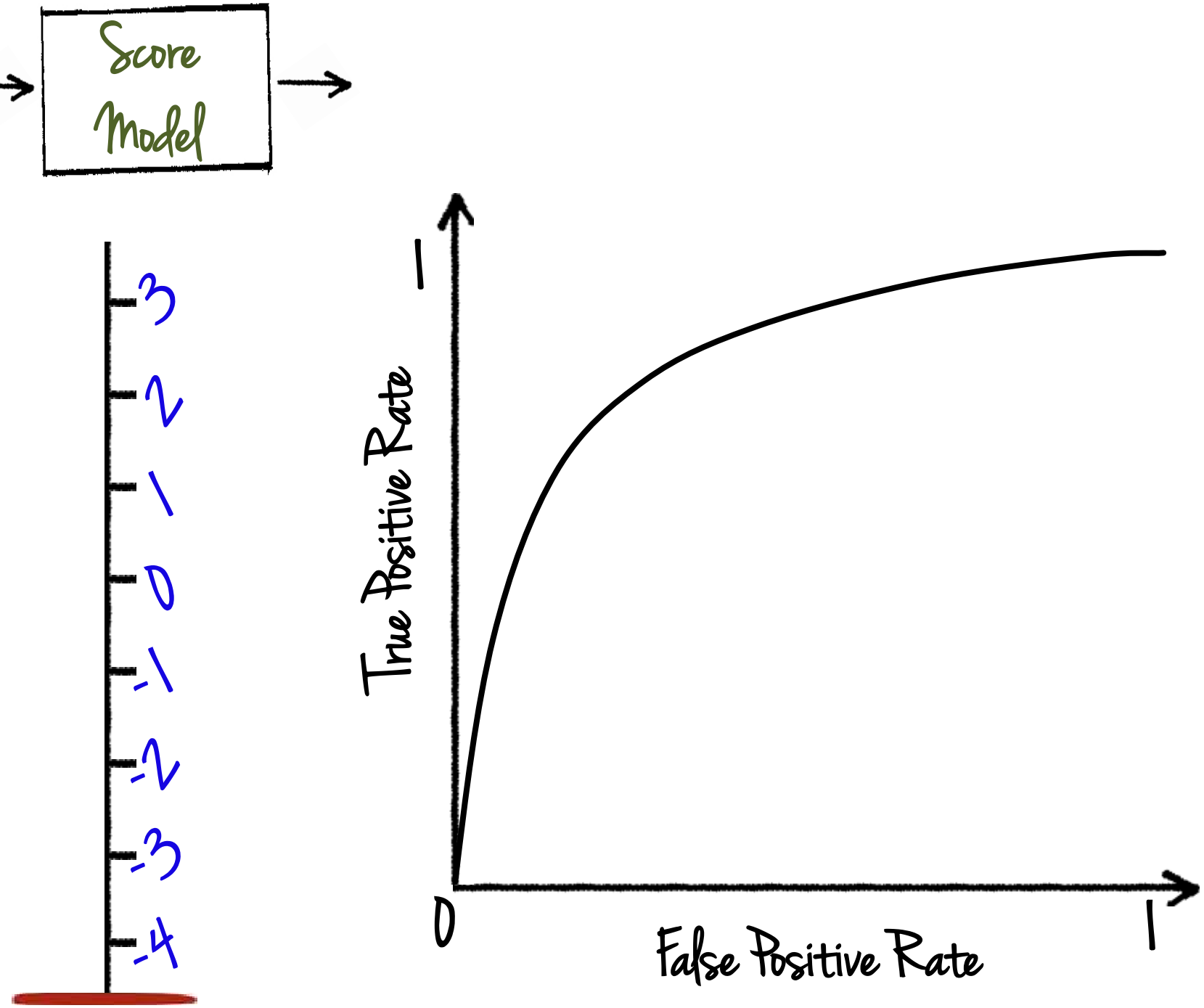


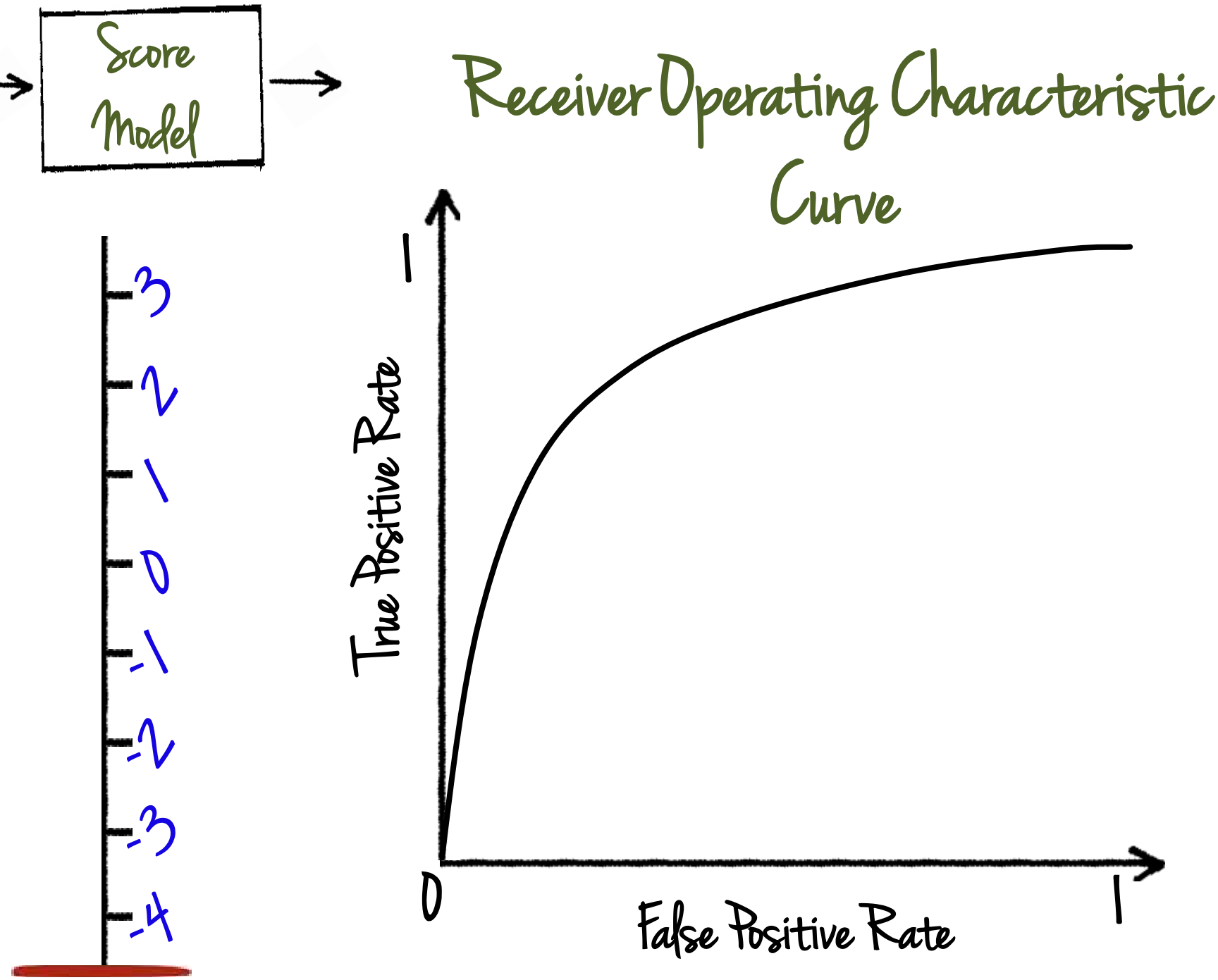
	Positive	Negative
Positive (Fraud)	?	
Negative (Genuine)	?	

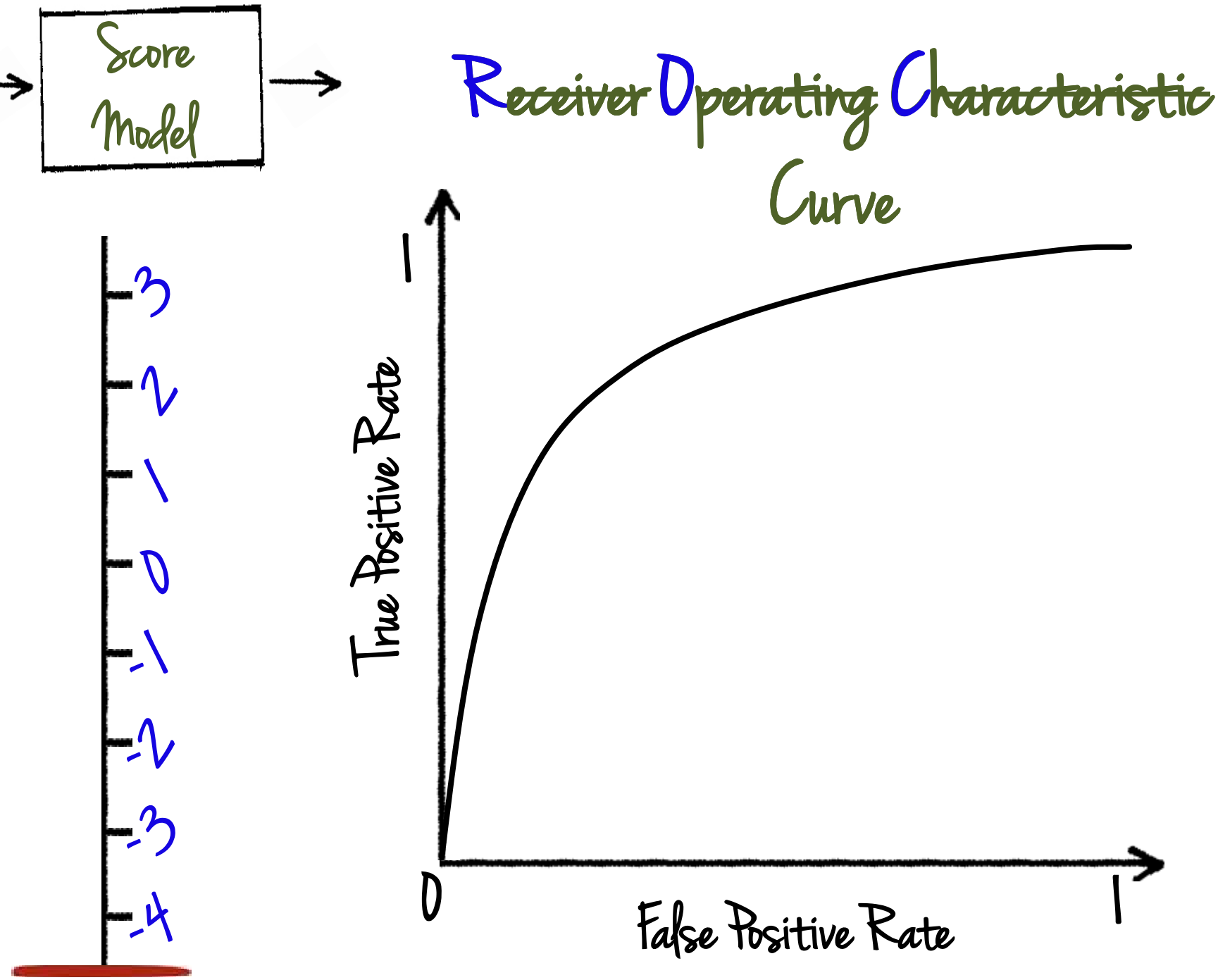

Score
Model
3
2
1
0
-1
-2
-3
-4
1
True Positive Rate
Model-x
0
1
False Positive Rate

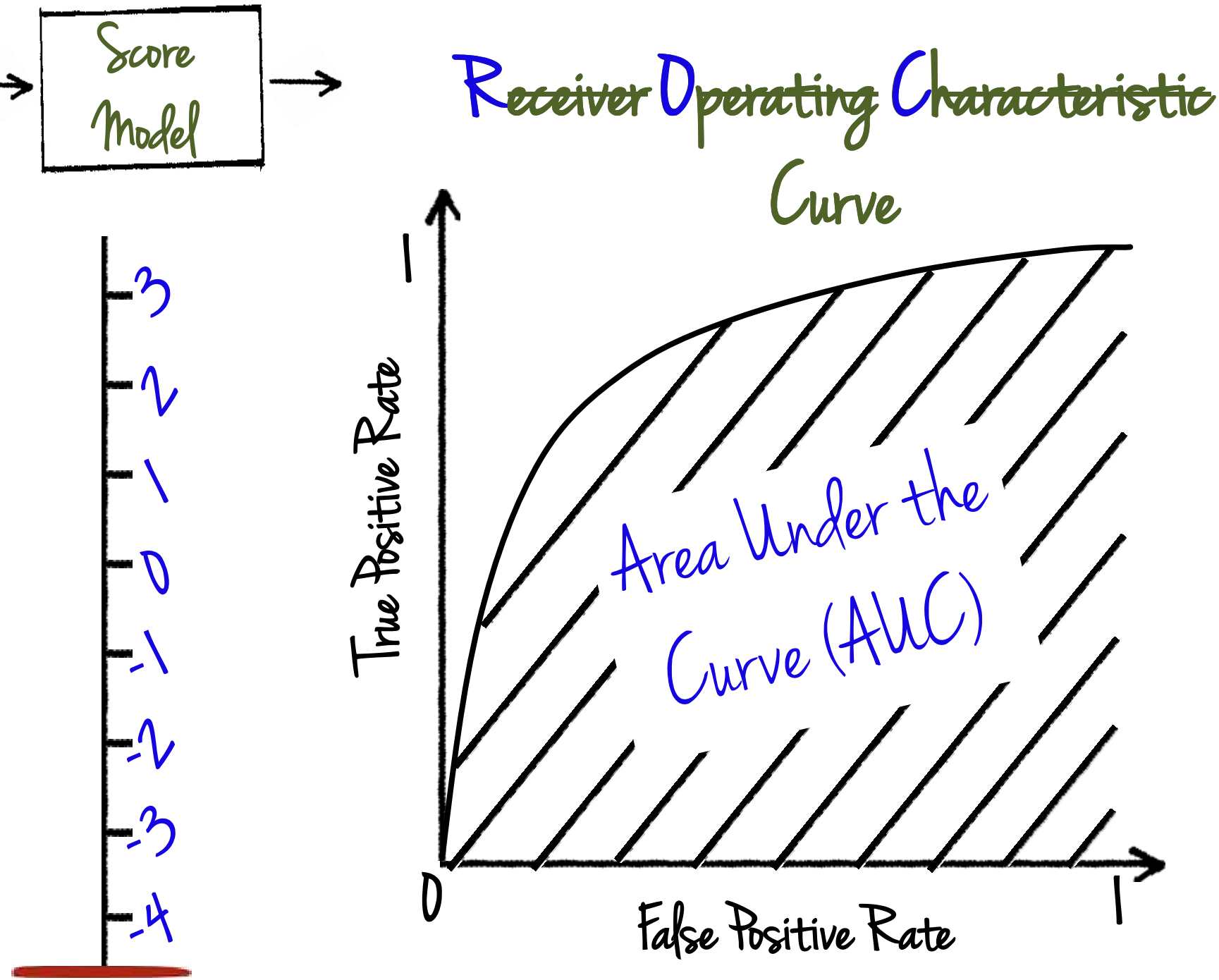

Score
Model
Model-y
3
2
1
0
-1
-2
-3
-4
1
True Positive Rate
0
1
False Positive Rate

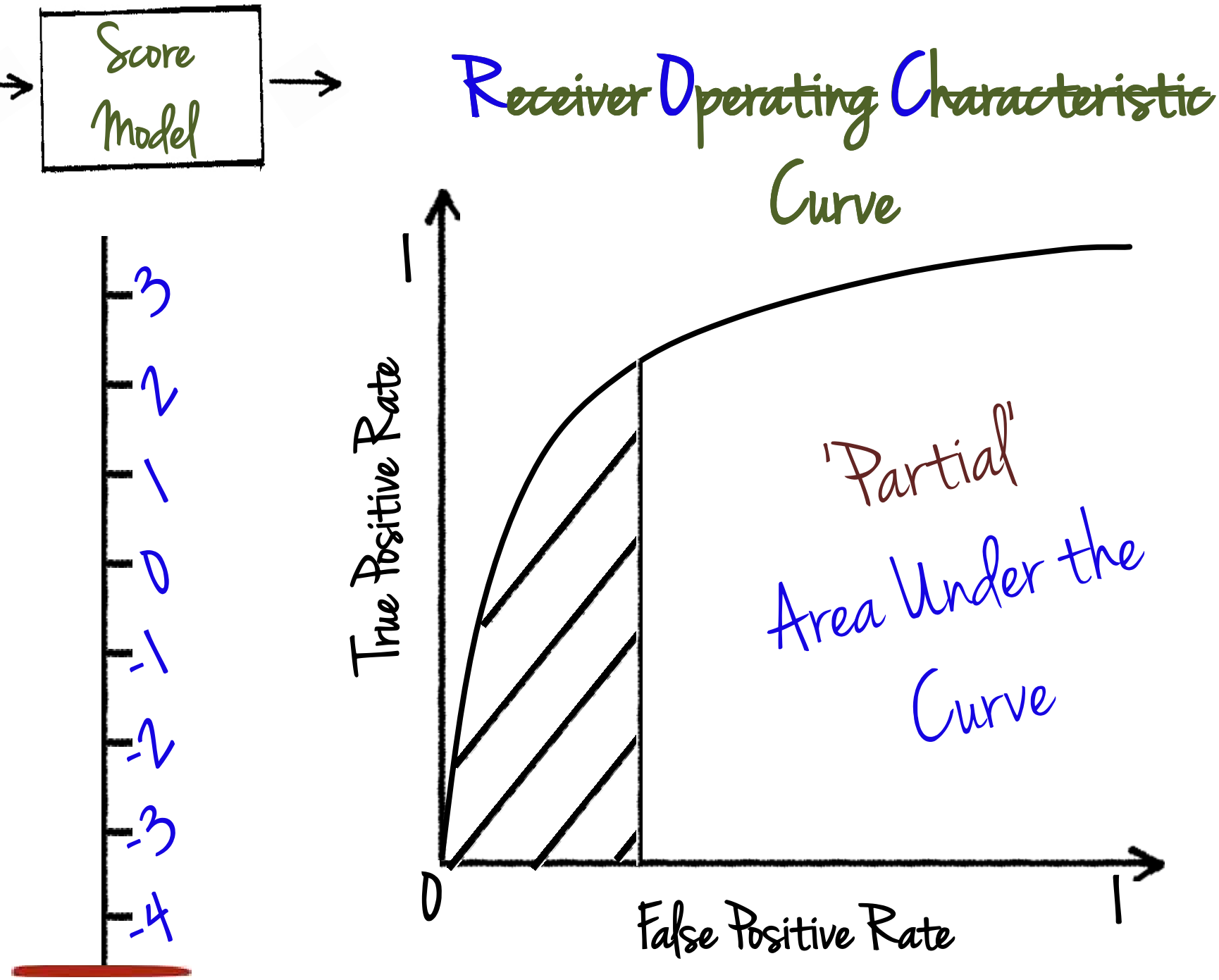

Score
Model
3
2
1
0
-1
-2
-3
-4
1
True Positive Rate
Model-2
0
1
False Positive Rate

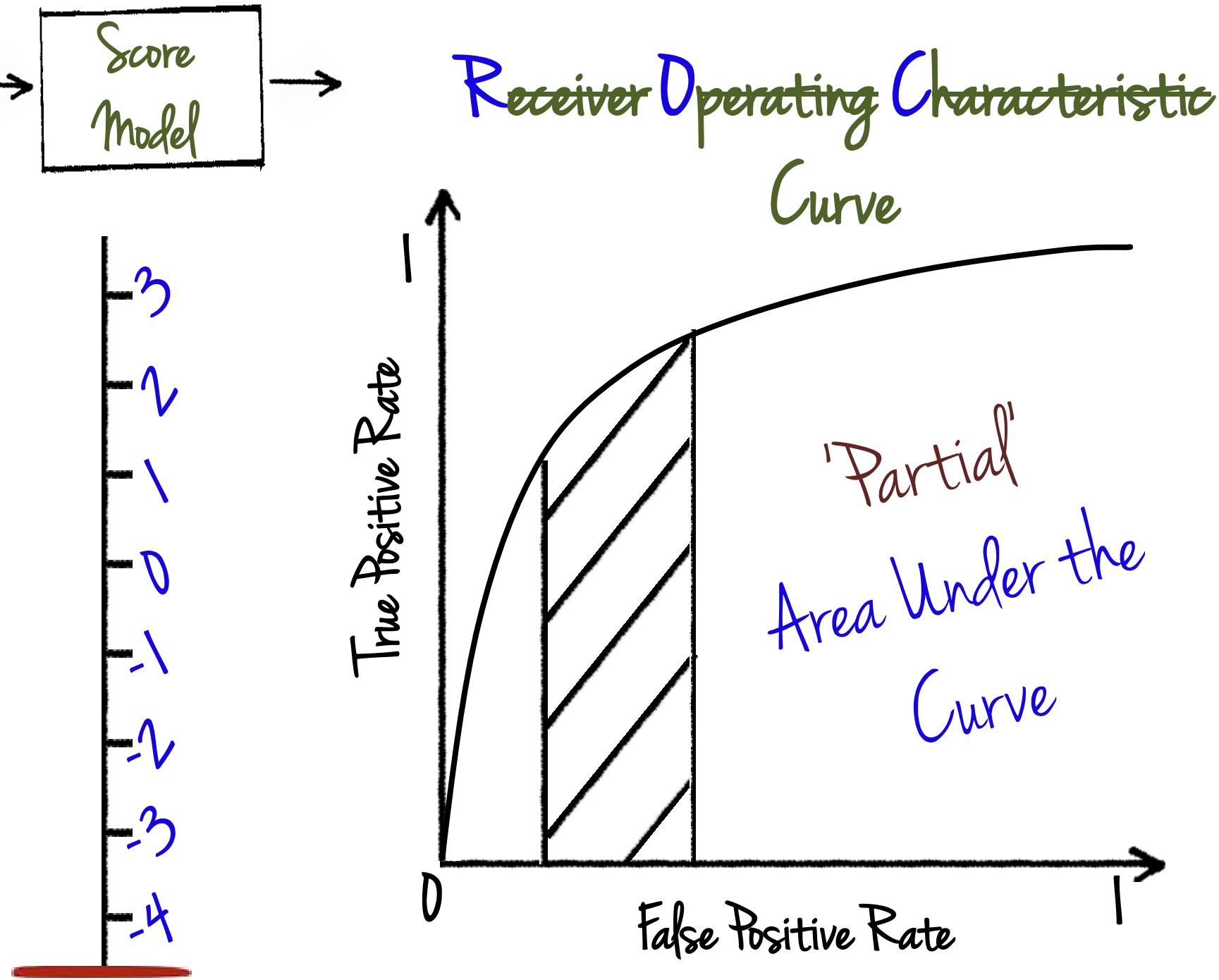

Score
Model
3
2
1
0
-1
-2
-3
-4
True Positive Rate
1
0
1
False Positive Rate


Score
Model
Receiver Operating Characteristic
Curve
3
2
1
0
-1
-2
-3
-4
True Positive Rate
1
0
1
False Positive Rate

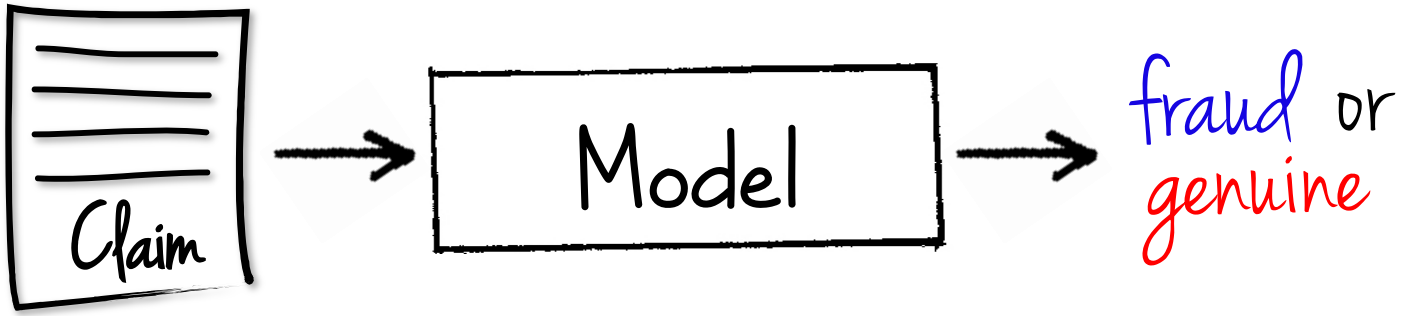

Score
Model
Receiver Operating Characteristic
Curve
3
2
1
0
-1
-2
-3
-4
1
True Positive Rate
0
1
False Positive Rate


Score
Model
3
2
1
0
-1
-2
-3
-4
Receiver Operating Characteristic
Curve
1
True Positive Rate
Area Under the
Curve (AUC)
0
1
False Positive Rate


Score
Model
Receiver Operating Characteristic
Curve
3
2
1
0
-1
-2
-3
-4
1
True Positive Rate
0
1
False Positive Rate
'Partial'
Area Under the
Curve

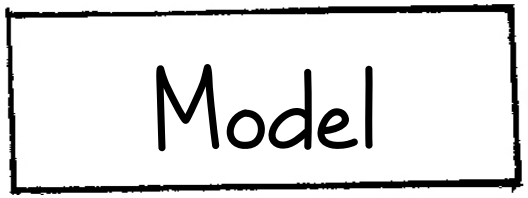

Score
Model
Receiver Operating Characteristic
Curve
3
2
1
0
-1
-2
-3
-4
1
True Positive Rate
0
False Positive Rate
1
'Partial'
Area Under the
Curve

Fraud Detection



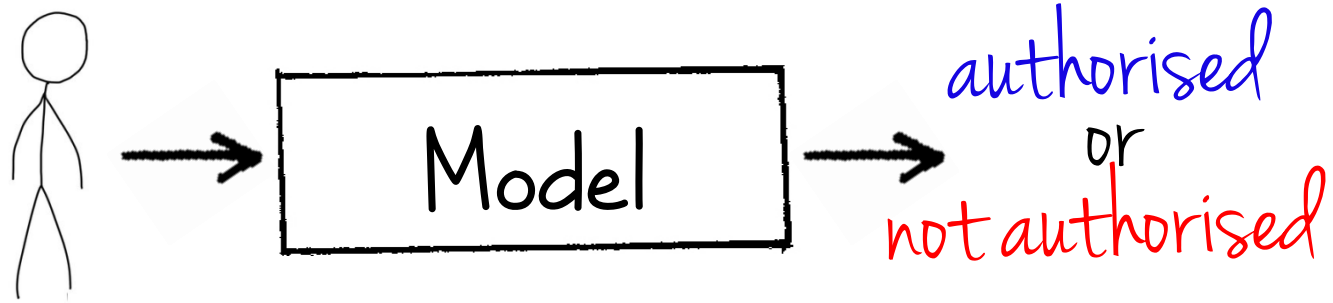
constructaquote.com

Fraud Detection



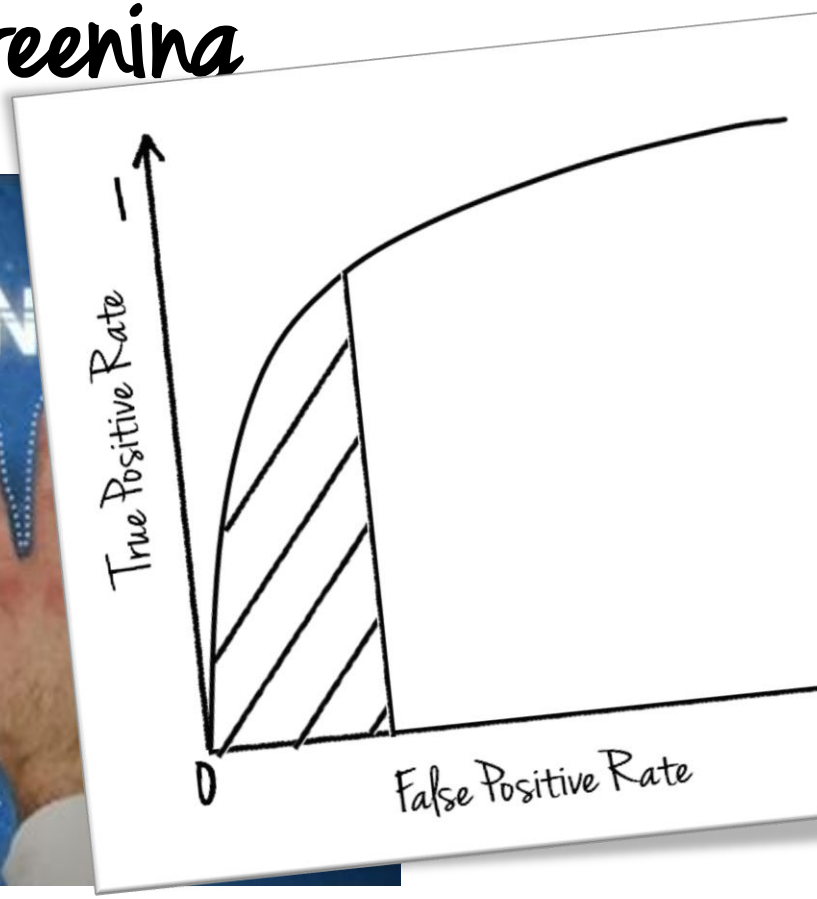
fraud or genuine

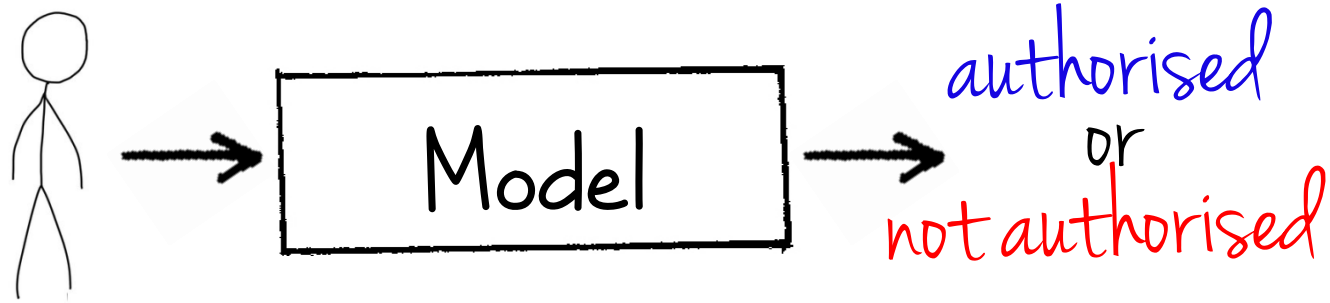
Biometric Screening



www.google.co.in/imghp

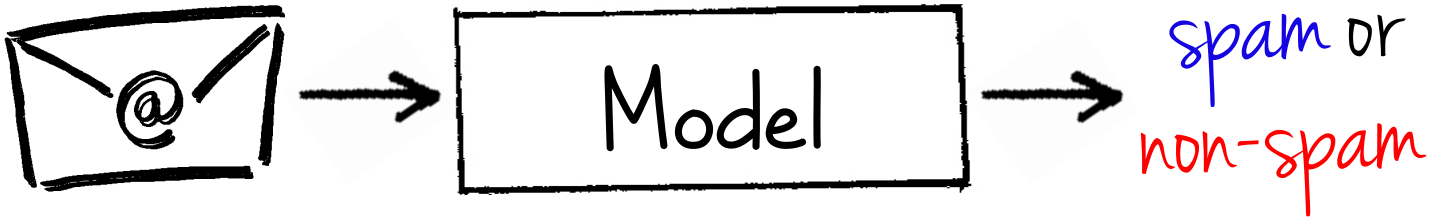
Biometric Screening





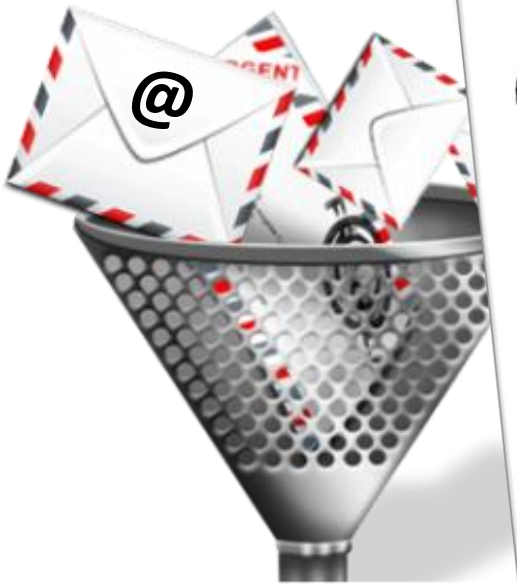
www.google.co.in/imghp

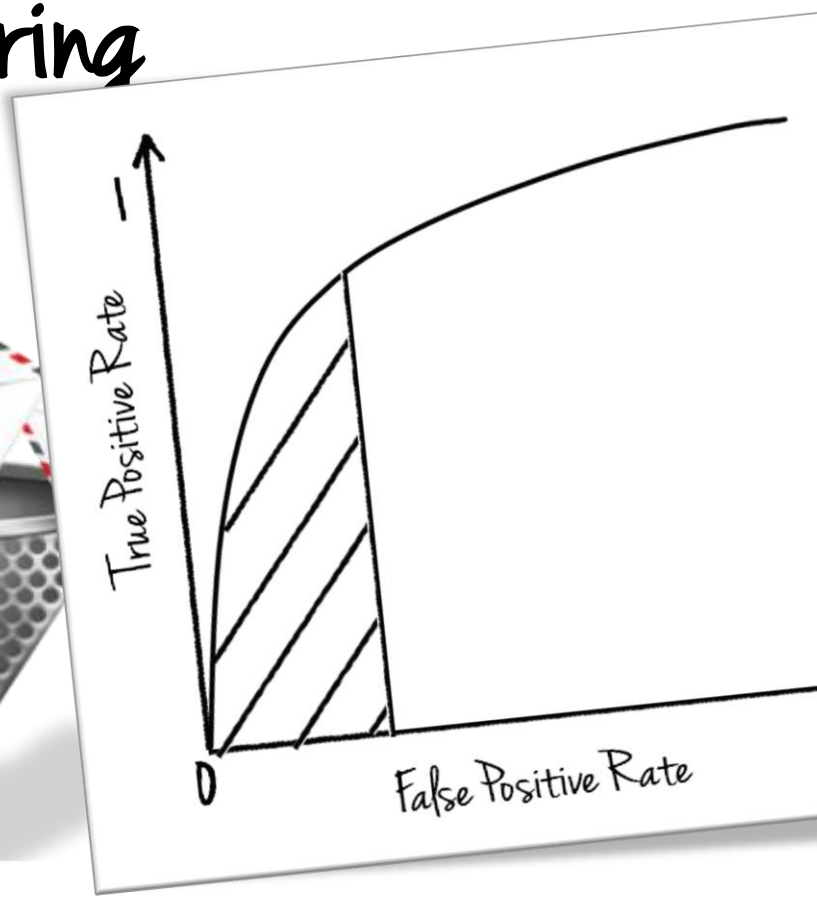
Spam Filtering

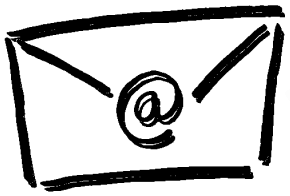


allspammedup.com

Spam Filtering





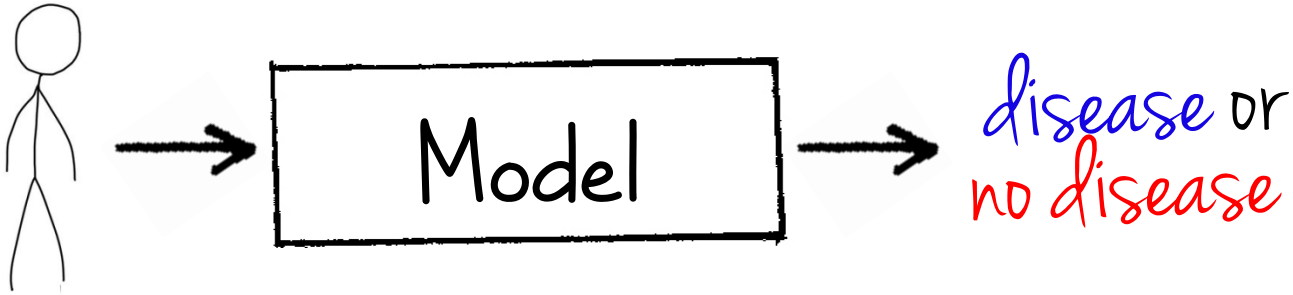


→ Model →

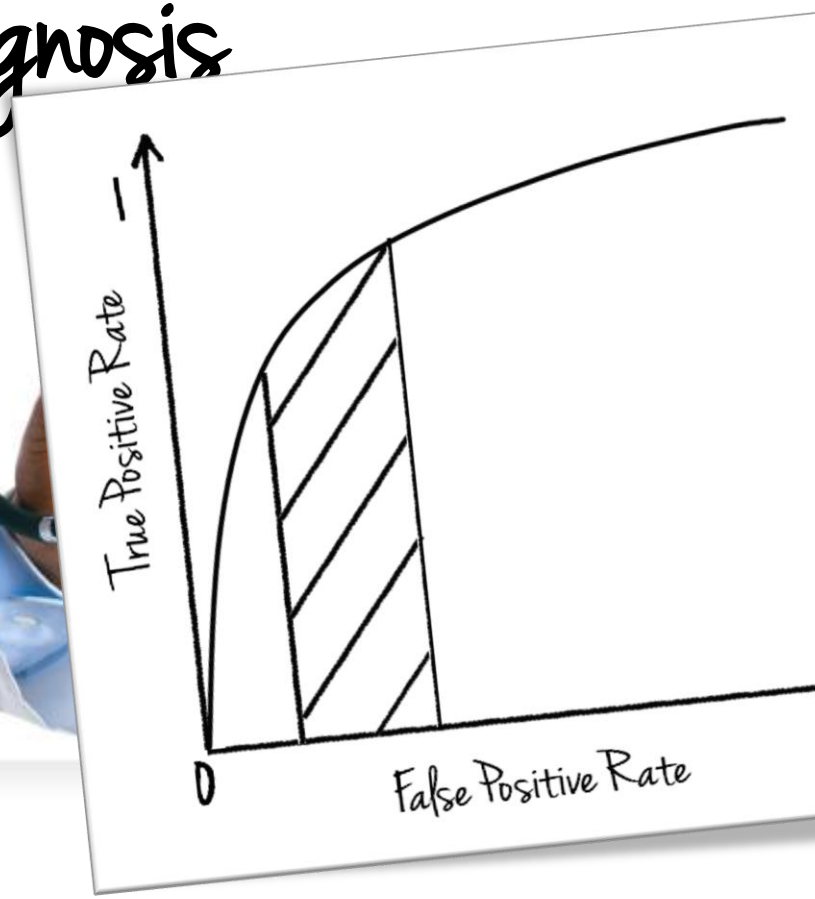
spam or
non-spam

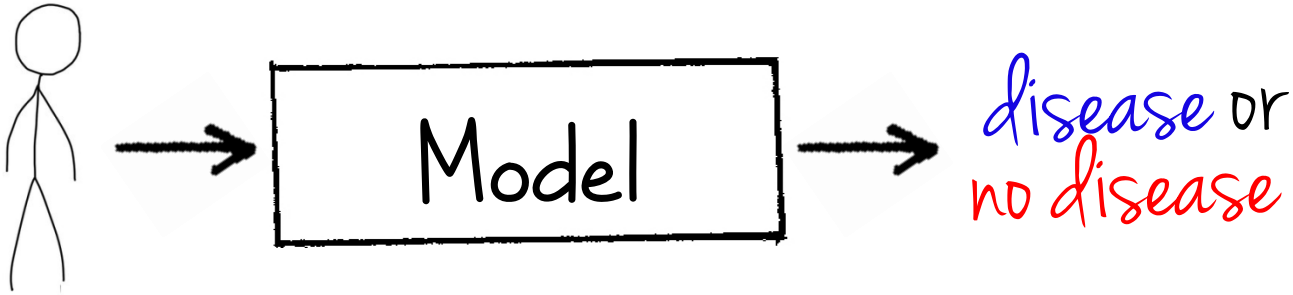
allspammedup.com

Medical Diagnosis

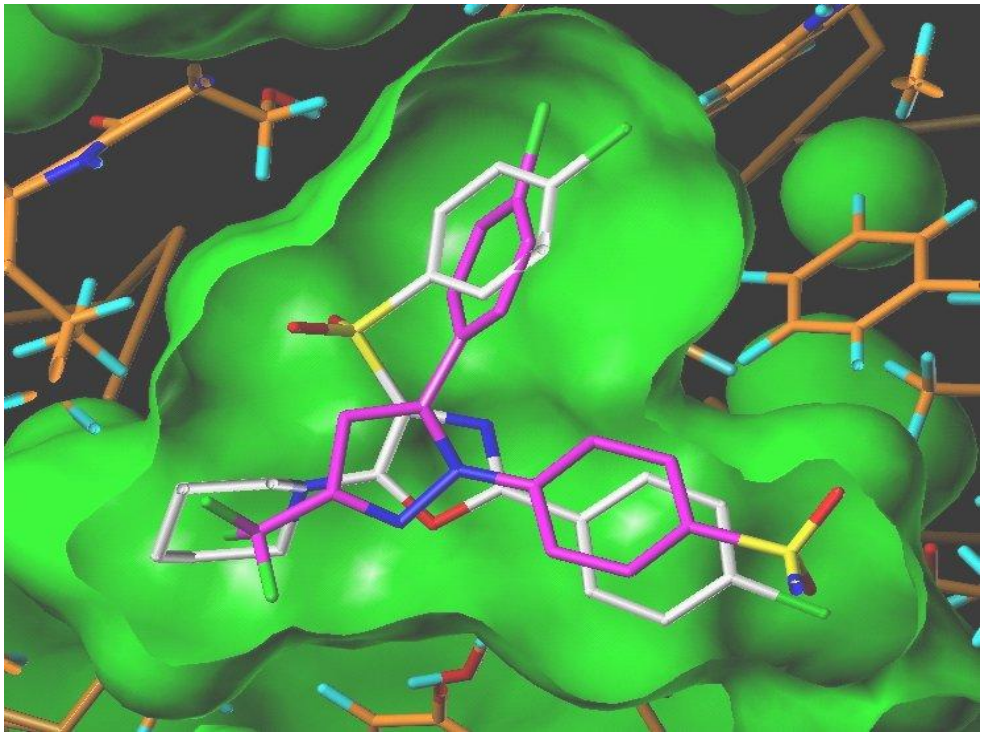


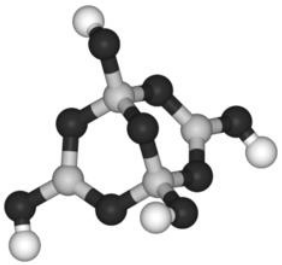
Medical Diagnosis





Drug Discovery





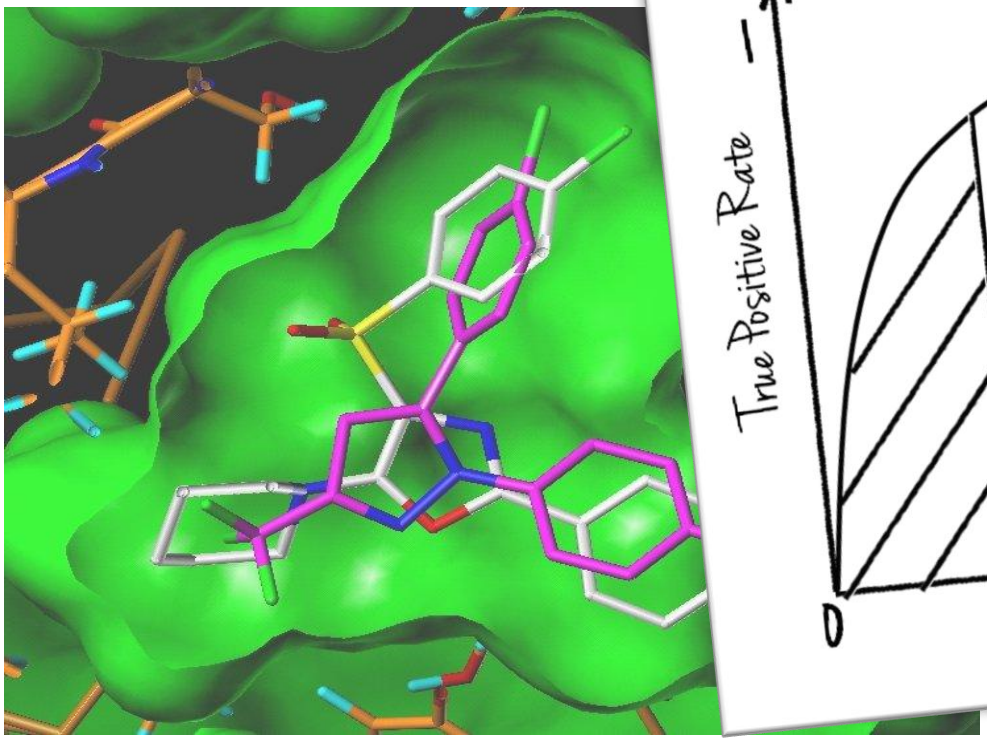
Model

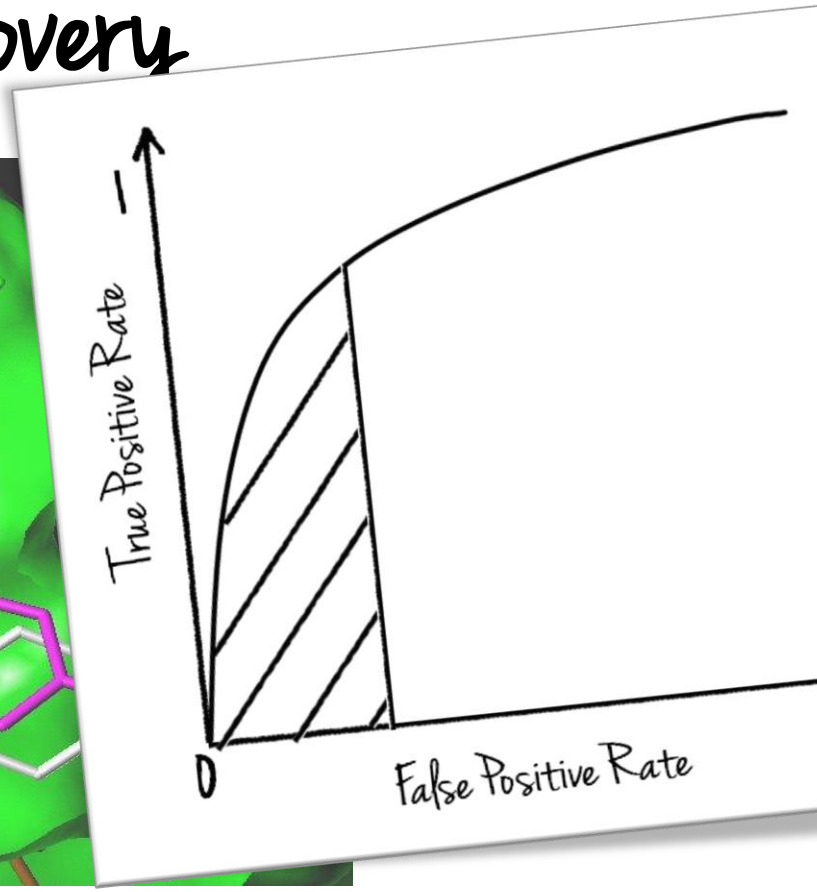
active or

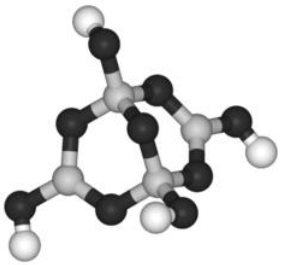
inactive

www.google.co.in/imghp

Drug Discovery





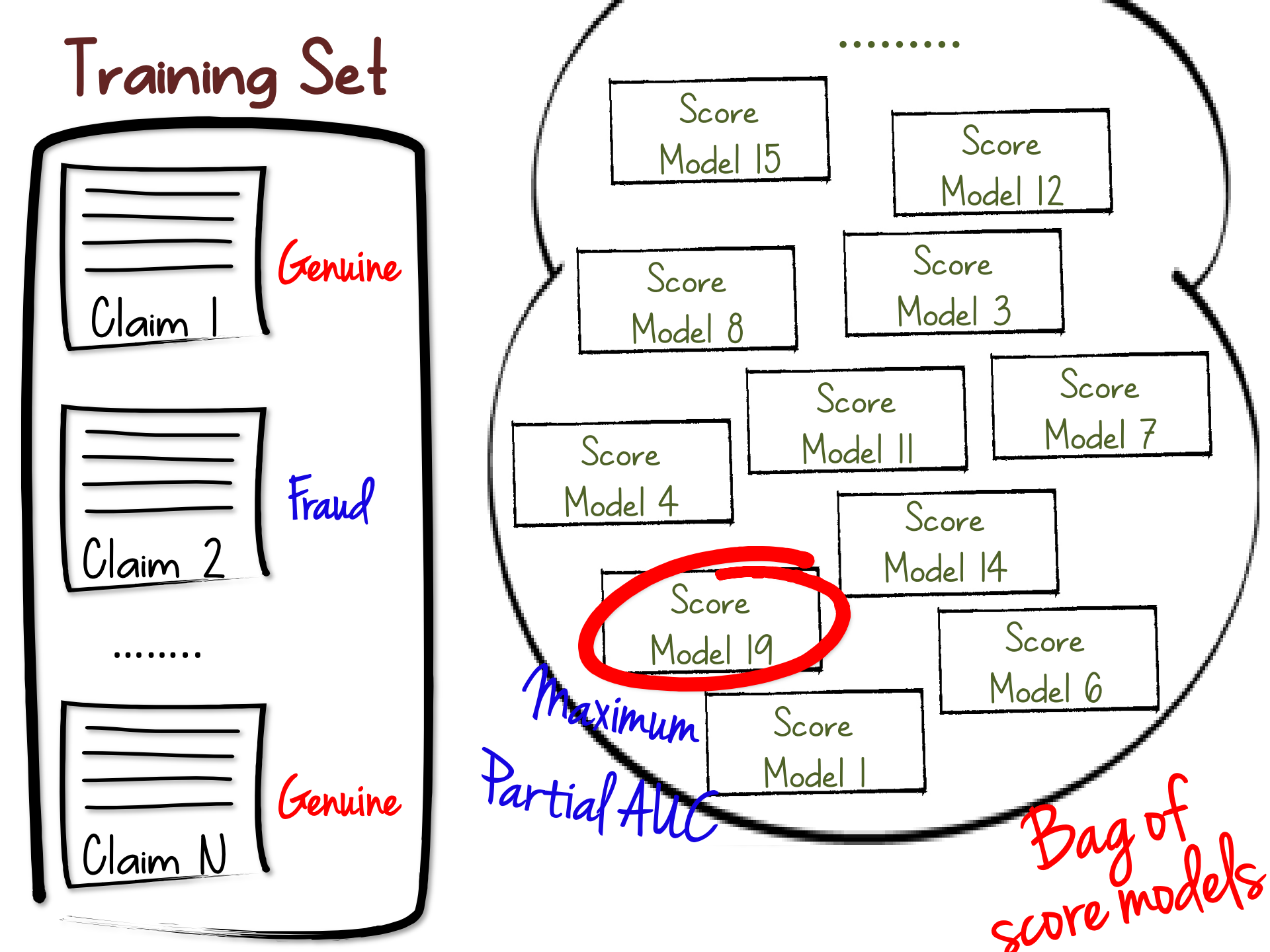


→ Model →

active or

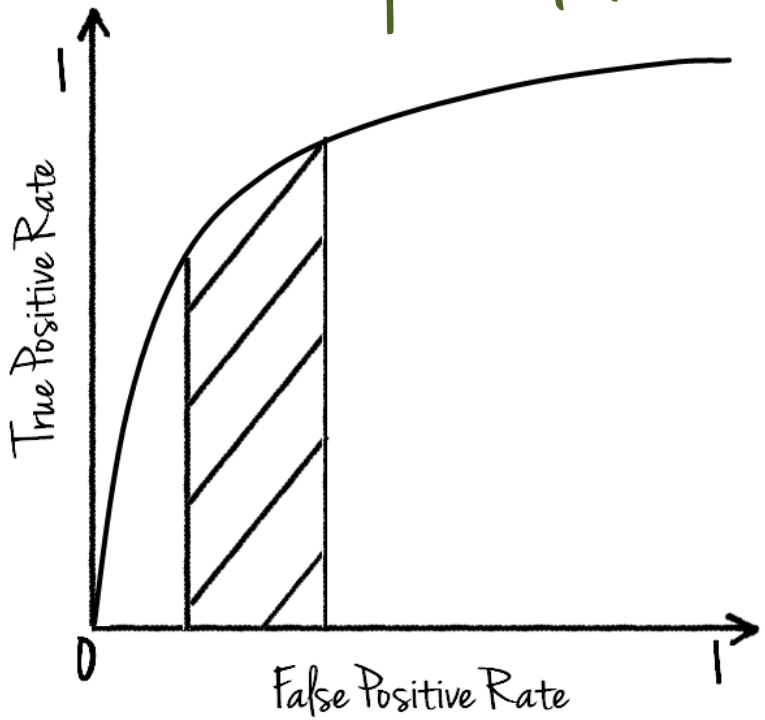
inactive

www.google.co.in/imghp


Training Set
Claim 1
Genuine
Claim 2
Fraud
.....
Claim N
Genuine
.....
Score Model 15
Score Model 12
Score Model 8
Score Model 3
Score Model 11
Score Model 7
Score Model 4
Score Model 14
Score Model 19
Score Model 6
Score Model 1
Maximum Partial AUC
Bag of score models

Optimization Problem

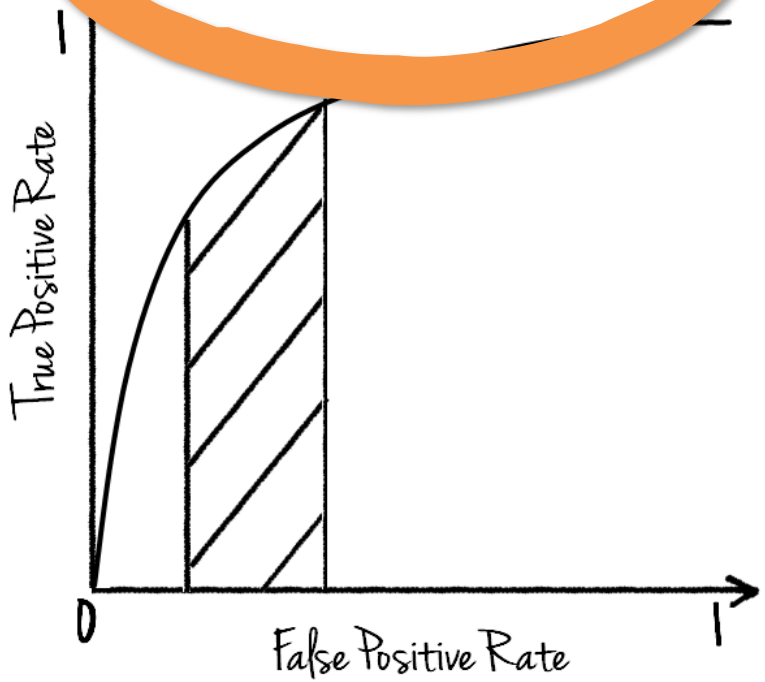
find score model with maximum 'partial AUC'



Optimization Problem

find score model with maximum 'partial AUC'

Intractable / Hard



Proxy Optimization Problem

(using structural SVMs)

Proxy Optimization Problem

(using structural SVMs)

maximize { 'partial AUC' }

Proxy Optimization Problem

(using structural SVMs)

~~maximize {'partial AUC'}~~

maximize {an approximation of 'partial AUC'}

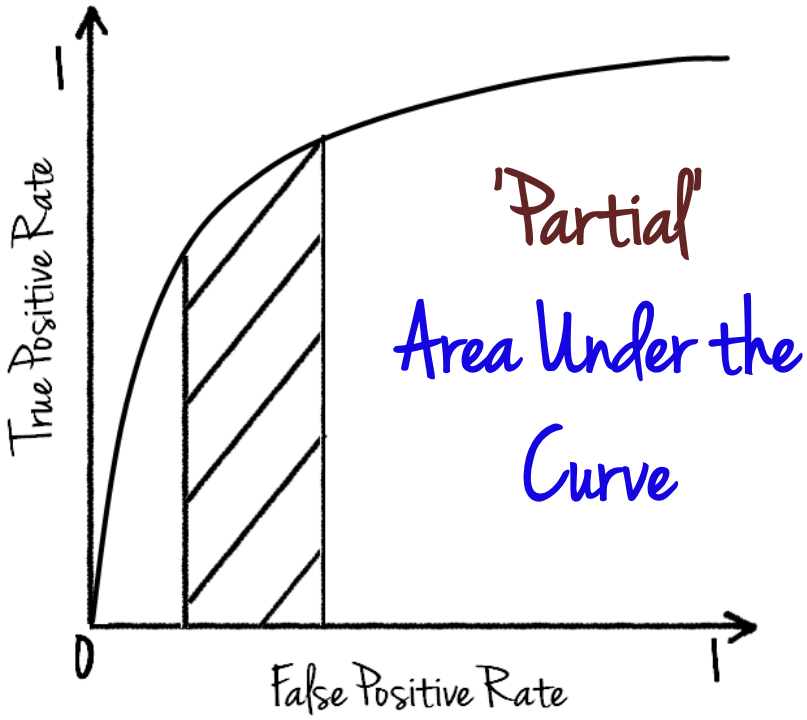
SVMpAUC-struct (ICML 2013)

SVMpAUC-tight (KDD 2013)

Narasimhan, H. and Agarwal, S. "*A structural SVM based approach for optimizing partial AUC*", ICML 2013.

Narasimhan, H. and Agarwal, S. *"SVM_pAUC^tight: A new support vector method for optimizing partial AUC based on a tight convex upper bound"*, KDD 2013.

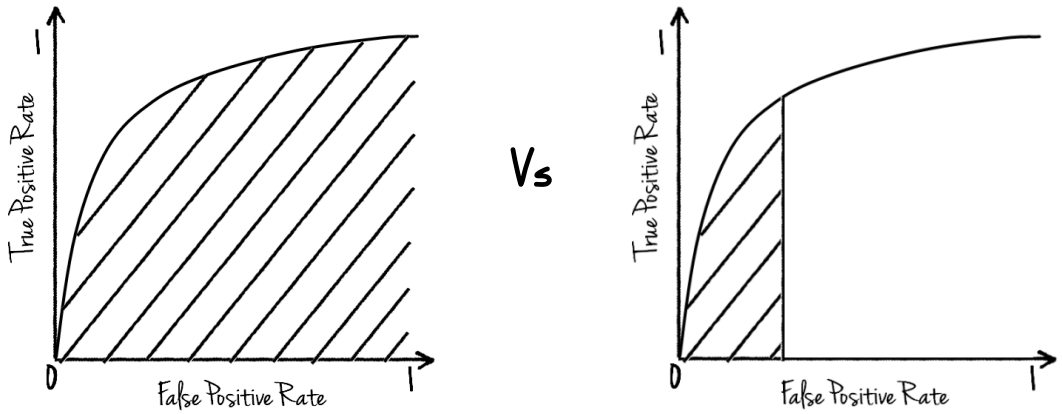
Experiments



'Partial'
Area Under the
Curve

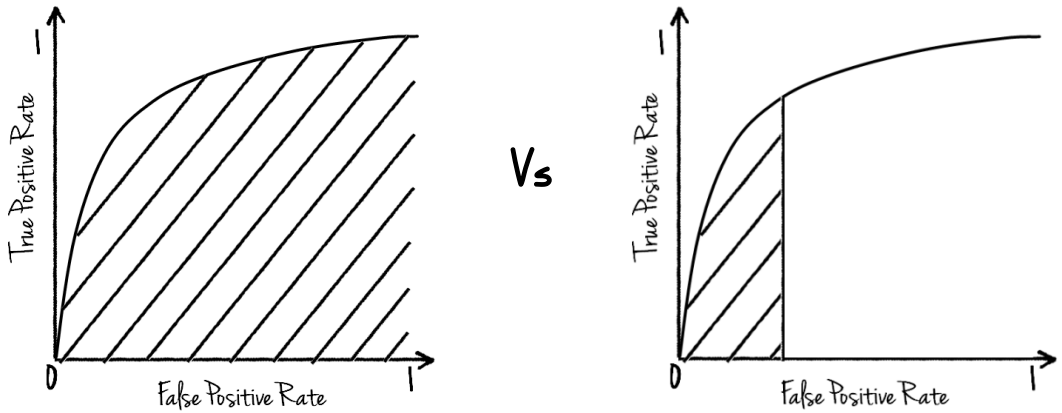
Experimental Results

- Baseline Methods:
  - Method for 'Full AUC' (Joachims, 2005)



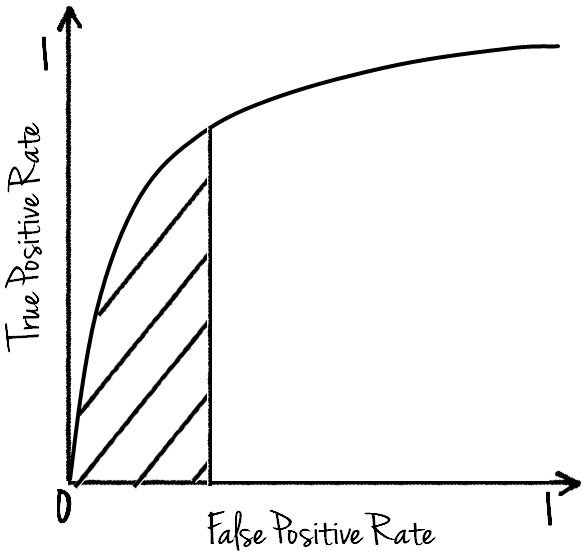
Experimental Results

- Baseline Methods:
  - Method for 'Full AUC' (Joachims, 2005)



  - Baseline 1 (Wu et al., 2008)
  - Baseline 2 (Komori & Eguchi, 2010)
  - Baseline 3 (Ricamato & Tortorella, 2011)

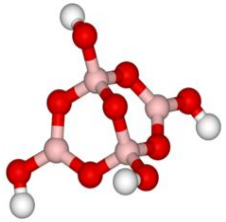
Experimental Results



Drug Discovery

50 active compounds / 2092 inactive compounds

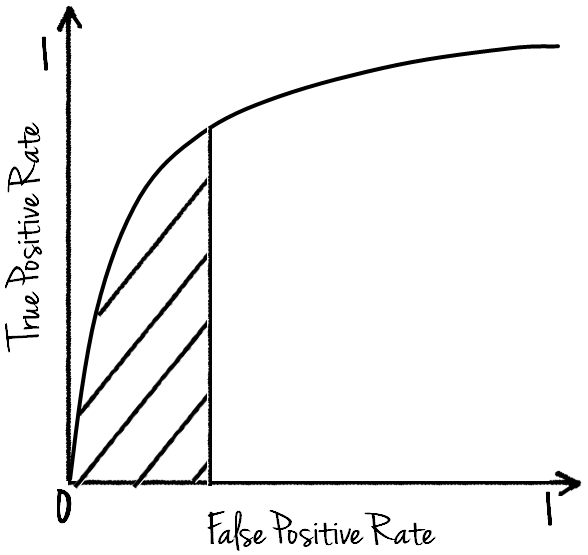
	Partial AUC in [0, 0.1]
Partial AUC Method	**65.25**
Full AUC Method	62.64
Baseline 1	63.80
Baseline 2	43.89
Baseline 3	8.33



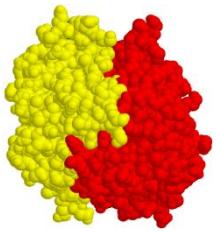
Experimental Results

Protein-Protein Interaction Prediction

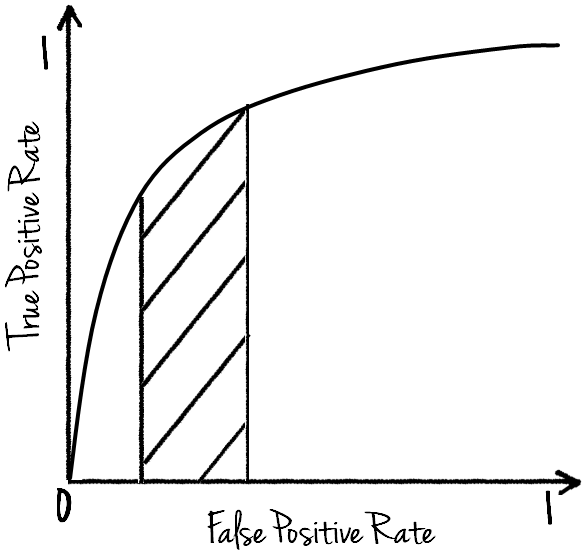
~3×10^3 interacting pairs / ~2×10^5 non-interacting pairs



	Partial AUC in [0, 0.1]
Partial AUC Method	**51.79**
Full AUC Method	39.72
Baseline 1	44.51
Baseline 2	48.65
Baseline 3	47.33



Experimental Results

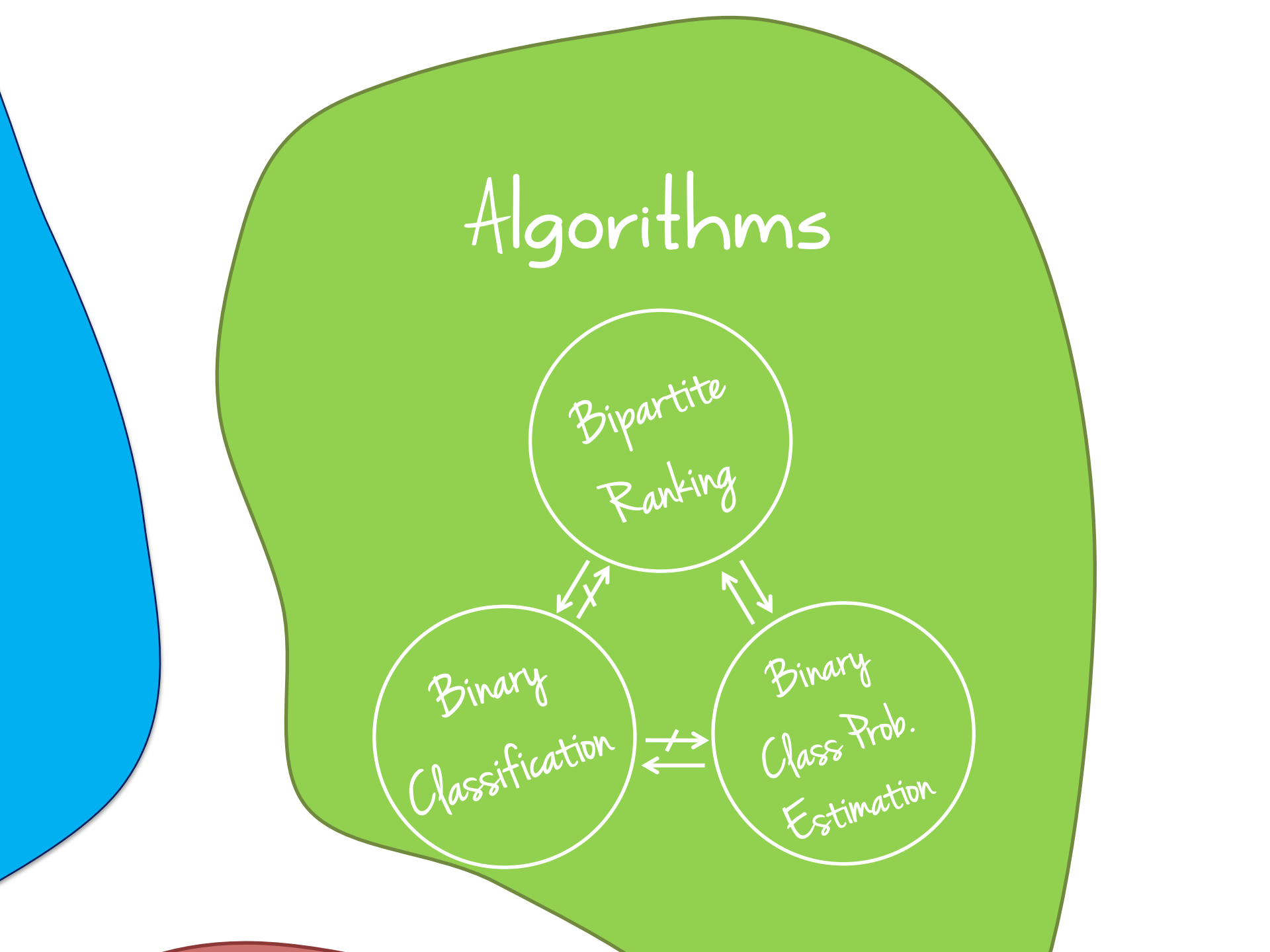


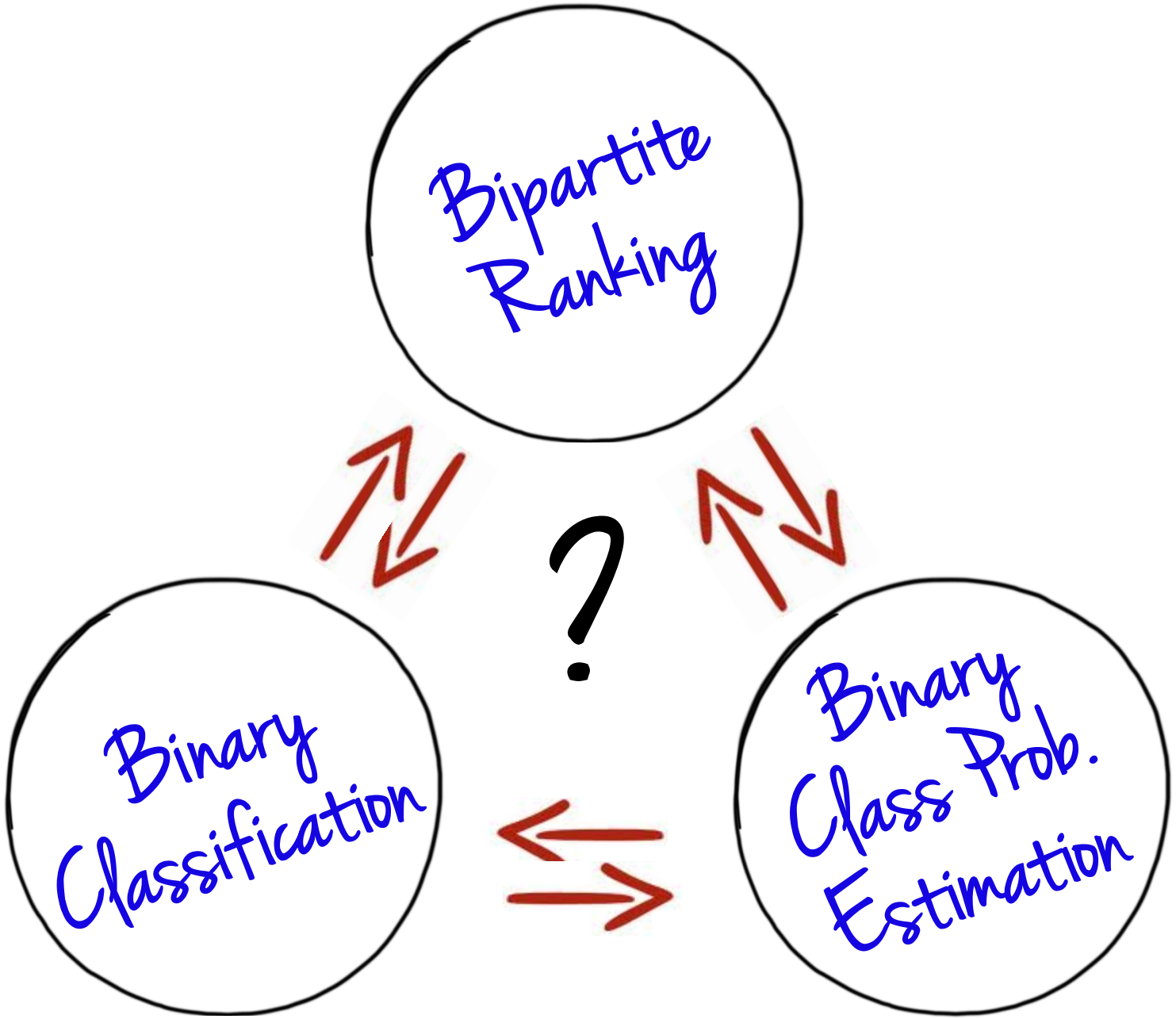
KDD Cup 2008

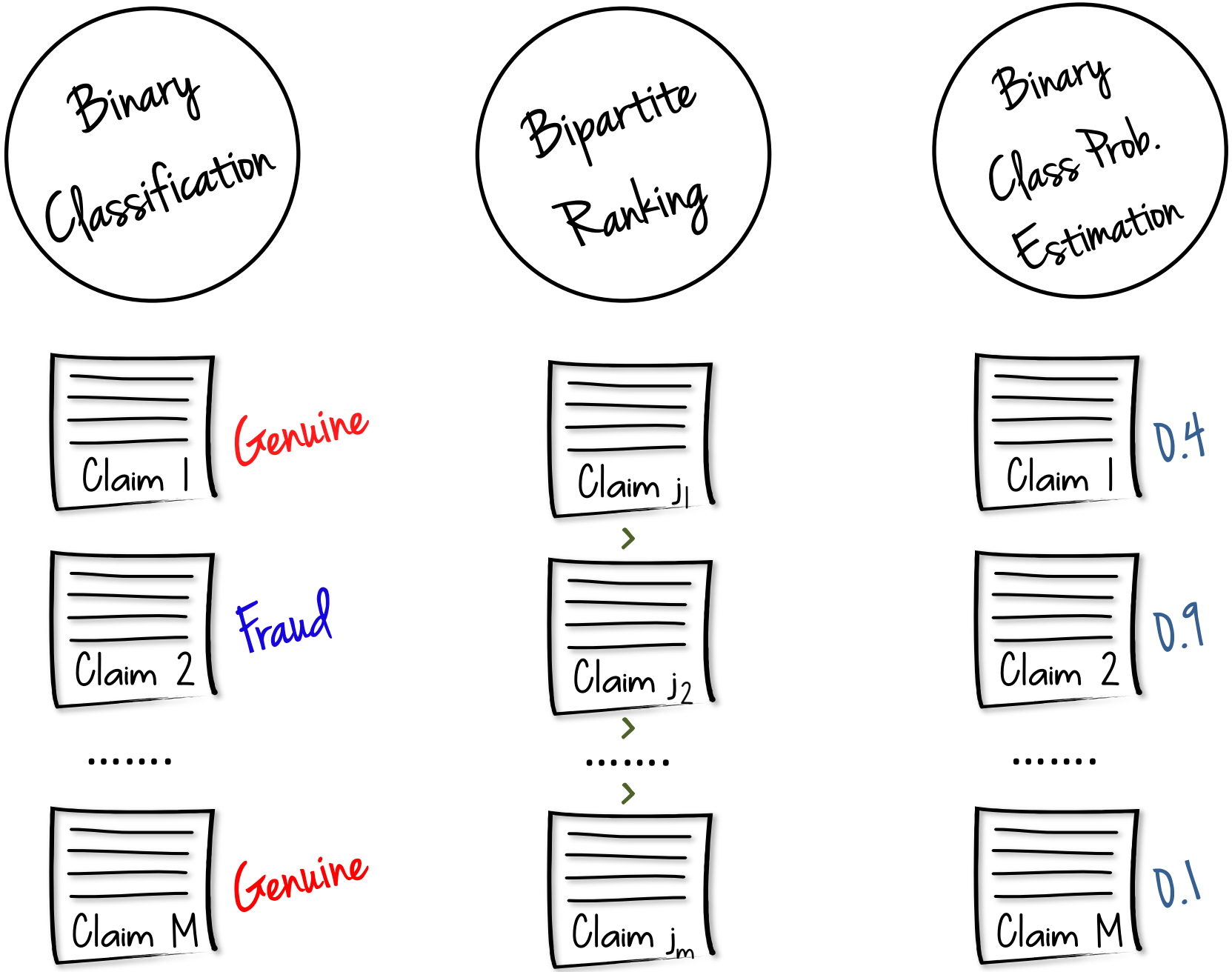
Breast Cancer Detection

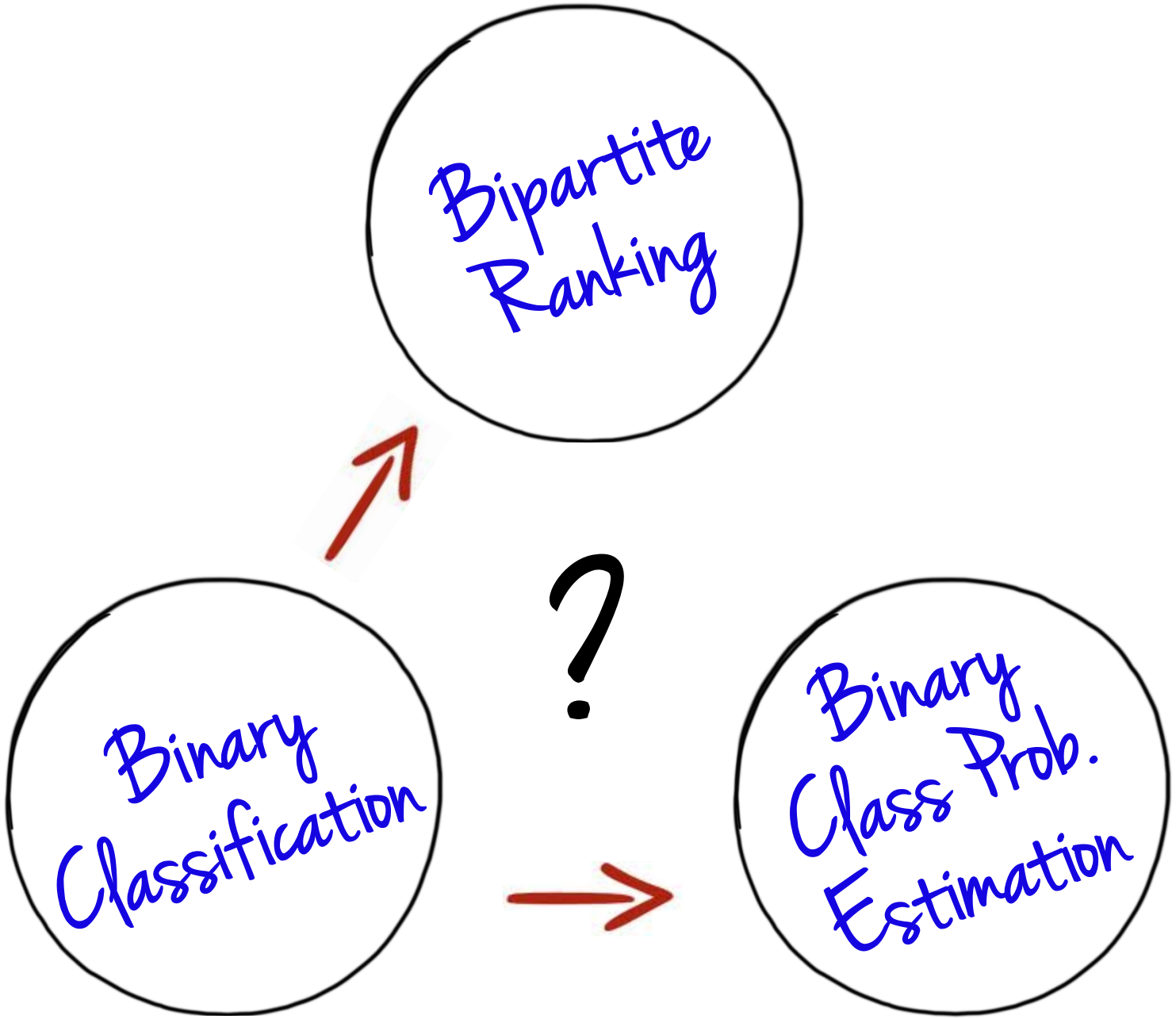
~600 malignant ROIs / ~10^5 benign ROIs

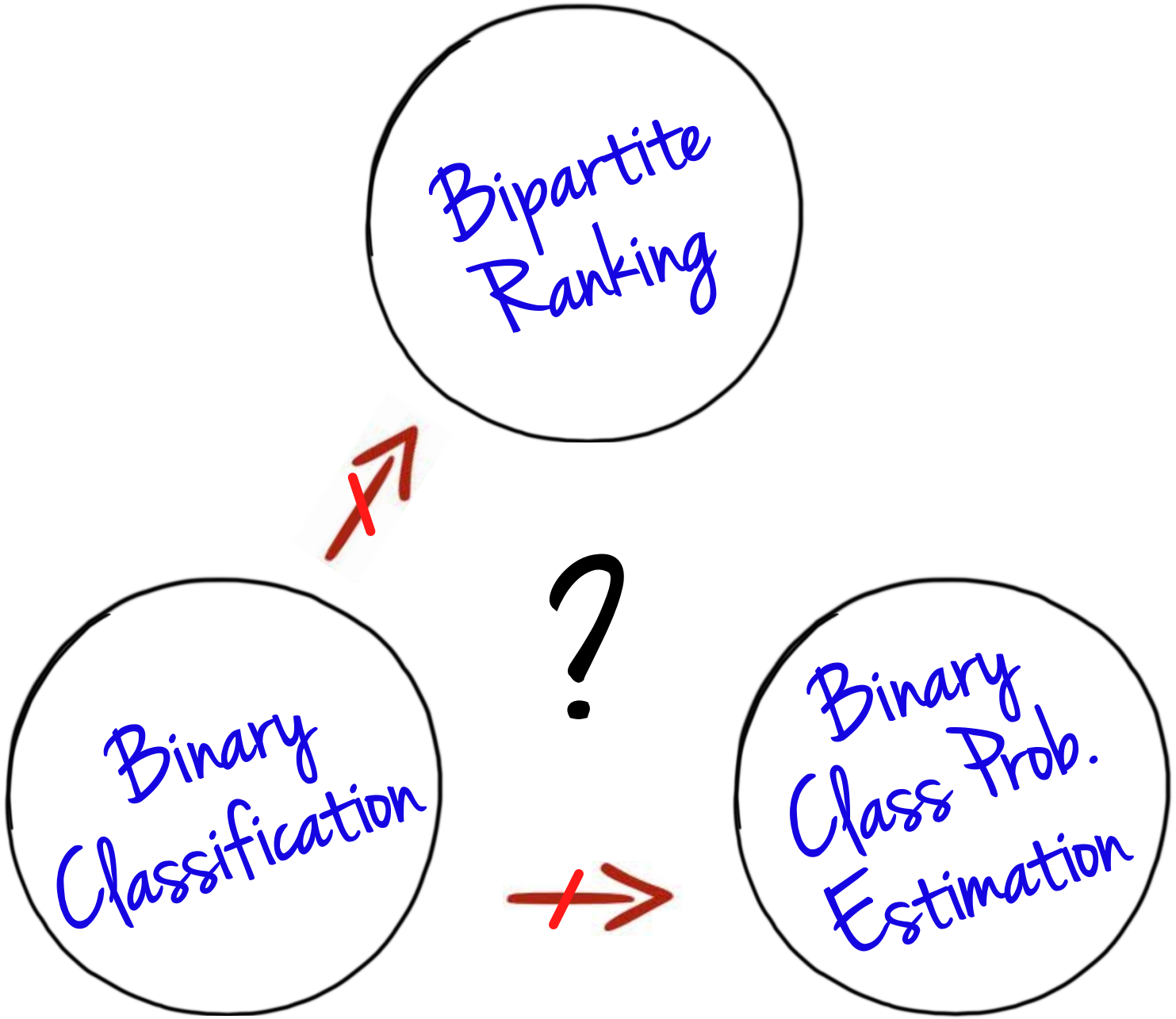
	Partial AUC in [0.2s, 0.3s]
Partial AUC Method	**51.44**
Full AUC Method	50.50
Baseline 1	48.06
Baseline 2	46.99

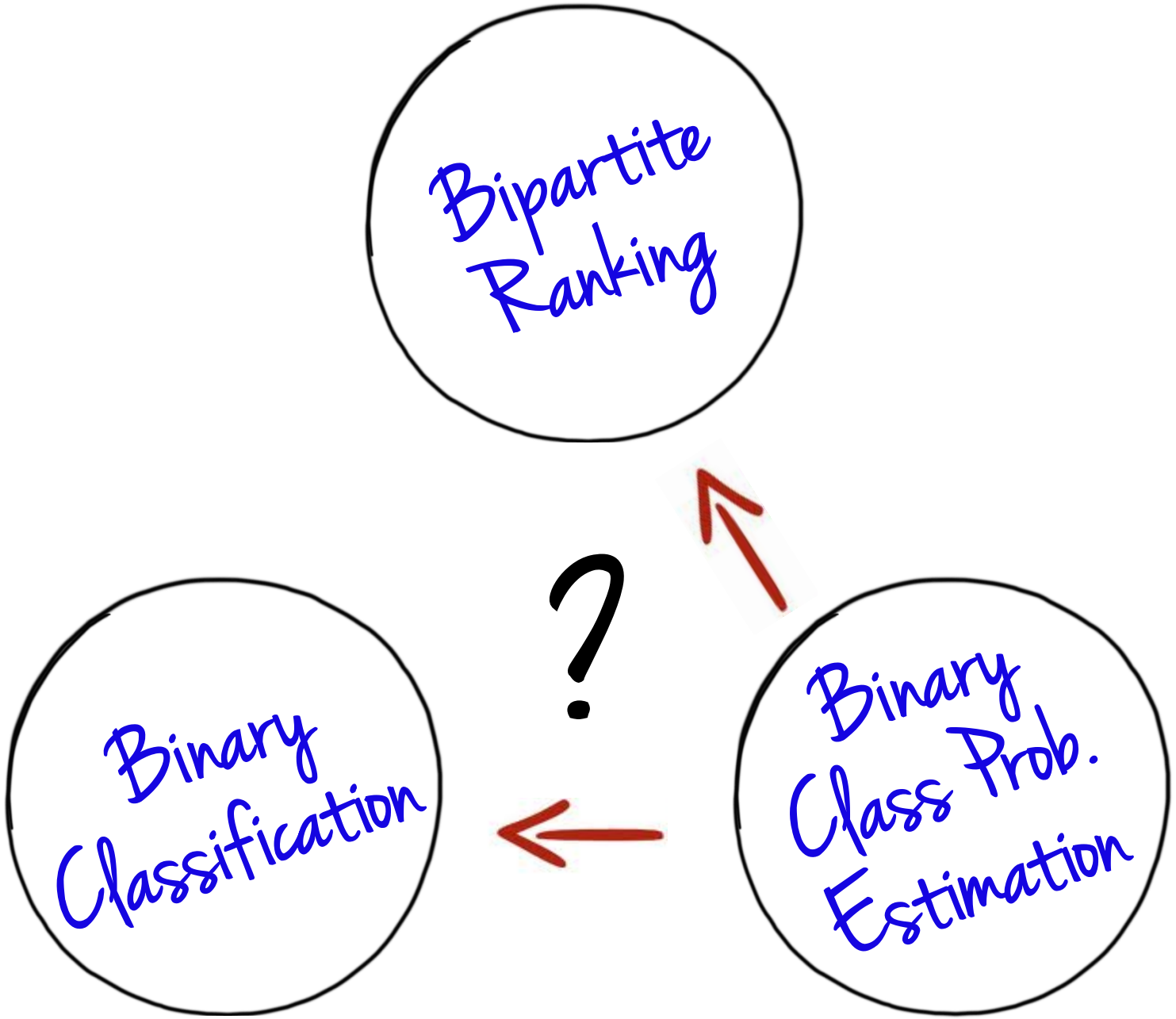

Algorithms
Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation

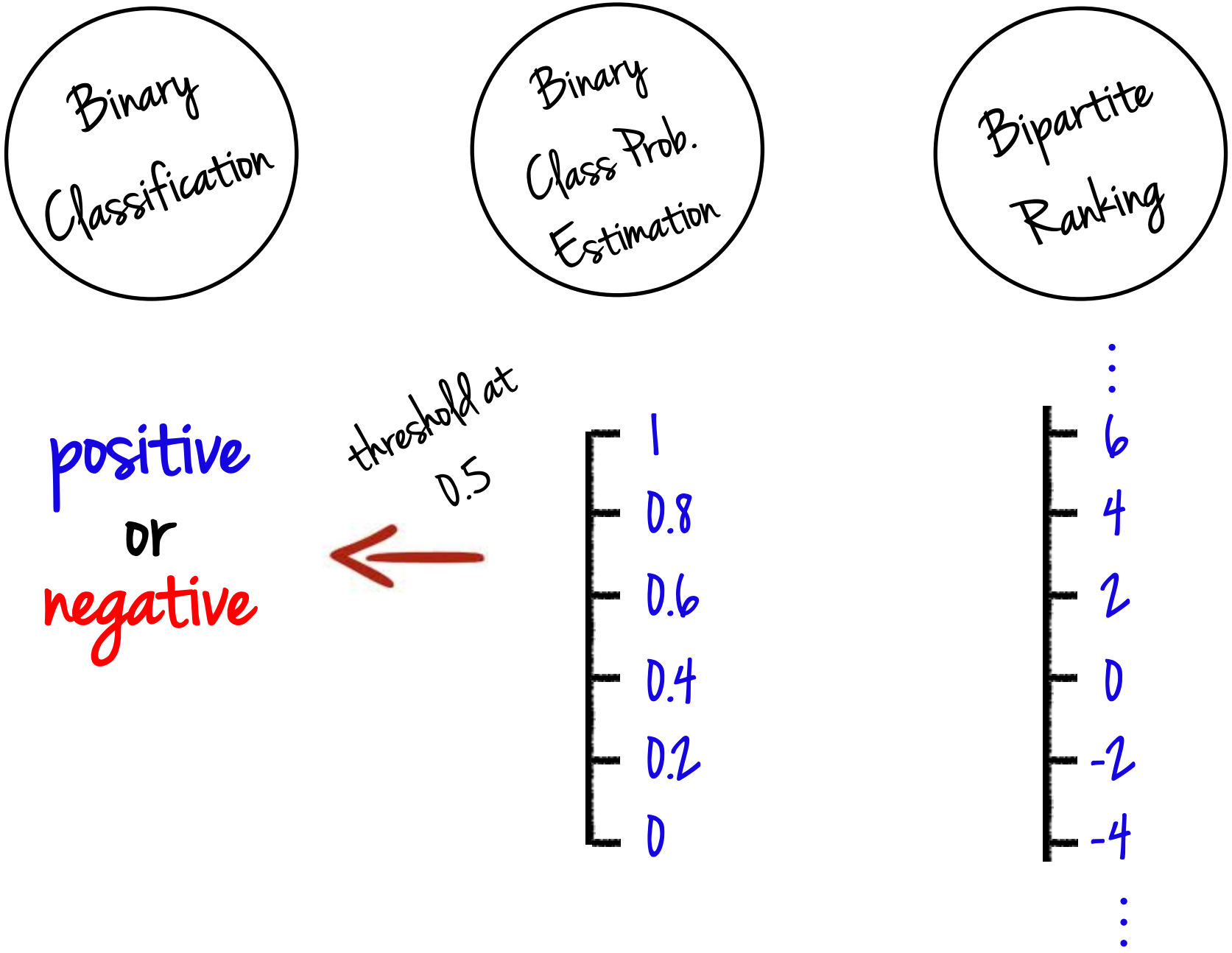

Bipartite Ranking
?
Binary Classification
Binary Class Prob. Estimation

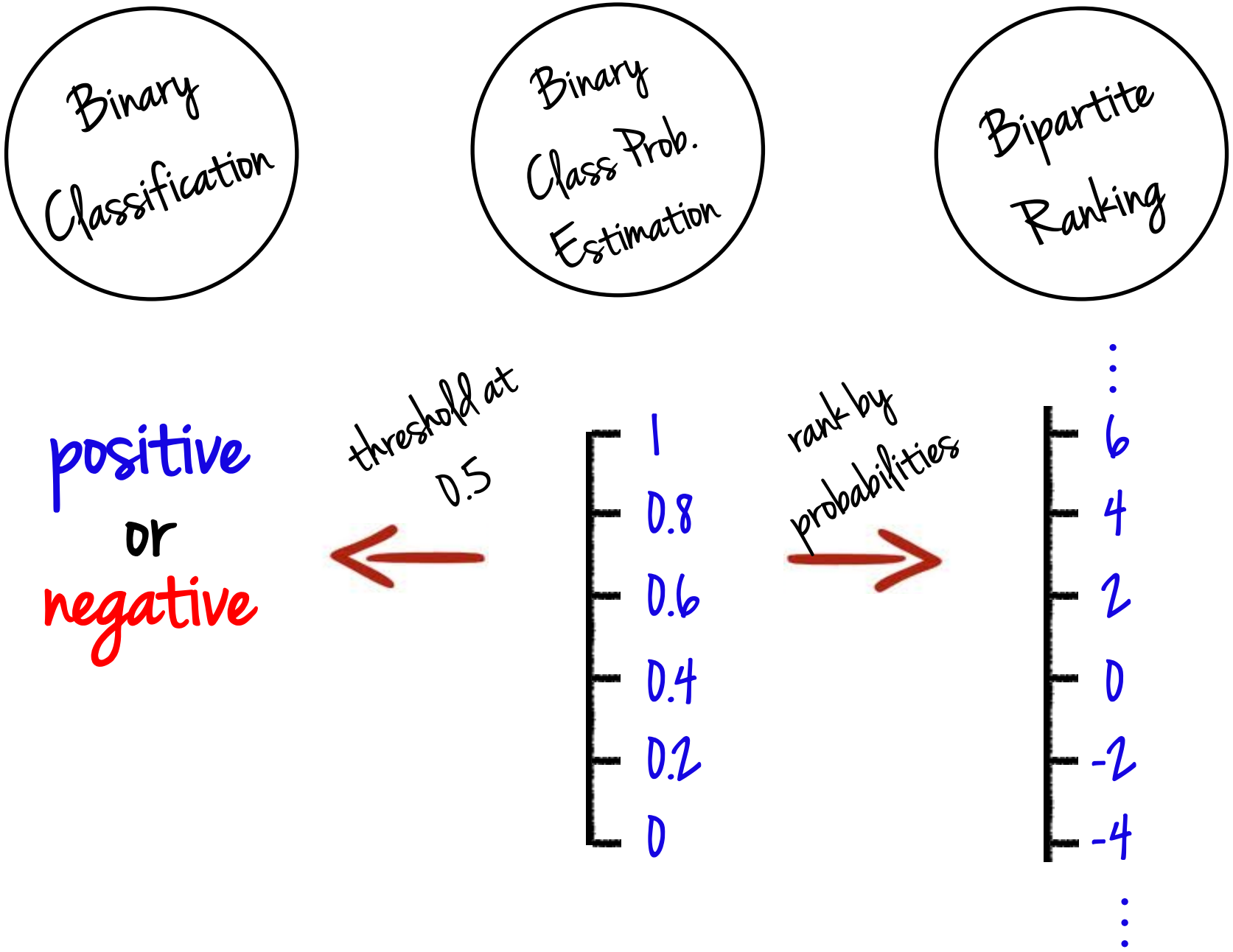

Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
Claim 1
Genuine
Claim 2
Fraud
.....
Claim M
Genuine
Claim j_1
>
Claim j_2
>
.....
>
Claim j_m
Claim 1
0.4
Claim 2
0.9
.....
Claim M
0.1

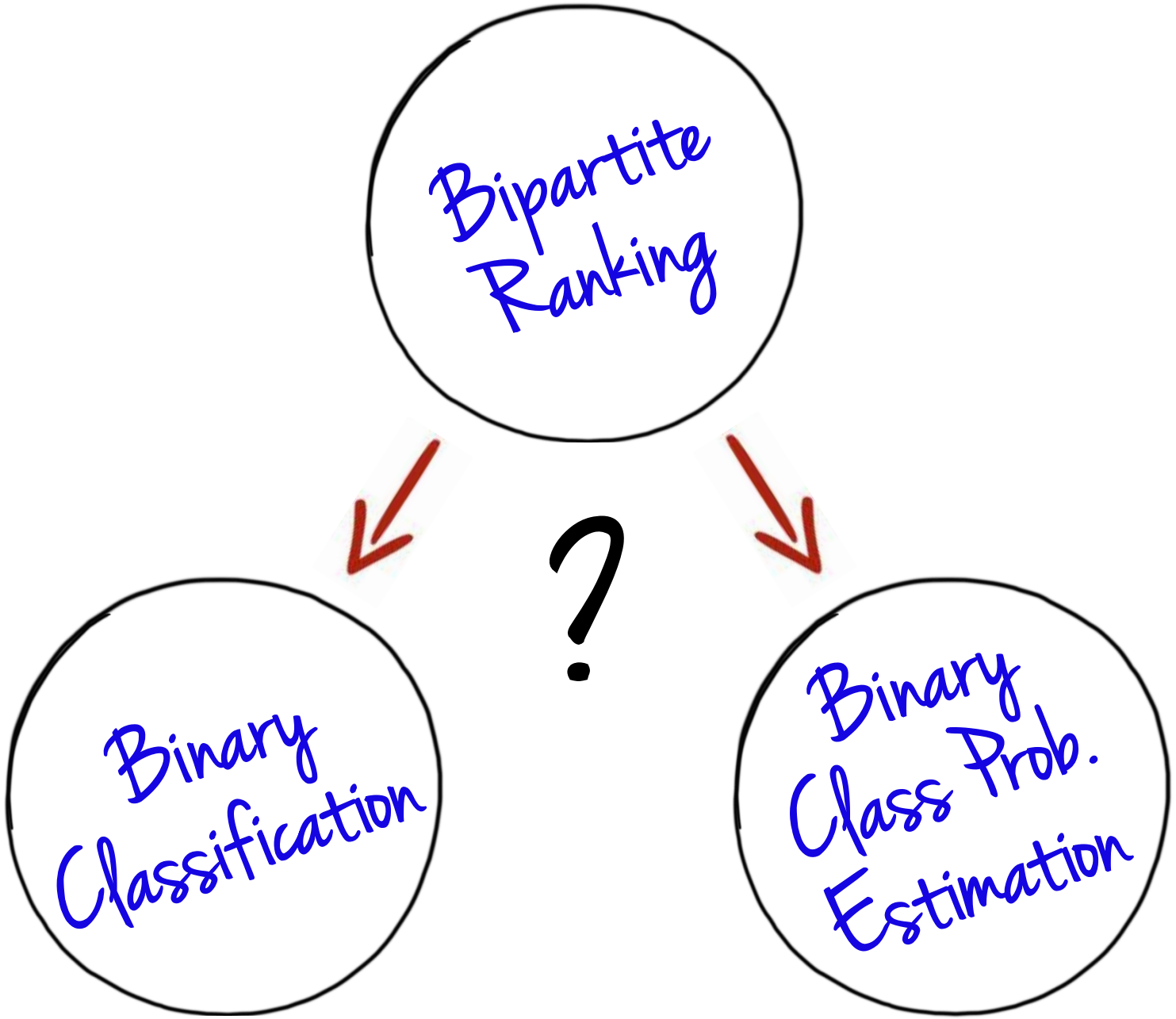

Bipartite Ranking
Binary Classification
?
Binary Class Prob. Estimation

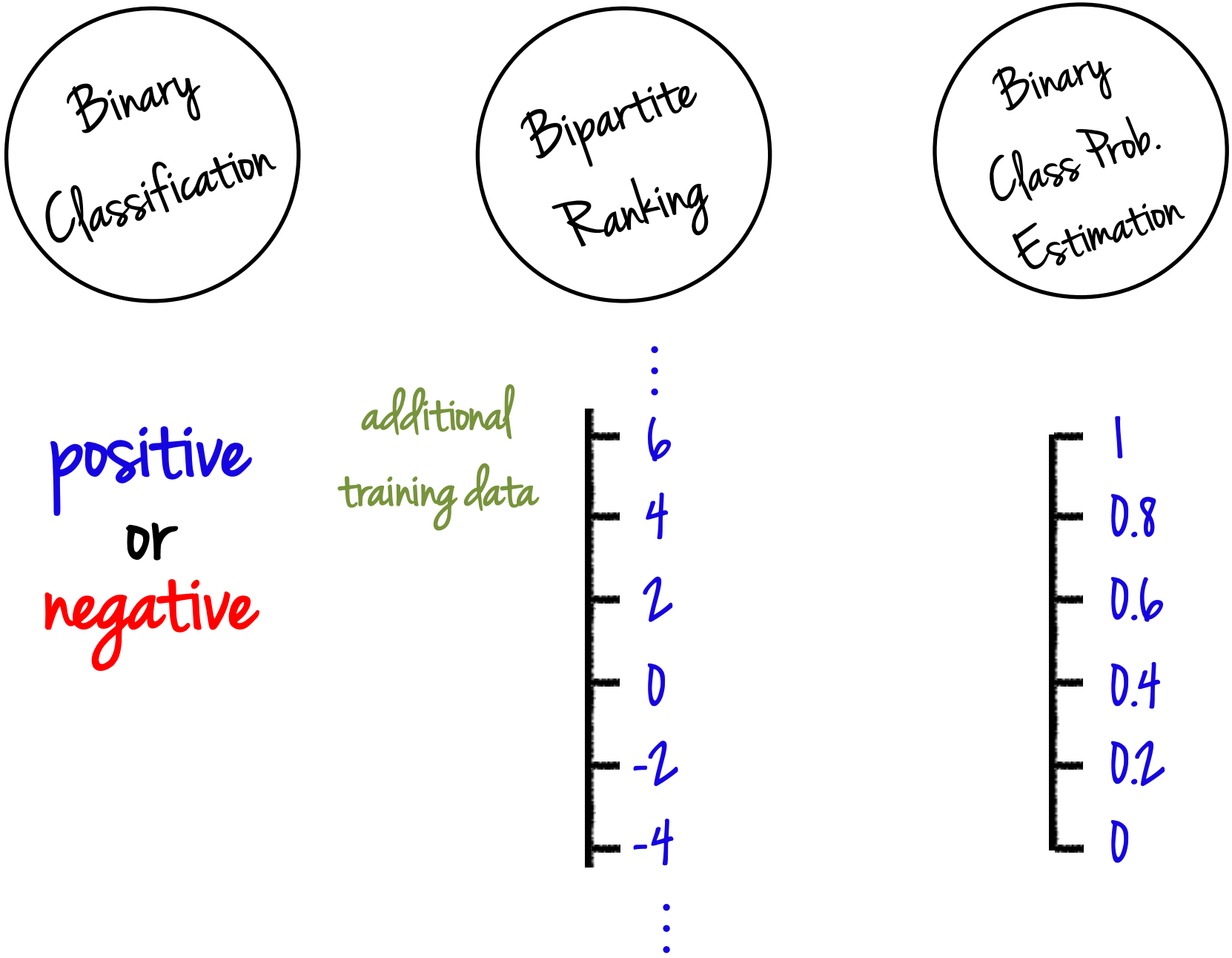

Bipartite Ranking
Binary Classification
?
Binary Class Prob. Estimation

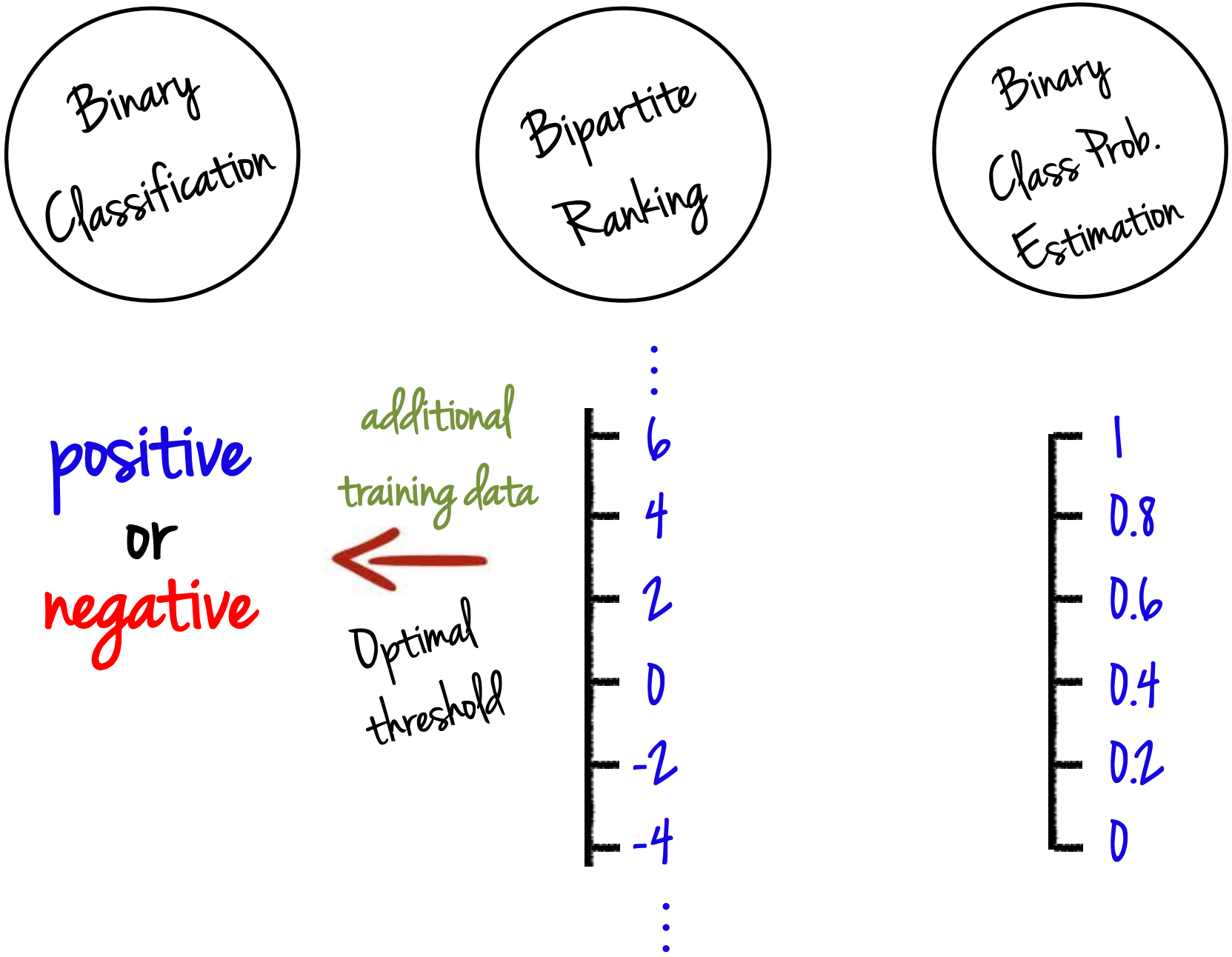

Bipartite Ranking
?
Binary Classification
Binary Class Prob. Estimation

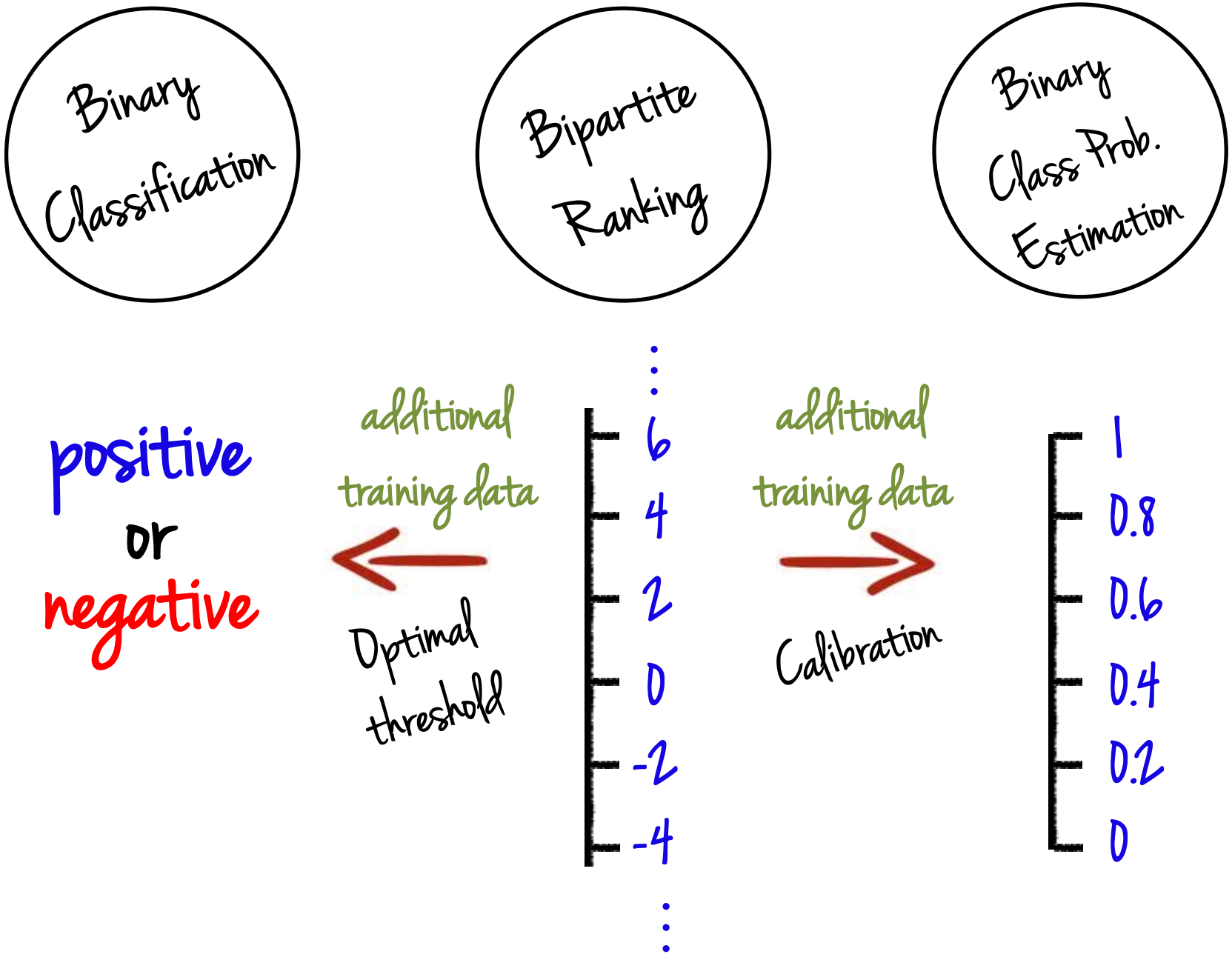

Binary Classification
Binary Class Prob. Estimation
Bipartite Ranking
positive
or
negative
threshold at 0.5
1
0.8
0.6
0.4
0.2
0
6
4
2
0
-2
-4


Binary Classification
Binary Class Prob. Estimation
Bipartite Ranking
positive
or
negative
threshold at 0.5
1
0.8
0.6
0.4
0.2
0
rank by probabilities
6
4
2
0
-2
-4

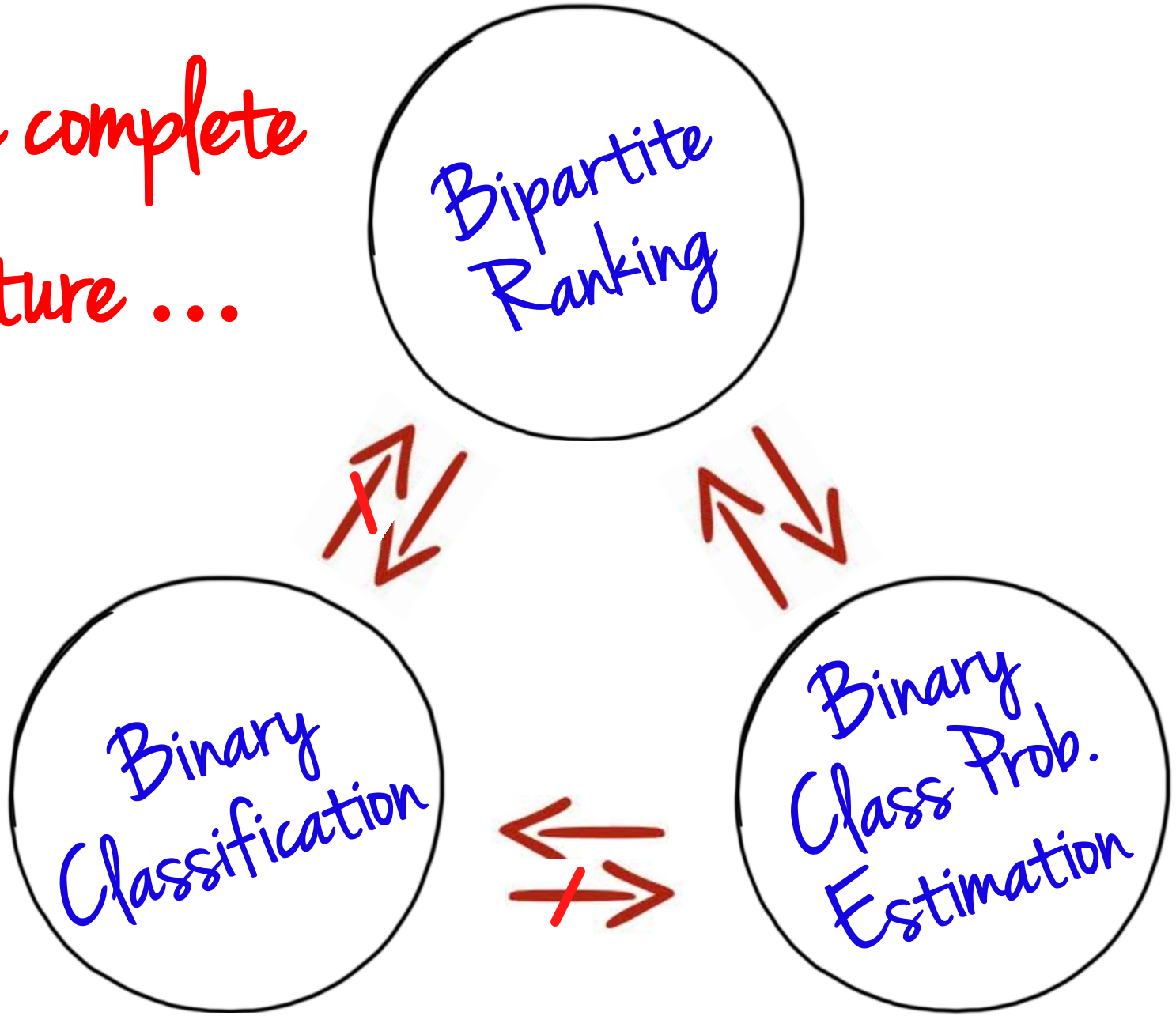

Bipartite Ranking
?
Binary Classification
Binary Class Prob. Estimation


Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
positive
or
negative
additional training data
6
4
2
0
-2
-4
1
0.8
0.6
0.4
0.2
0

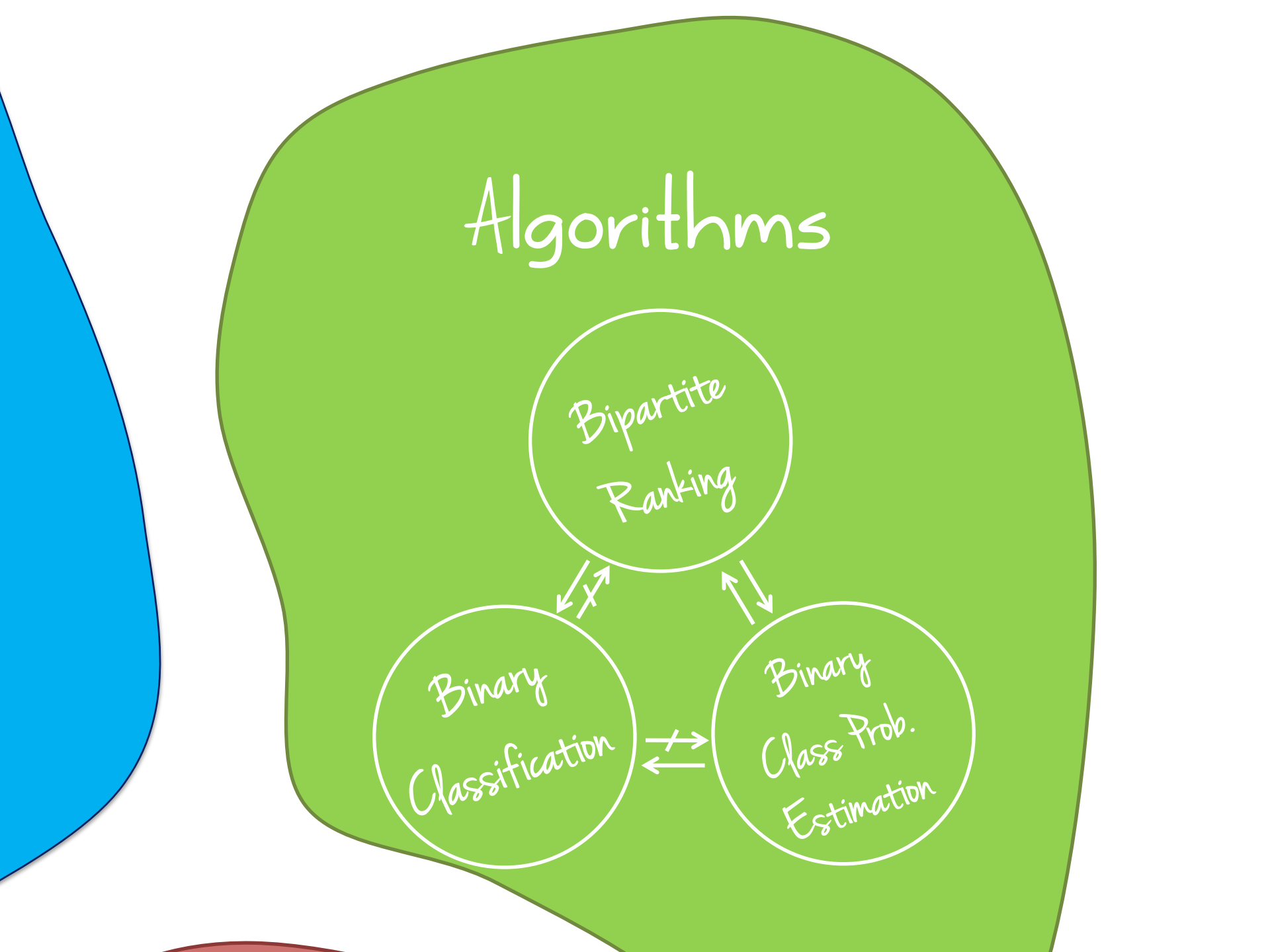

Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
positive
or
negative
additional training data
Optimal threshold
6
4
2
0
-2
-4
1
0.8
0.6
0.4
0.2
0

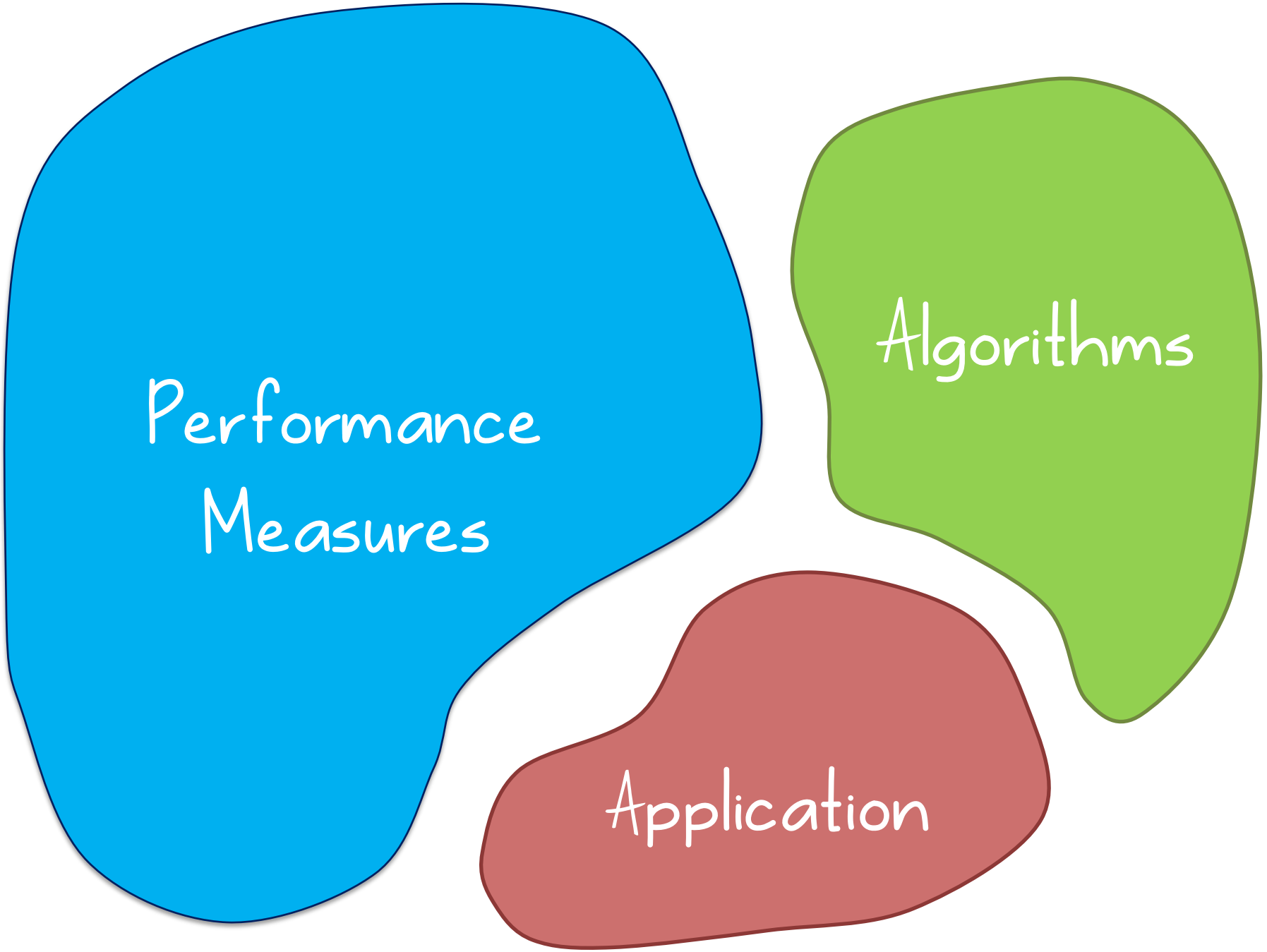

Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
positive
or
negative
additional training data
Optimal threshold
additional training data
Calibration
6
4
2
0
-2
-4
1
0.8
0.6
0.4
0.2
0

the complete picture …

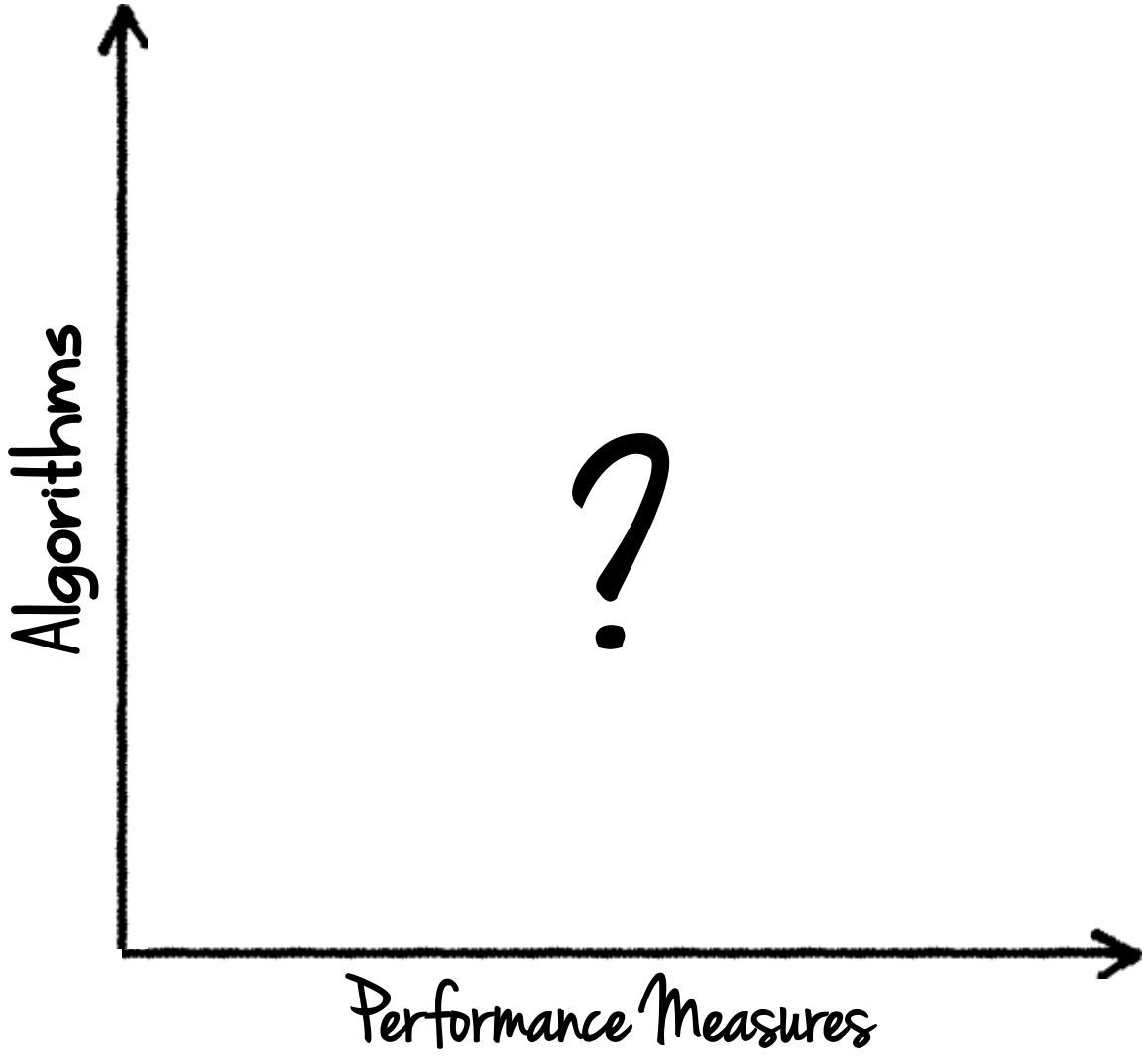


Narasimhan, H. and Agarwal, S. "*On the relationship between binary classification, bipartite ranking, and binary class probability estimation*". NIPS 2013.

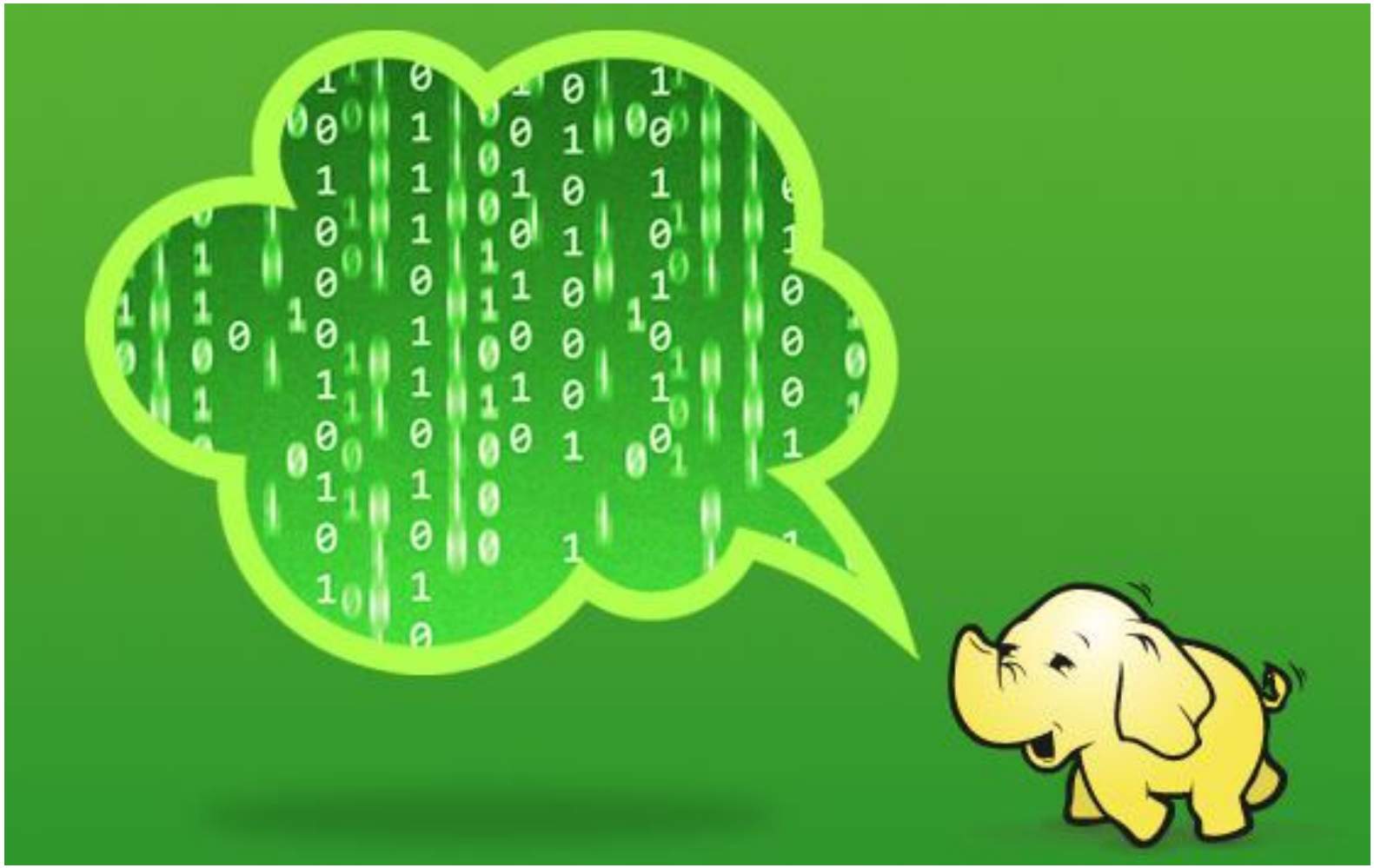

Algorithms
Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation


Performance Measures
Algorithms
Application

take away message ...



open challenges …



www.siliconangle.com

open challenges …

Big Data?

Performance Measure	Run-time	References
F-measure PBREP Precision@K	$O(N^2)$	Joachims (2005);
AUC Partial AUC Pos@Top	$O(N \log N)$	Joachims (2005, 2006); Narasimhan & Agarwal (2013 a,b)
Average Precision	$O(N^2)$	Yue et al. (2007)
Discounted Cumulative Gain (DCG) *	$O(Nk)$	Chakrabarti et al. (2008)
Mean Reciprocal Rank (MRR) *	$O(N \log N + k^2)$	Chakrabarti et al. (2008)

Acknowledgements

Machine Learning and Learning Theory Group @ IISc

Shivani Agarwal

Harish Guruprasad Ramasamy

Siddarth Ramamohan

Arun Rajkumar

Rohit Vaish

Arpit Agarwal

Saneem Ahmed

Suprovat Ghoshal

Aadirupa Saha

Acknowledgements

Collaborators

Sanjay Chawla, University of Sydney and NICTA, Sydney

Pradip K. Majumder, Mitra Biotech, Bangalore

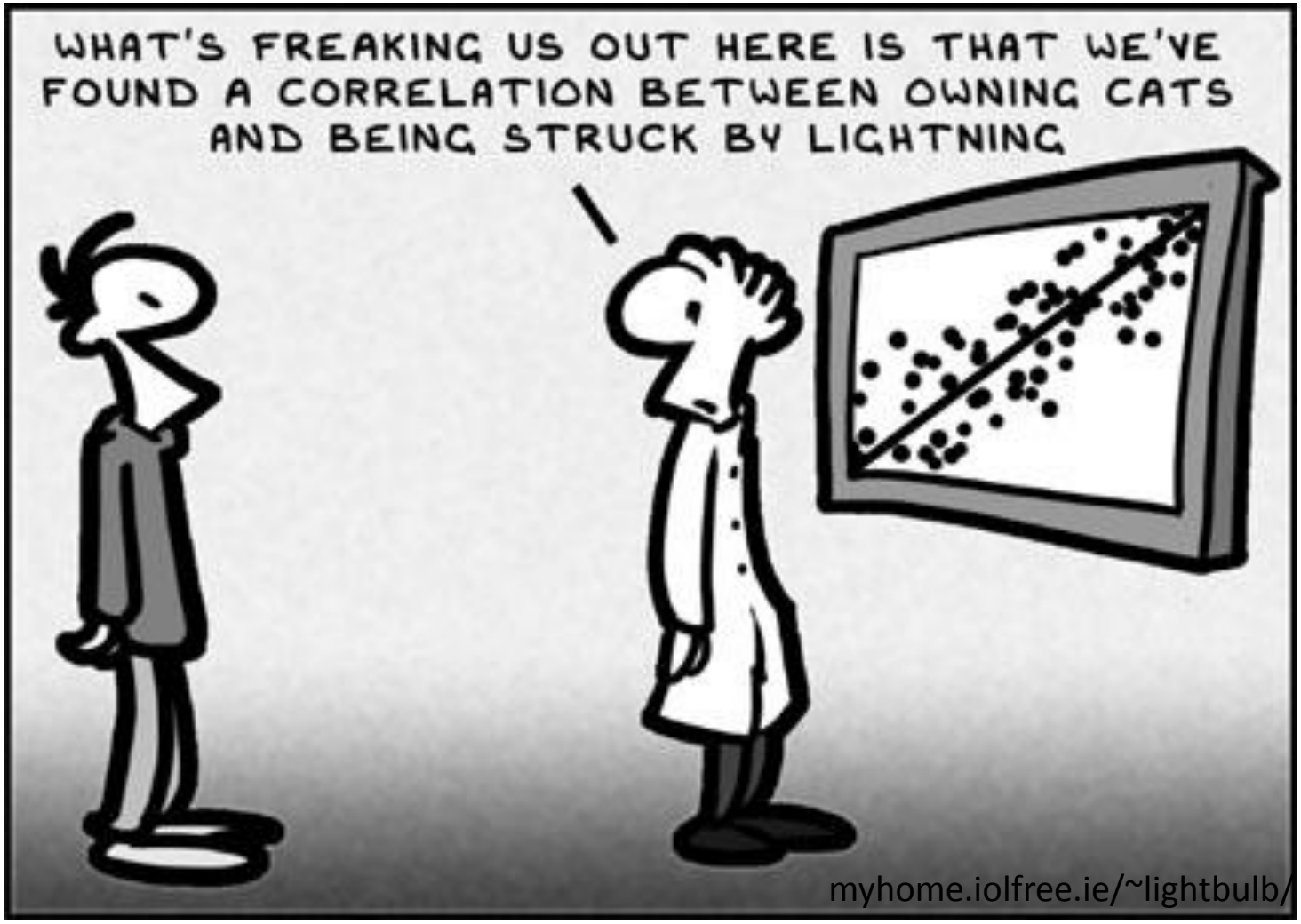
Biswanath Majumder, Mitra Biotech, Bangalore

Aditya Menon, NICTA, Sydney

Padhma Radhakrishnan, Mitra Biotech, Bangalore

Shiladitya Sengupta, Harvard Medical School, Boston

Mallikarjun Sundaram, Mitra Biotech, Bangalore


WHAT'S FREAKING US OUT HERE IS THAT WE'VE FOUND A CORRELATION BETWEEN OWNING CATS AND BEING STRUCK BY LIGHTNING
myhome.iolfree.ie/~lightbulb/

Questions?

Most images in this presentation were obtained using Google image search (http://www.go ogle.co.in/imghp)

Some of the hand-drawn shapes & clipart images were obtained from the following sites:

http://www.clker.com/

http:// www.free-power-point-templates.com /

http://neilslorance.wordpress.com/

http://yukaihan.deviantart.com/

The hand-written fonts used in this presentation were downloaded from:

http://www.dafont.com/