NATIONAL RESEARCH
UNIVERSITY

School of Data Analysis and Artificial
Intelligence Department of Computer Science

# DATA SCIENCE FOR BUSINESS

Lecture 4. Customer relationship management. Churn prediction. Classification.

Moscow, April 29th, 2022.

lzhukov@hse.ru

# CUSTOMER RELATIONSHIP MANAGEMENT

**CRM** an approach to managing a company's interaction with current and potential customers to improve business relationships with customers, focusing on customer retention and increasing sales


Company functions

- Marketing
- Sales
- Customer service support

# CUSTOMER RELATIONSHIP MANAGEMENT

**The service sector consists of the production of services instead of end products**

- Competitive landscape, easy to switch
- High customer acquisition cost
- Repetitive interaction - service
- Large customer LTV

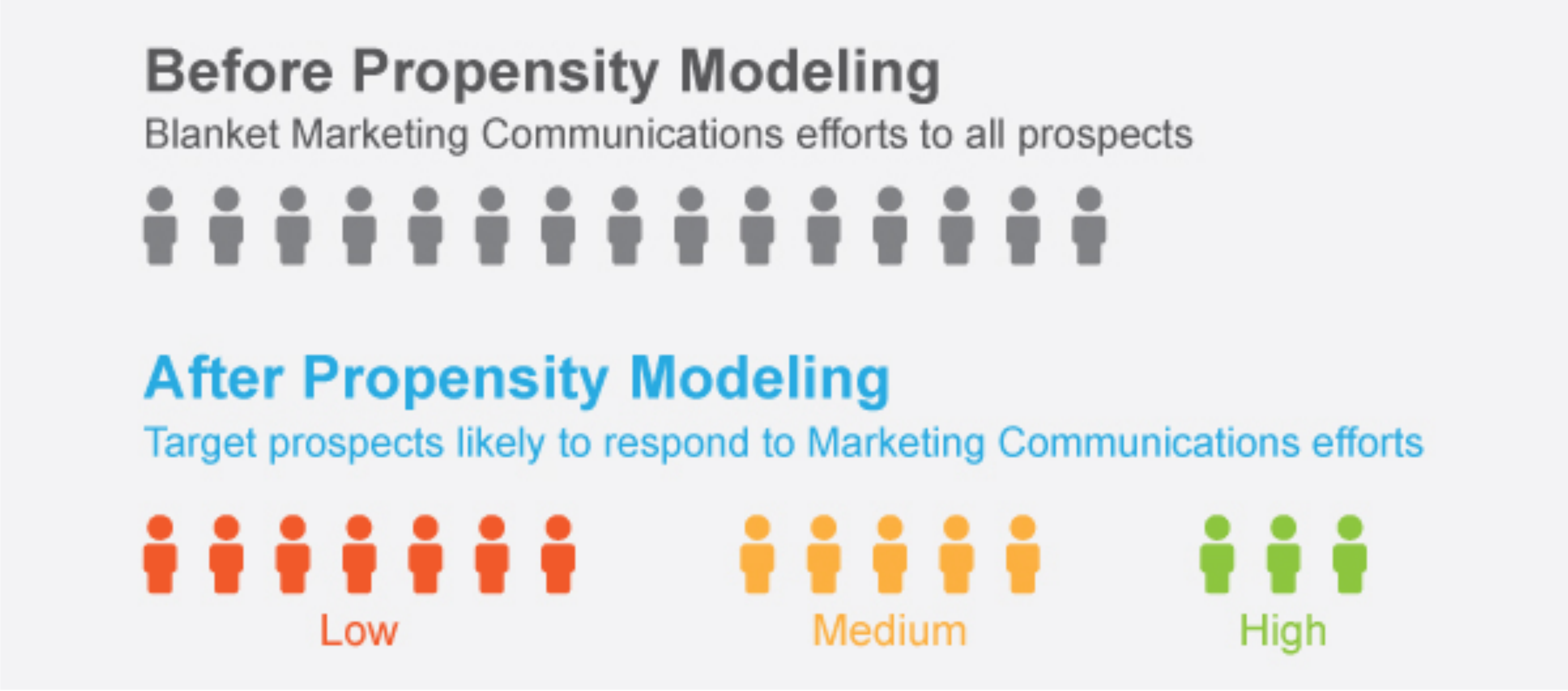
Service sector industries

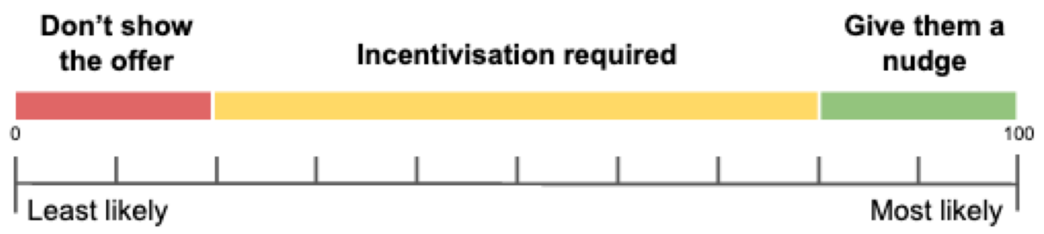
- **Telecommunication**
- Healthcare
- Education
- **Financial services (banking, insurance, investment**)
- Professional services (accounting, legal, consulting)
- Transportation
- Hospitality and tourism
- Mass media
- Real estate
- Utilities

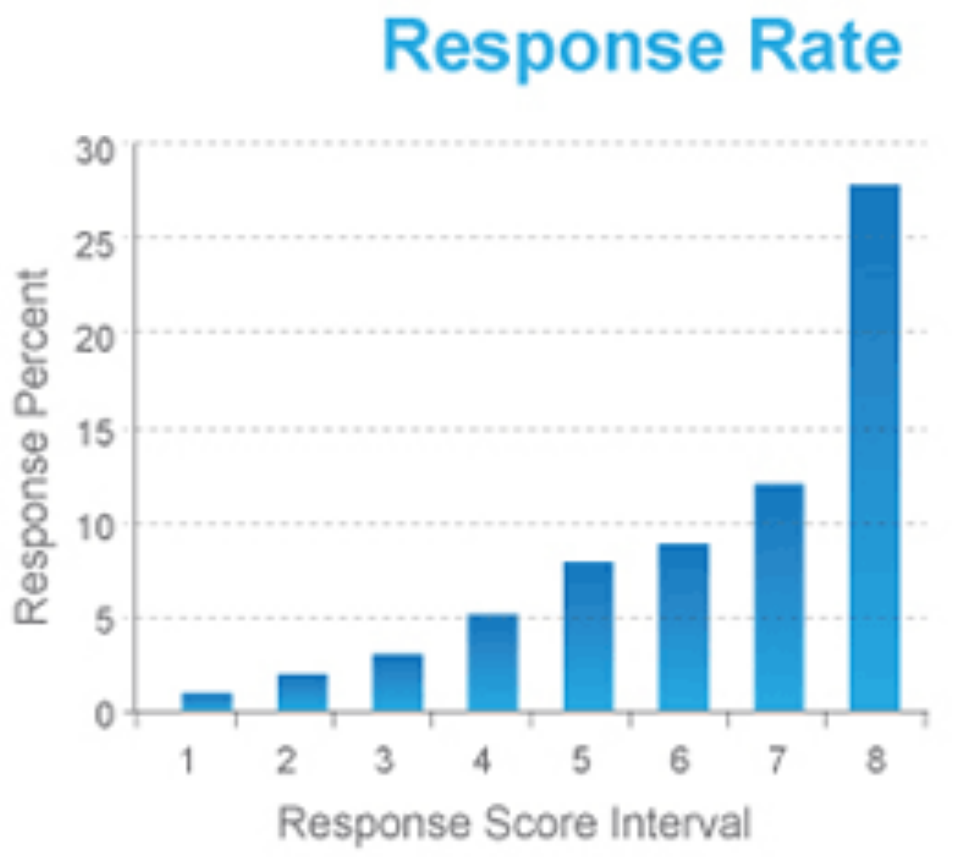
# PROPENSITY MODELING

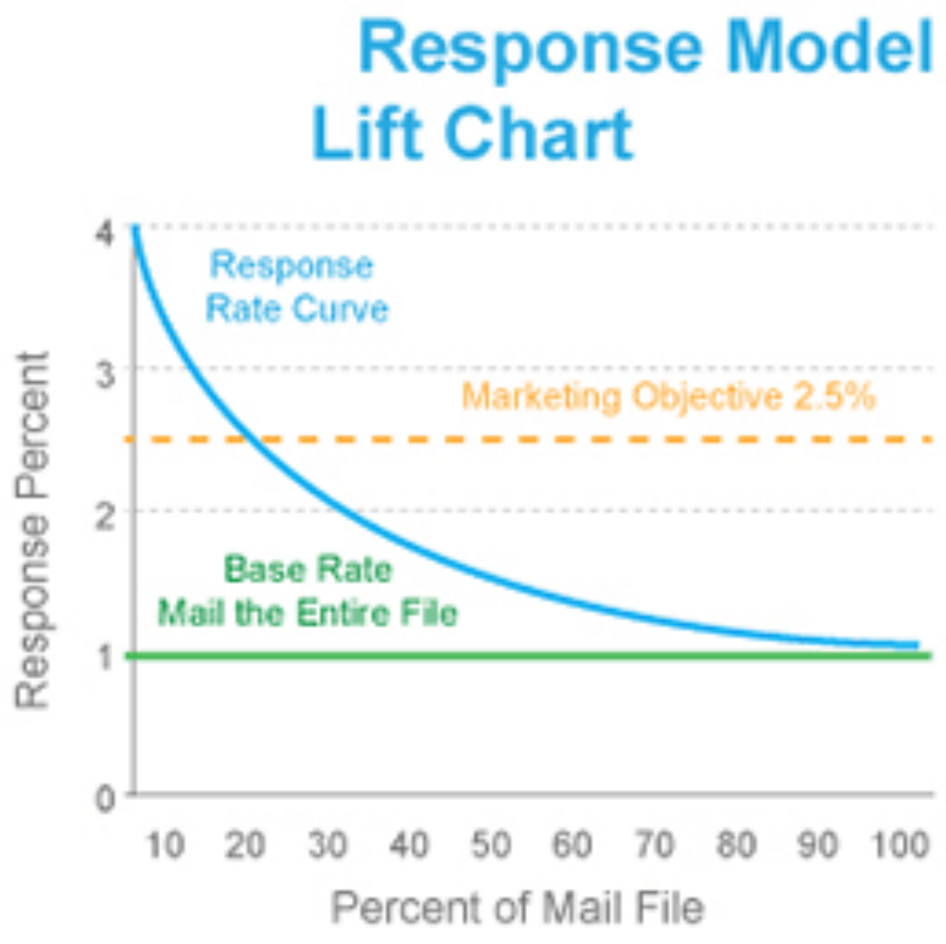
## Efficient customer communication and interaction

- Propensity to buy new service or product
- Propensity to respond to an offer
- Propensity to churn



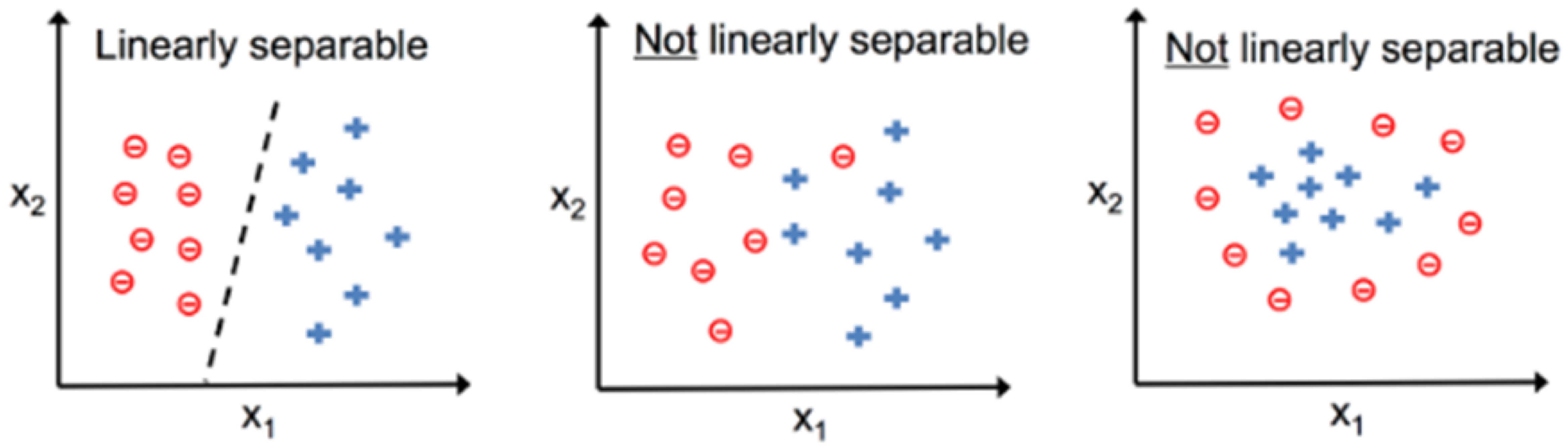
# CLASSIFICATION ALGORITHMS

## Classification algorithms

- Binary classification
- Multi-class classification

## Results:

- Class assignments
- Probability or score/ranking




## Examples of classification algorithms

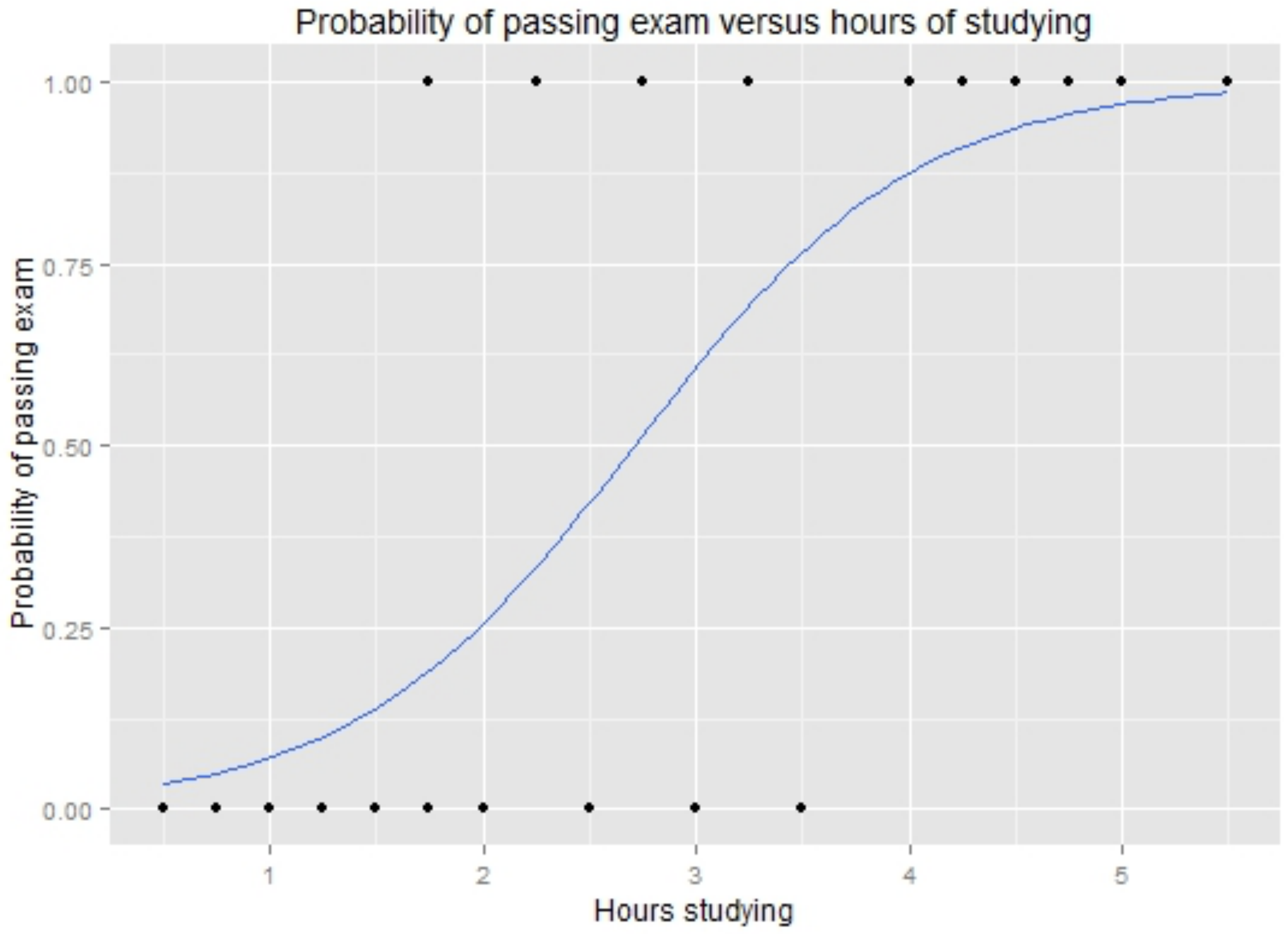
- Logistic regression
- Decision trees
- KNN k-nearest neighbors
- Naïve Bayes
- Ensemble methods:
- Random forest
- Gradient boosting decision trees (xgboost)
- Neural networks & DL

# LOGISTIC REGRESSION

Problem set-up:

- Target variable y – binary 0/1 (class)
- Predictors (x1,x,2,x3 ) – numerical

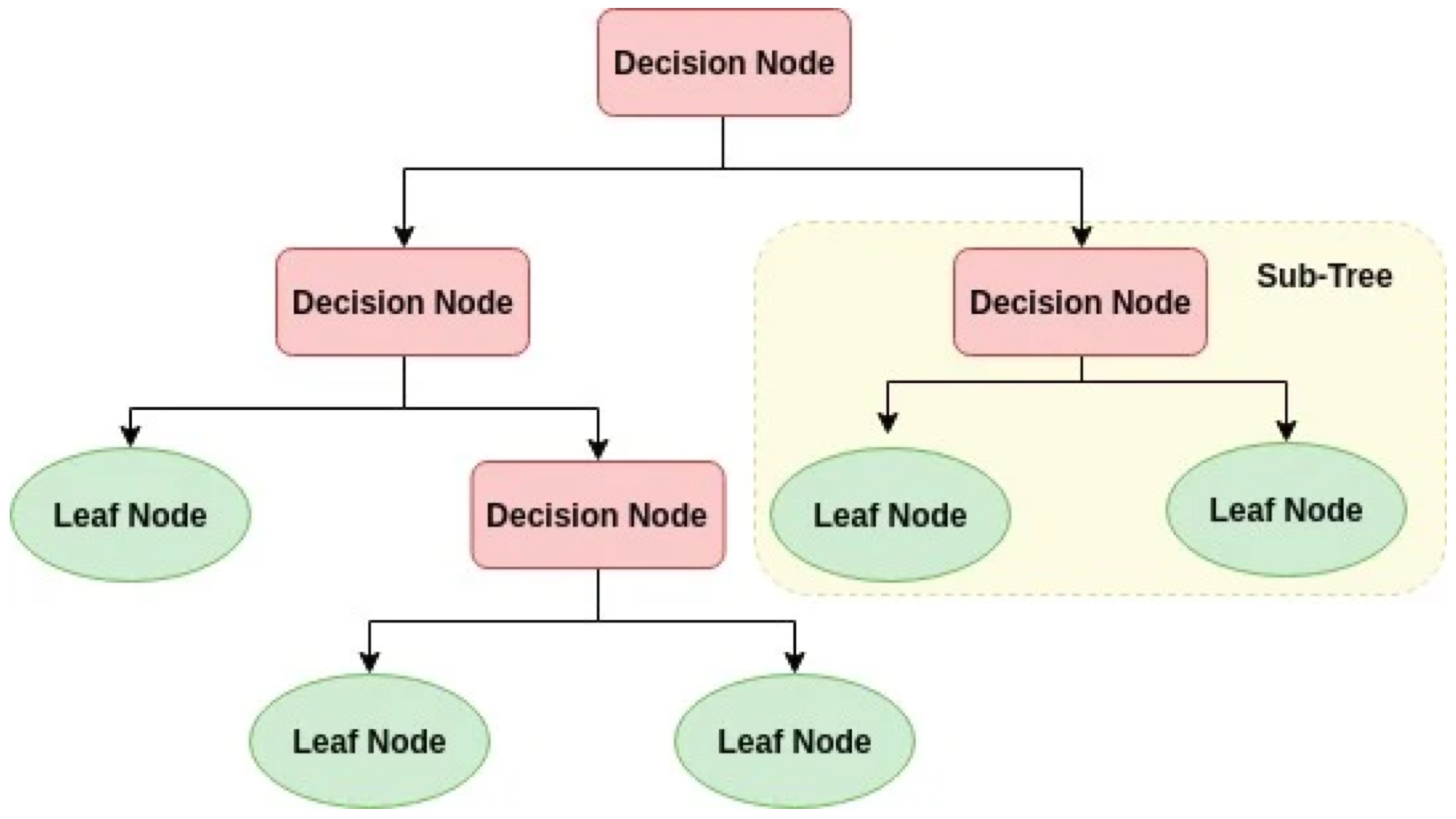
$$P(y^{(i)} = 1) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_p x_p^{(i)}))}$$

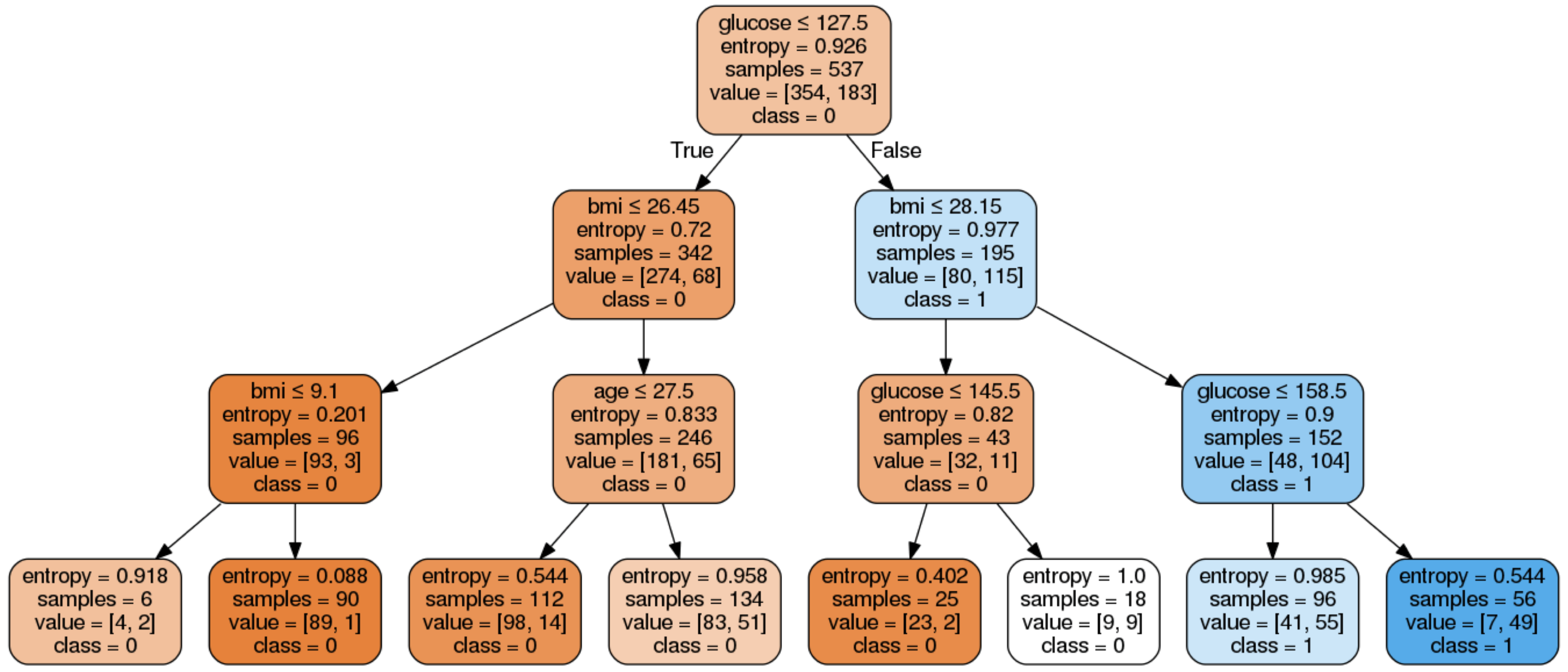


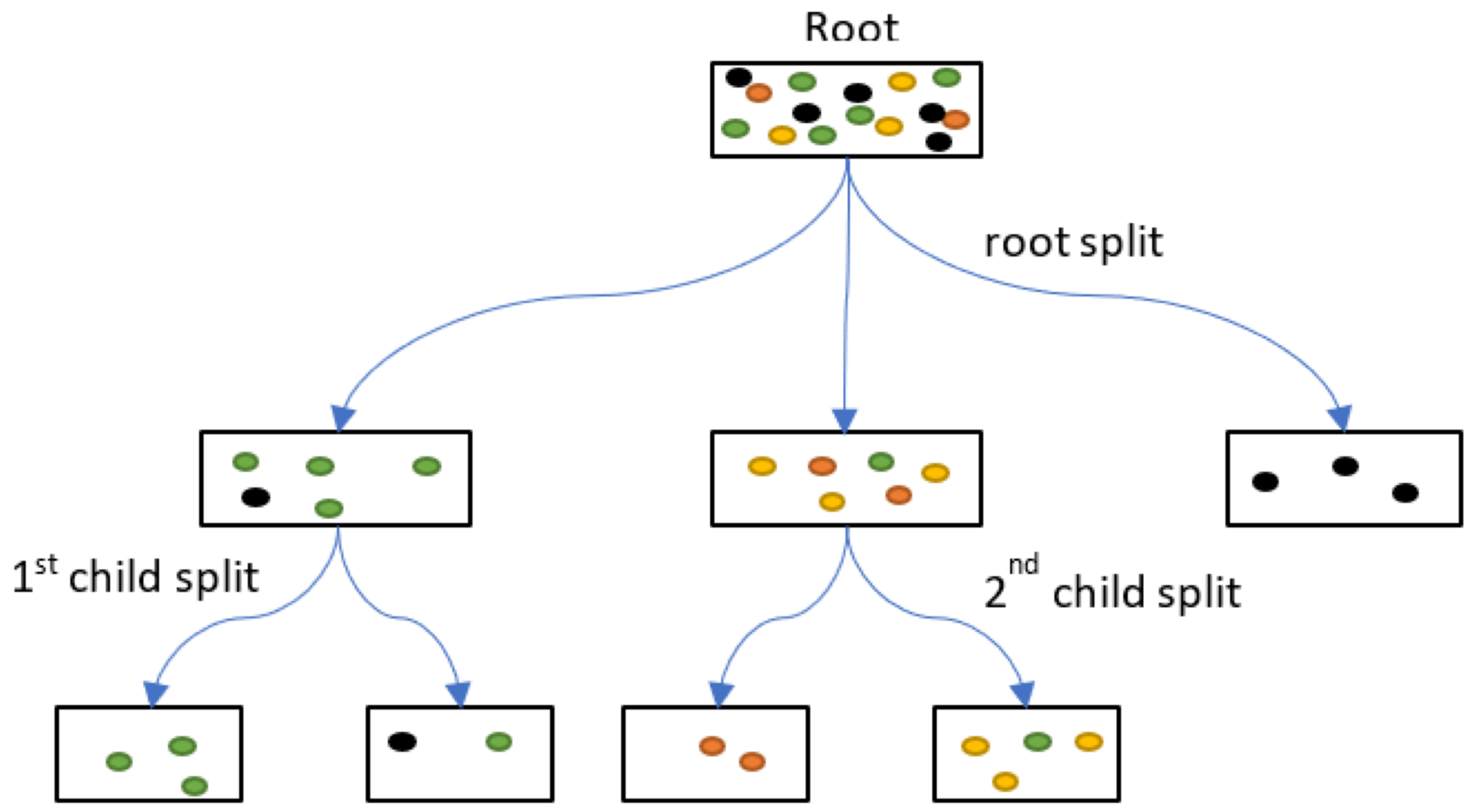
# DECISION TREE

Problem set-up:

- Target variable binary class / multi-class
- Predictors (x1,x,2,x3 ) – numerical or categorical



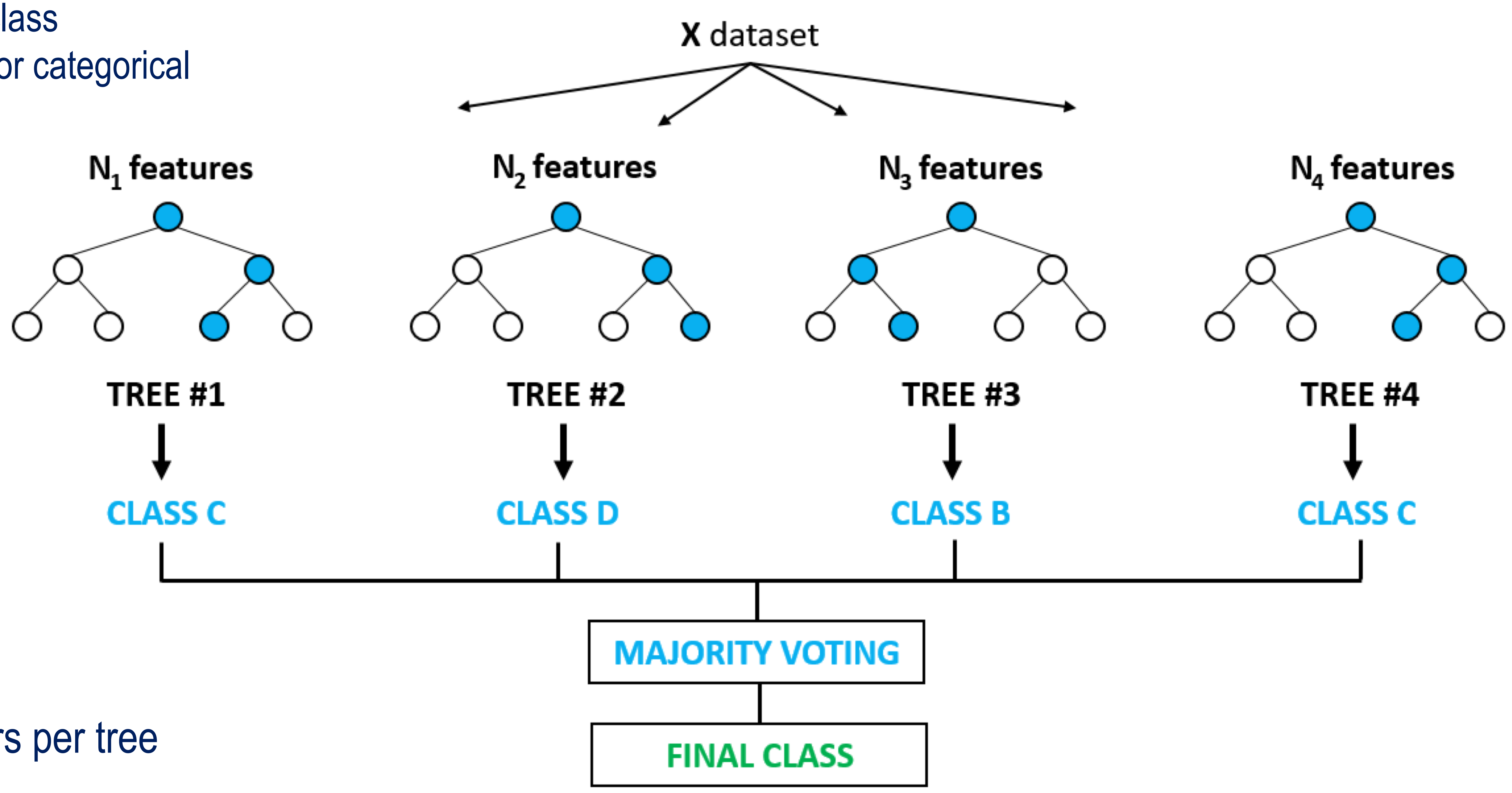




- Decision node – feature/attribute test and rule (numerical or categorical)
- Leaf node – outcome (class in classification)

# RANDOM FOREST

Problem set-up:

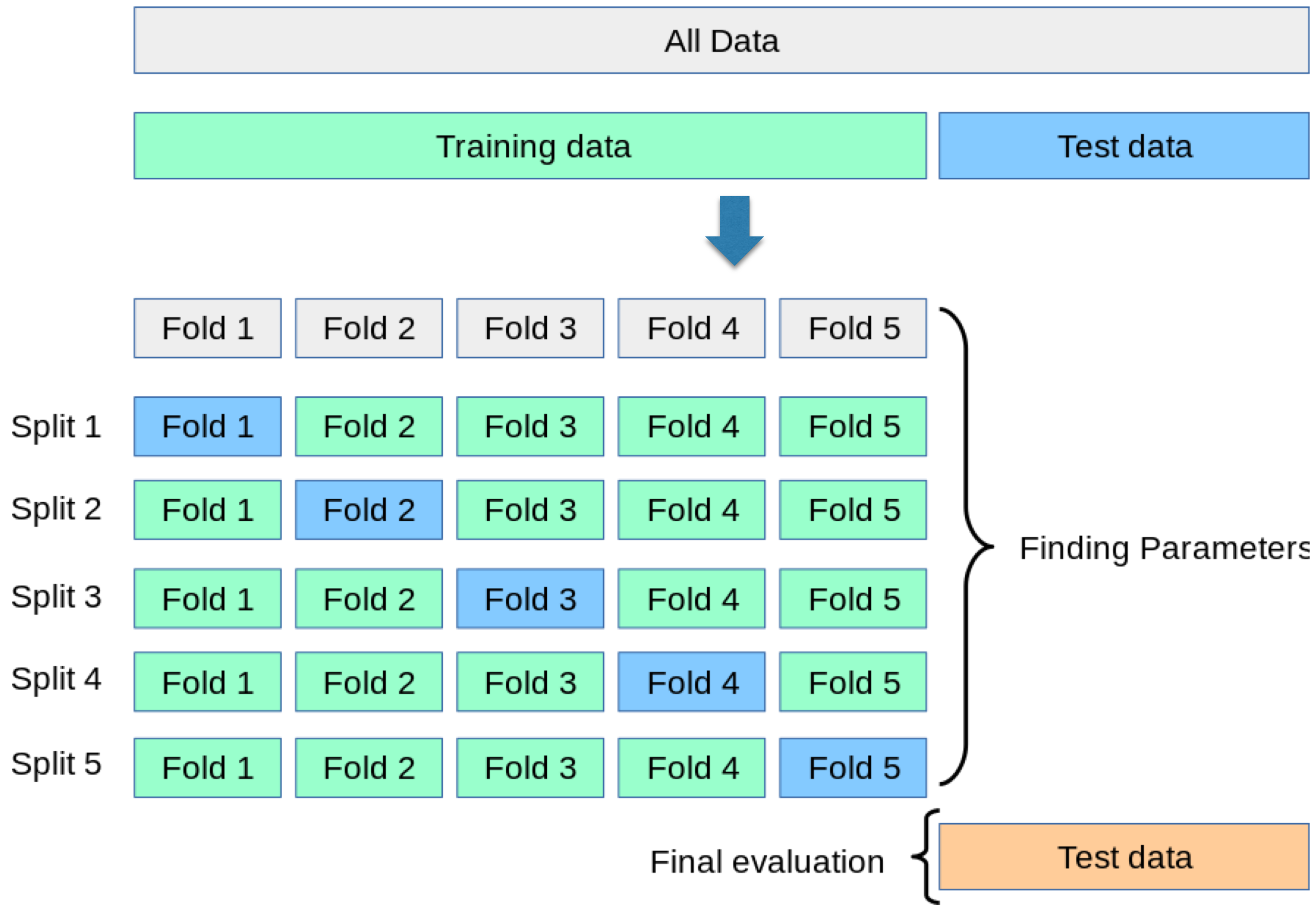
- Target variable binary class / multi-class
- Predictors (x1,x,2,x3 ) – numerical or categorical

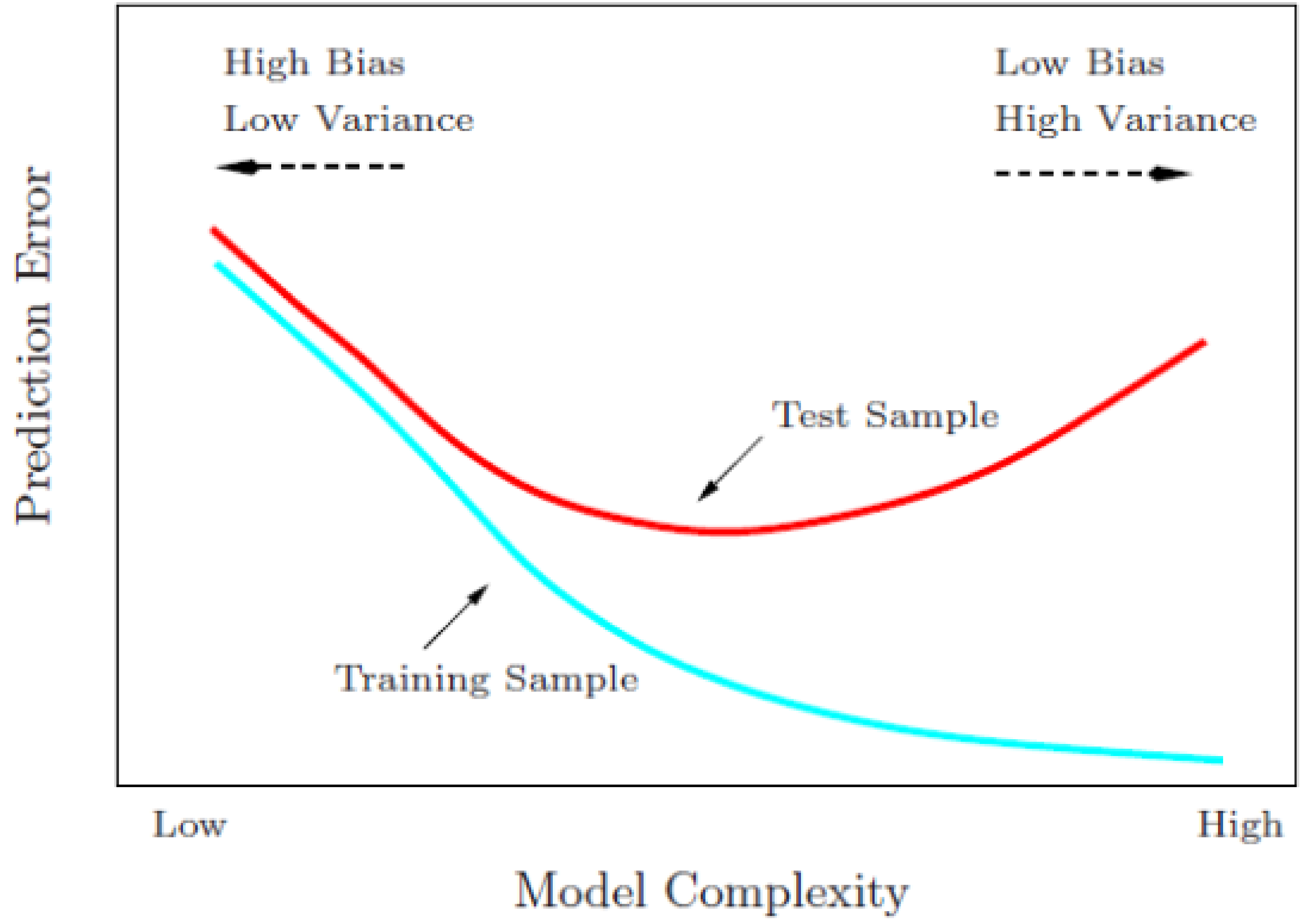



- Bootsrap sampling of examples
- Subsample of features/predictors per tree
- Voting for the result

# MODEL TRAINING AND TESTING

## Same rules apply as for regression algorithms



# CLASSIFICATION EVALUATION

## Quality metrics

| | | Actual: Positive | Actual: Negative |
|---|---|---|---|
| Predicted | Yes (+) | True positives TP | False Positives FP |
| | No (-) | False Negatives FN | True negatives TN |

True positive = Predict event and event happens

True negative = Predict event does not happen, nothing happens

False positive = Predict event and event does not happen (false alarm)

False negative = Fail to predict event that does happen (missed alarm)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

# CLASSIER EVALUATION

## Majority class classifier

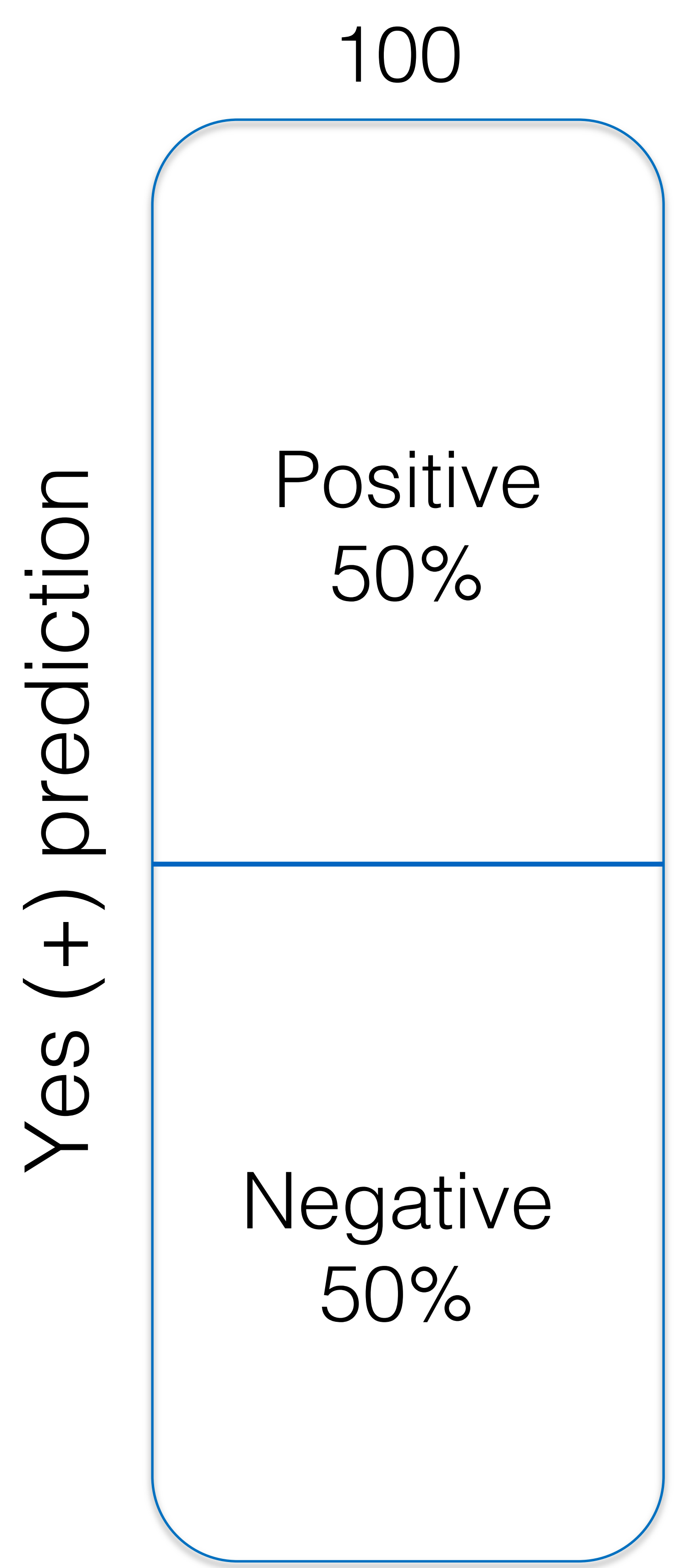
“Majority class” classifier (classify all to positive)

sample → Classifier → “positive”

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

True Positive Rate: tpr = TP / P

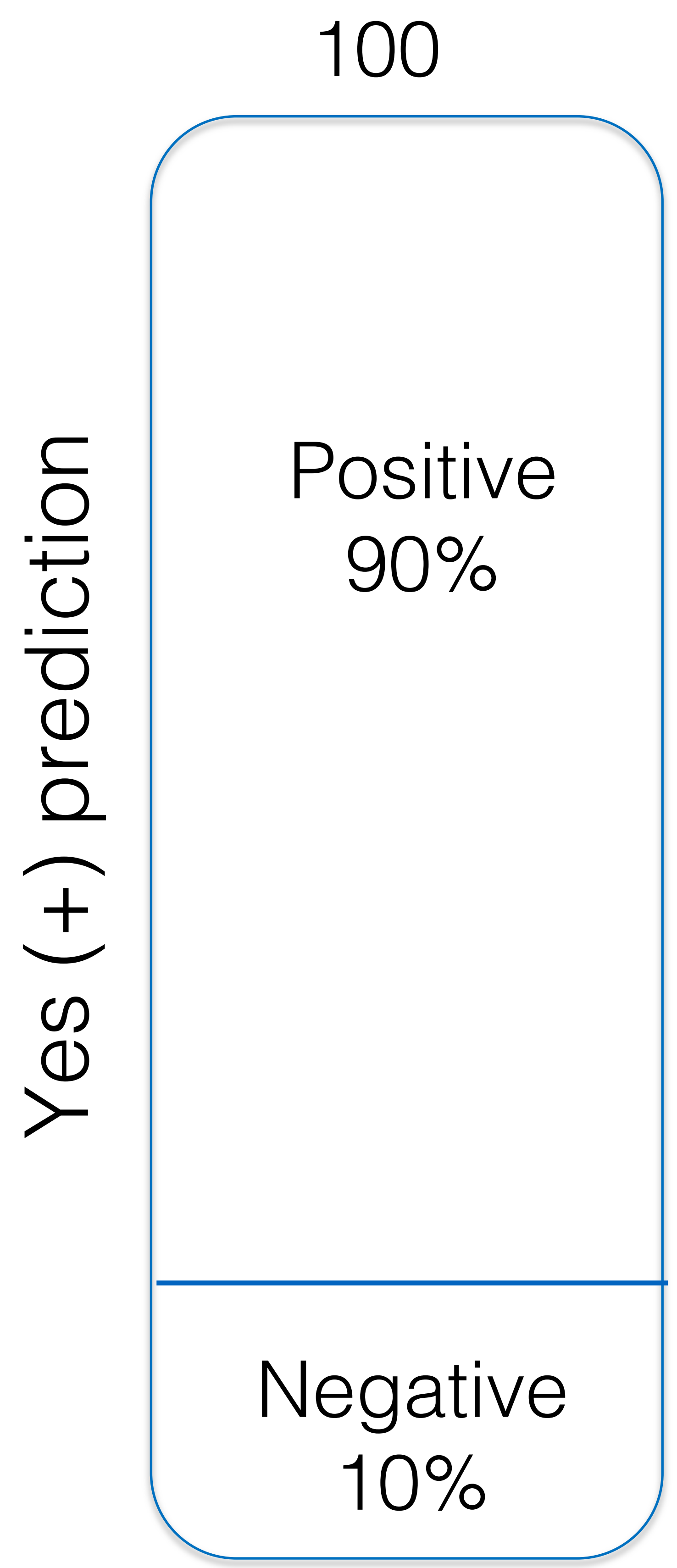
False Positive Rate: fpr = FP / N




Data set 1

Accuracy = 0.5

tpr = 1, fpr =1




Data set 2

Accuracy = 0.9

tpr = 1, fpr =1

# CLASSIFIER EVALUATION

## Normalized confusion matrix

| Predicted \ Actual | Positive | Negative |
|---|---|---|
| Yes (+) | True positives TP | False Positives FP |
| No (-) | False Negatives FN | True negatives TN |
| | P | N |

| Predicted \ Actual | Positive | Negative |
|---|---|---|
| Yes (+) | True positives rate tpr | False positives rate fpr |
| No (-) | False negatives rate fnr | True negatives rate tnr |

tpr = TP / P = TP /(TP+FN)

fpr = FP / N = FP /(FP+TN)

fnr = FN / P = 1 - tpr

tnr = TN / N = 1 - fpr

Accuracy = (TP +TN)/ (P+N) = tpr*P/(P+N) + (1-fpr)* N/(P+N)

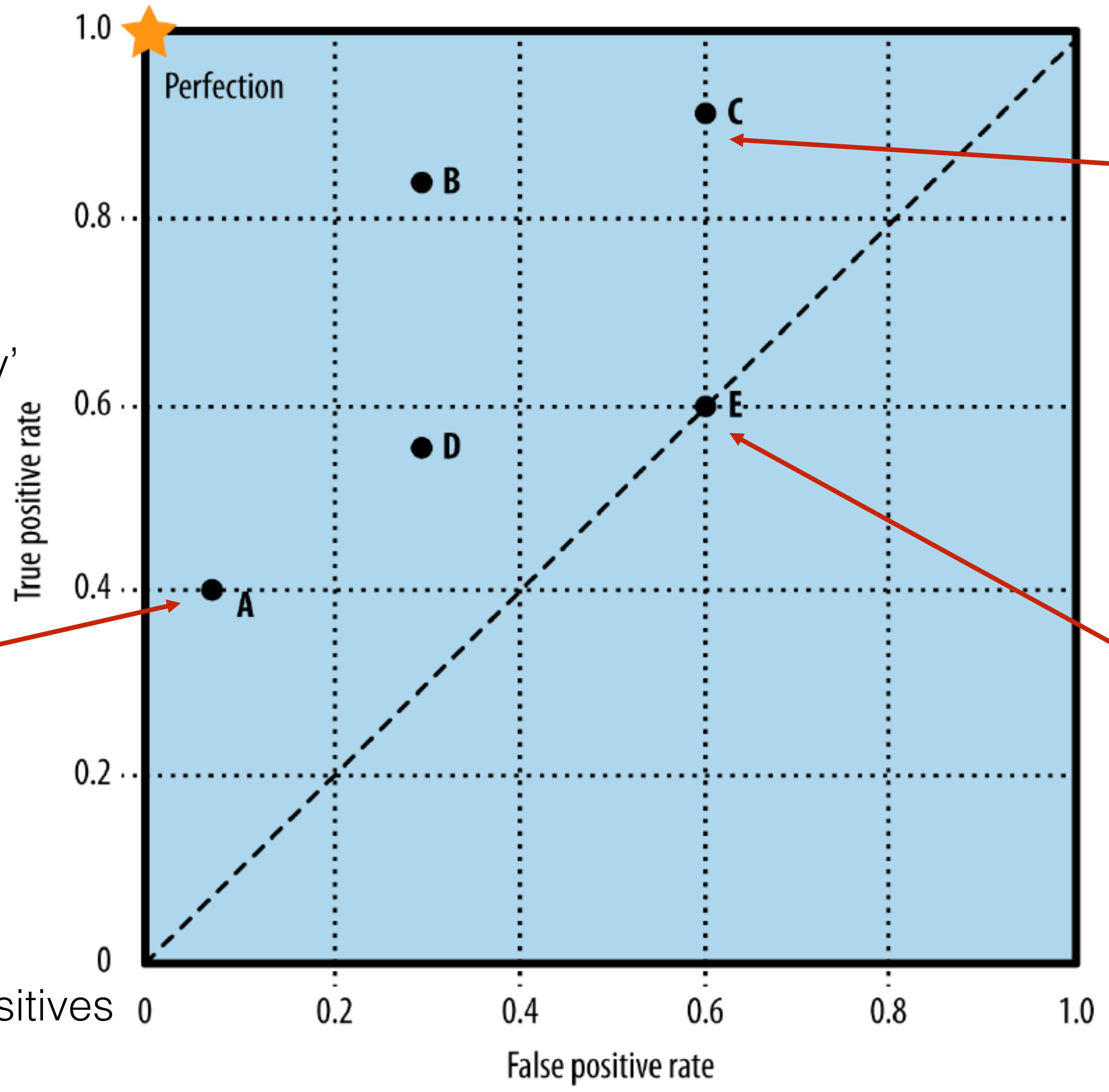
# CLASSIFIER EVALUATION

## ROC space

instance is **positive**, prediction = positive:
tpr = TP/P = 'specificity' = 'recall'
(% of positive detected)

instance is **negative**, prediction = positive:
fpr = FP/N = 'false alarm rate' = 1 – 'sensitivity'
(% of negatives wrongly reacted to)

all positives




no positives

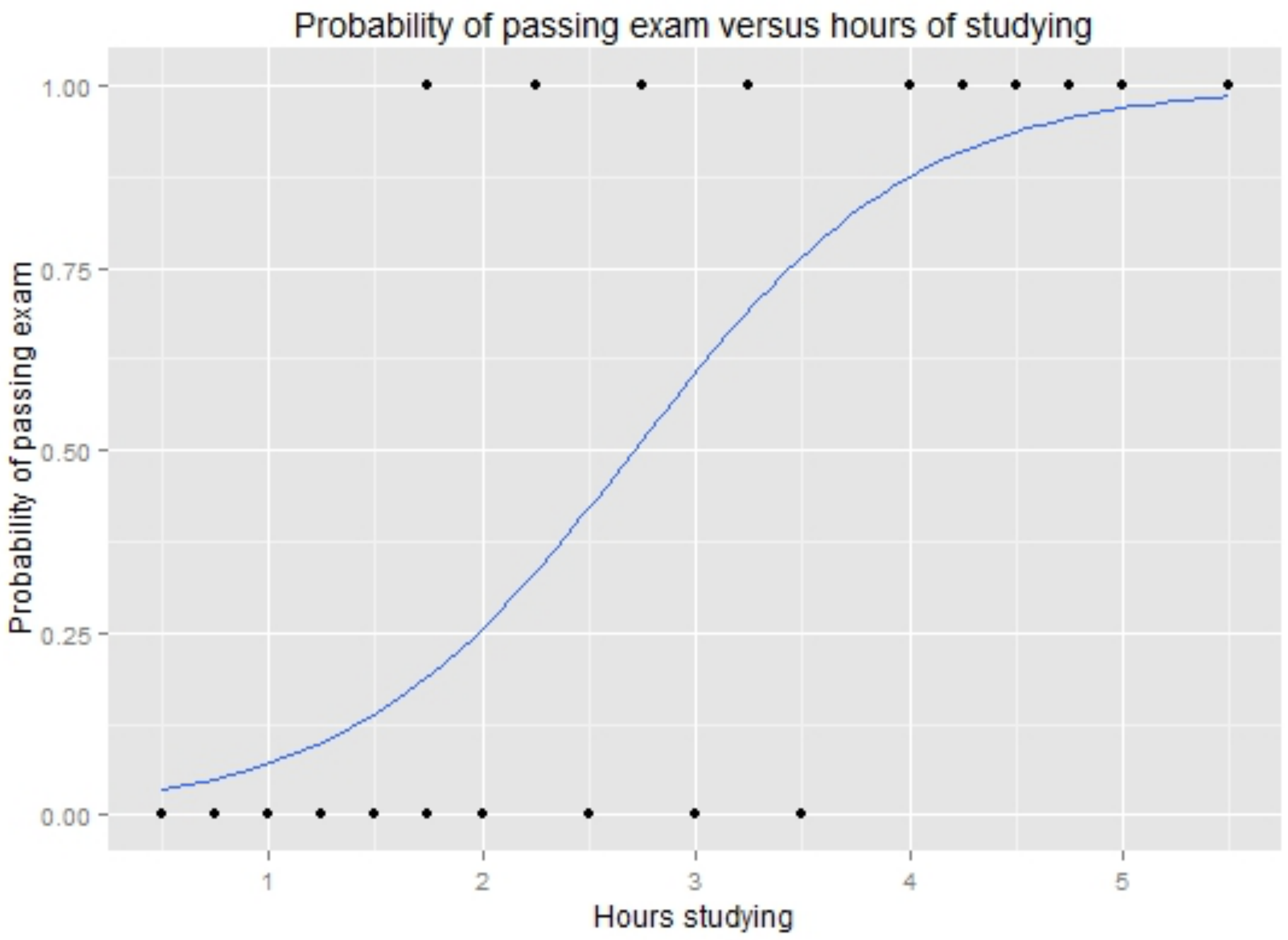
| | P | N |
|---|---|---|
| Y | 0.9 | 0.6 |
| N | 0.1 | 0.4 |

| | P | N |
|---|---|---|
| Y | 0.4 | 0.08 |
| N | 0.6 | 0.92 |

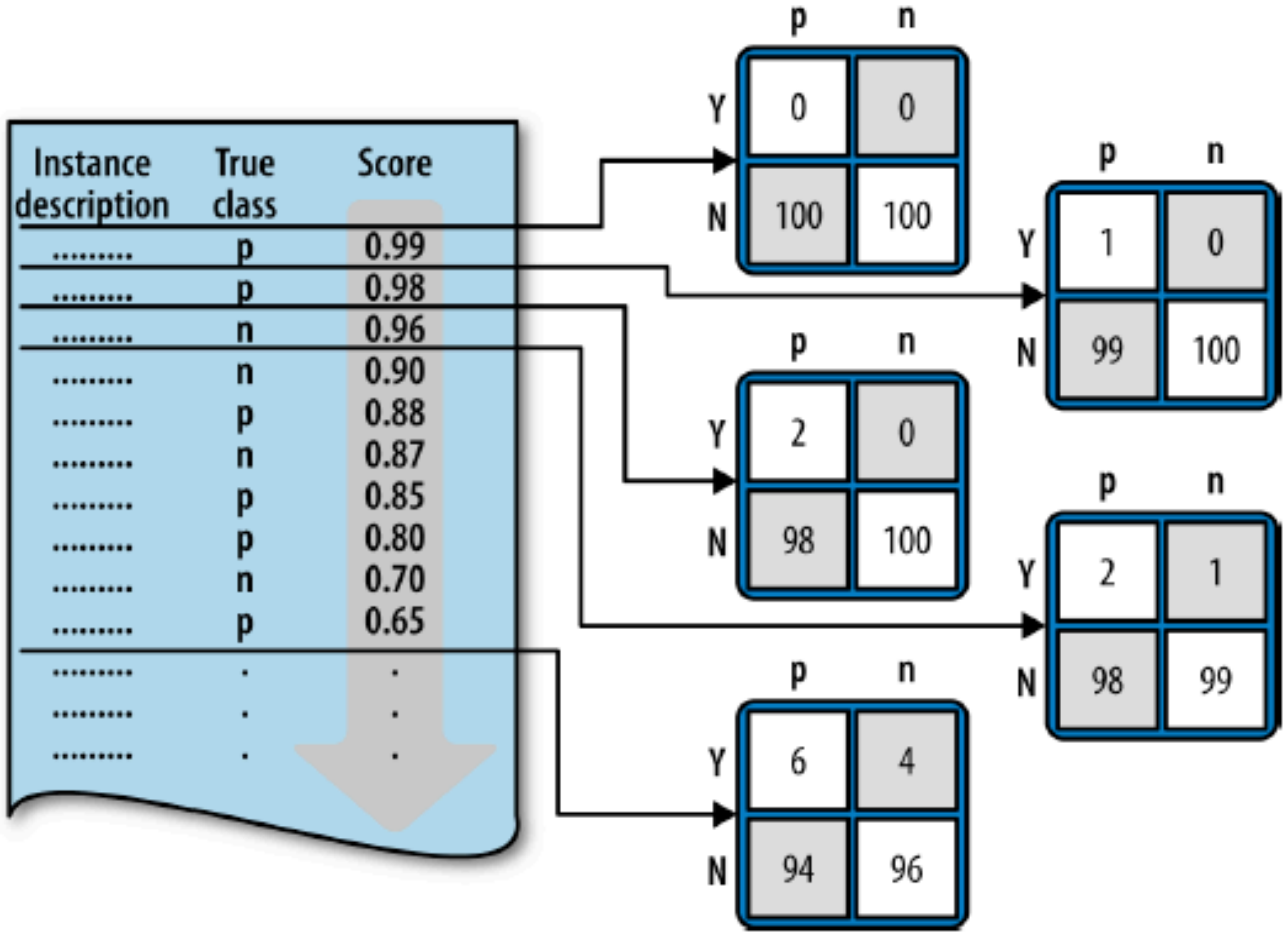
| | P | N |
|---|---|---|
| Y | 0.6 | 0.6 |
| N | 0.4 | 0.4 |

# RANKING CLASSIFIERS

## Threshold on ranking



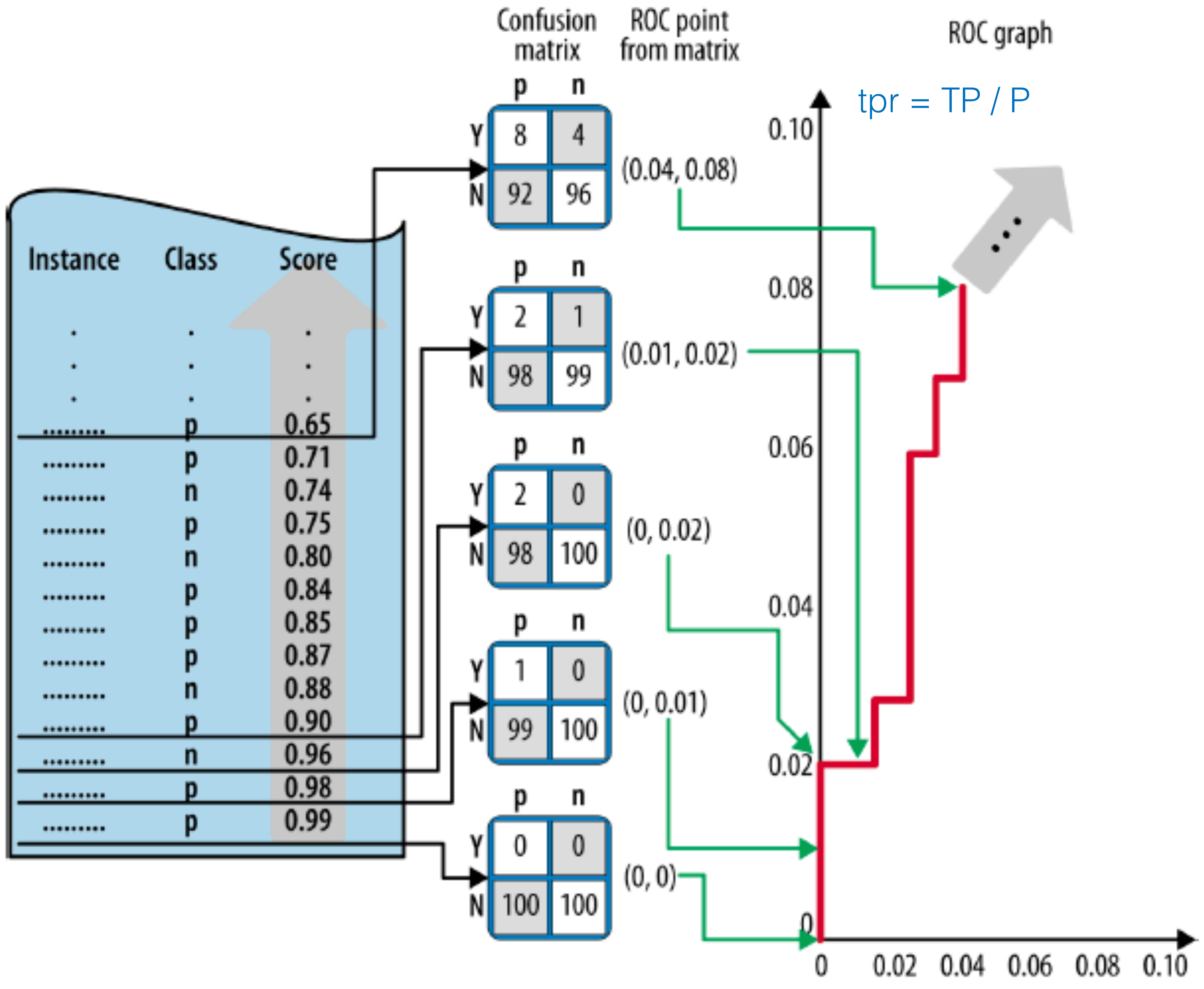

$$P(y^{(i)} = 1) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_p x_p^{(i)}))}$$



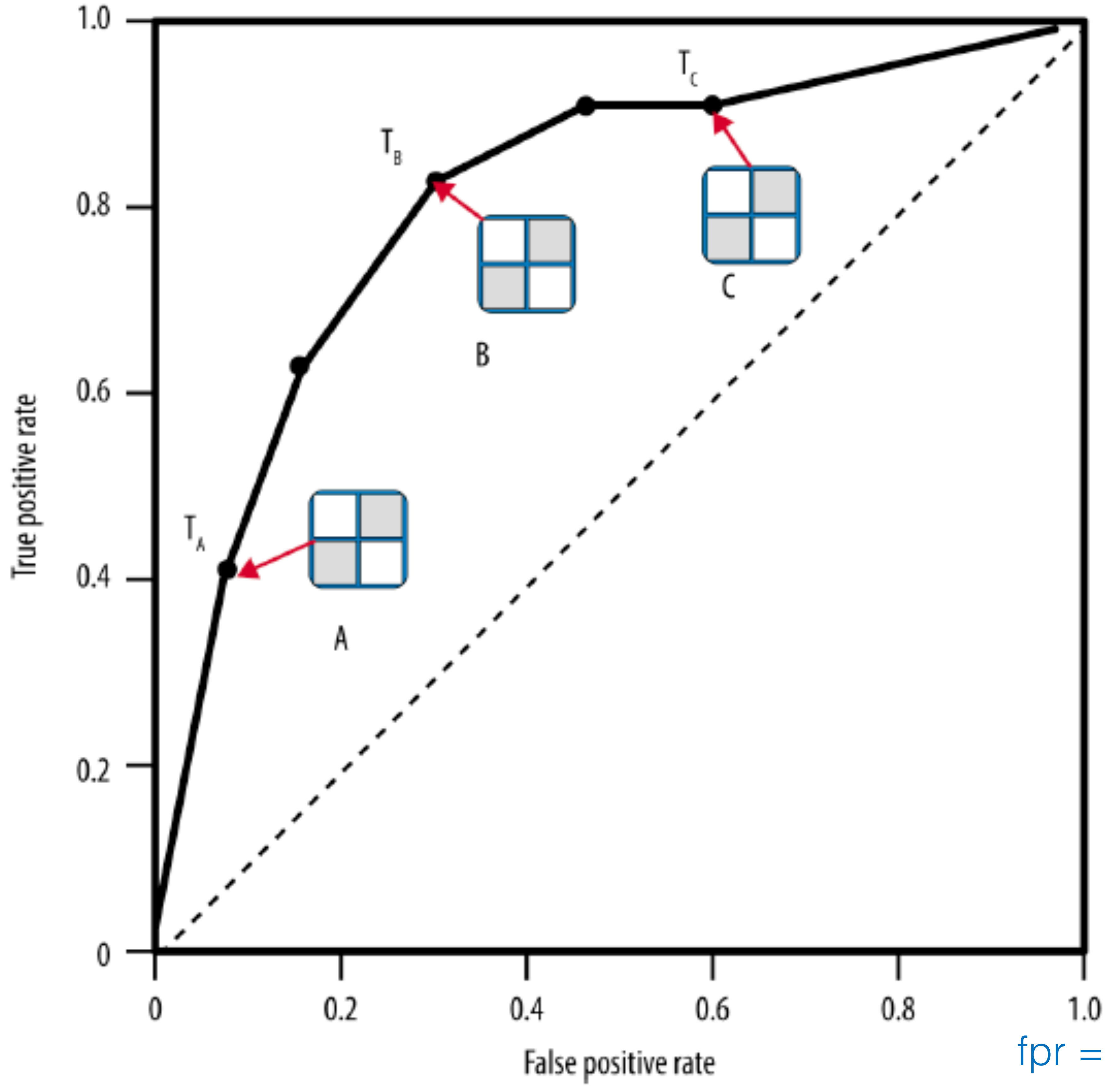

above threshold- positive class; below threshold – negative class

# RECEIVER OPERATING CURVE (ROC)

tpr = TP / P




fpr = FP / N

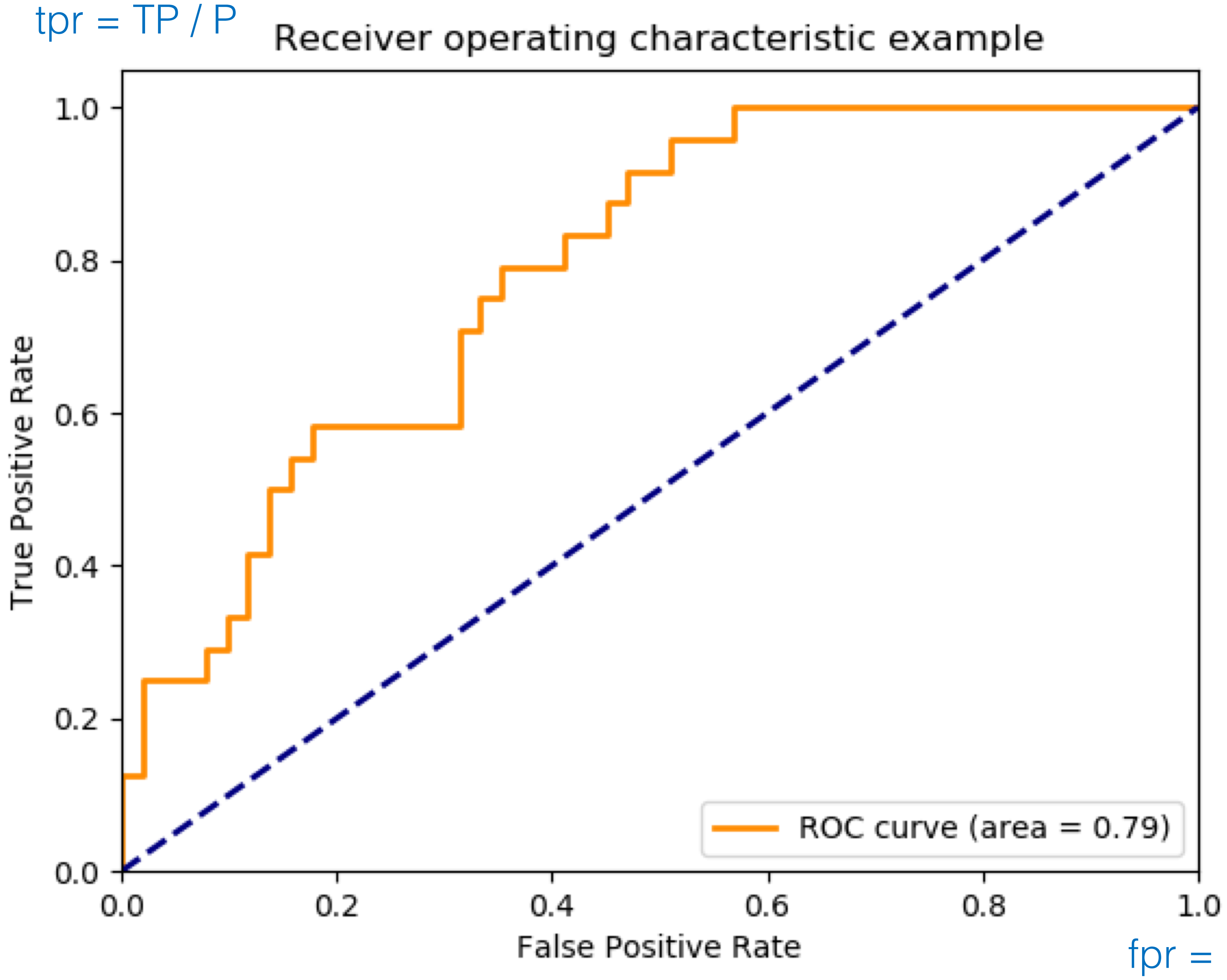



fpr = FP / N

# RECEIVER OPERATING CURVE (ROC)

## Classifier quality metric – ROC AUC

tpr = TP / P




fpr = FP / N

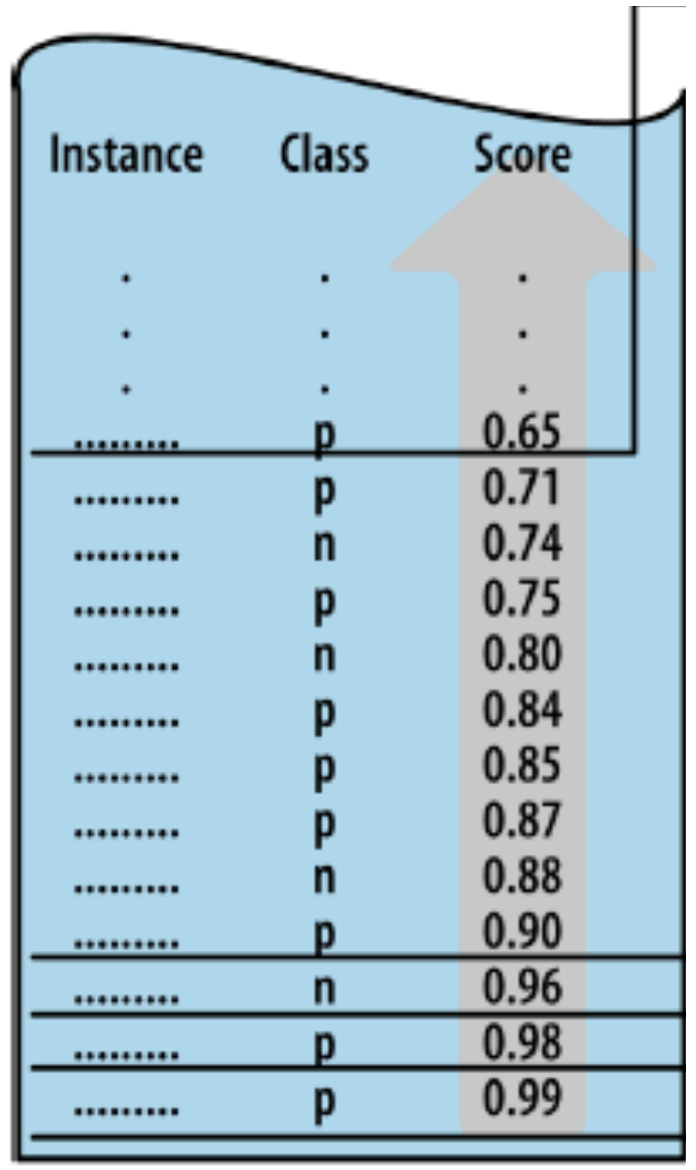
ROC AUC = area under the ROC curve

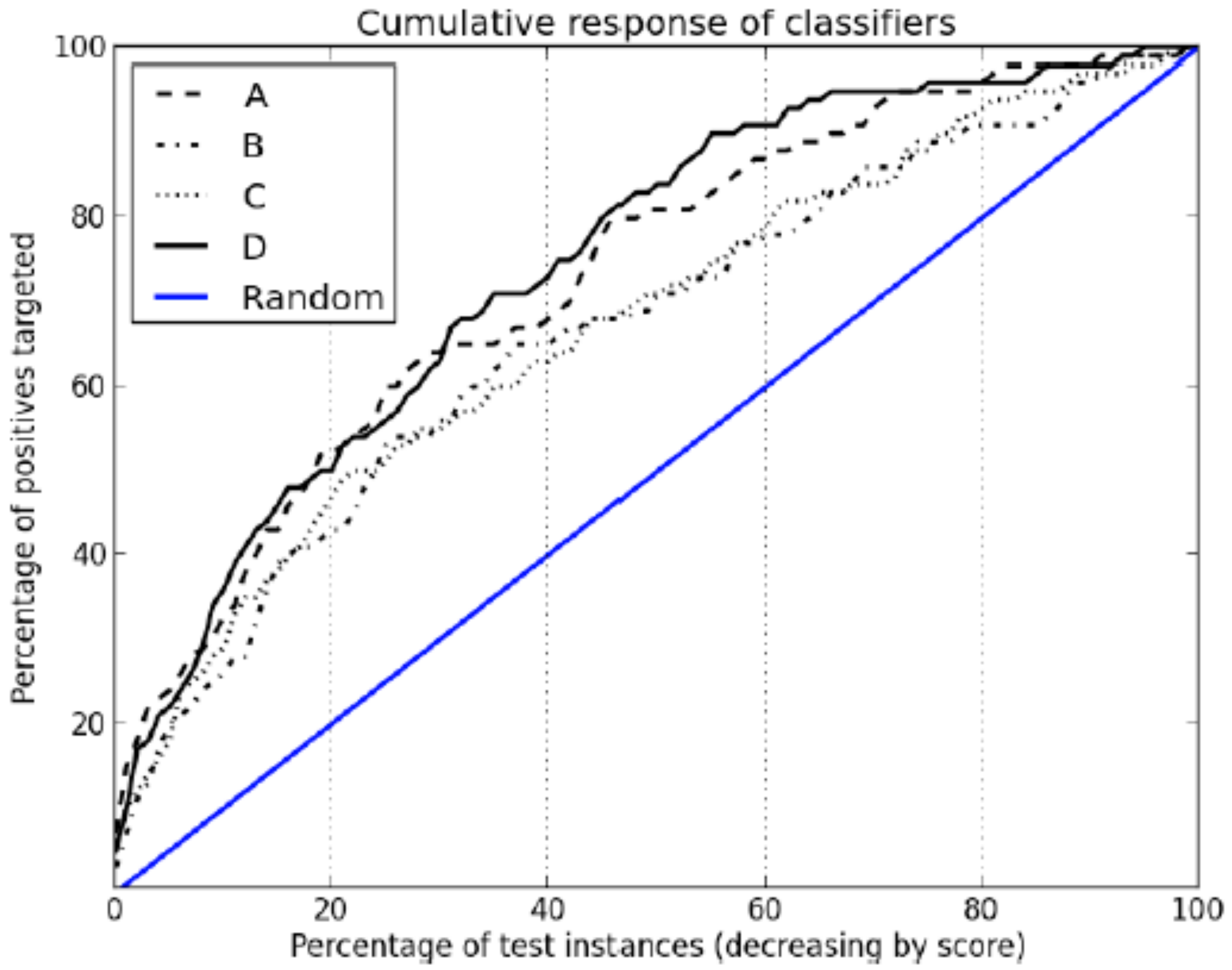
ROC AUC = 1 perfect
ROC AUC =0.5 random (bad)

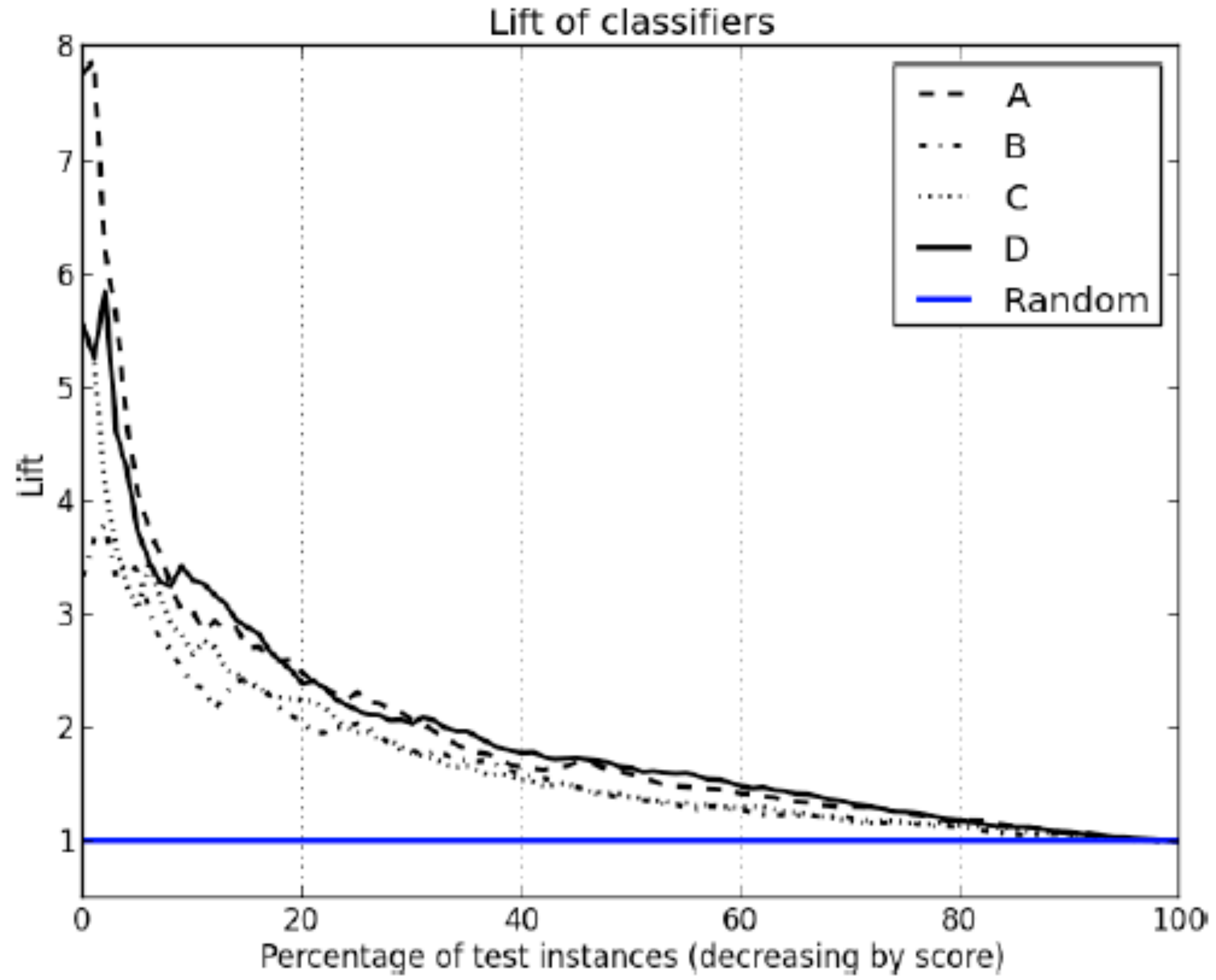
# CUMULATIVE RESPONSE AND LIFT CURVES

## Ranking classifiers

tpr = TP / P



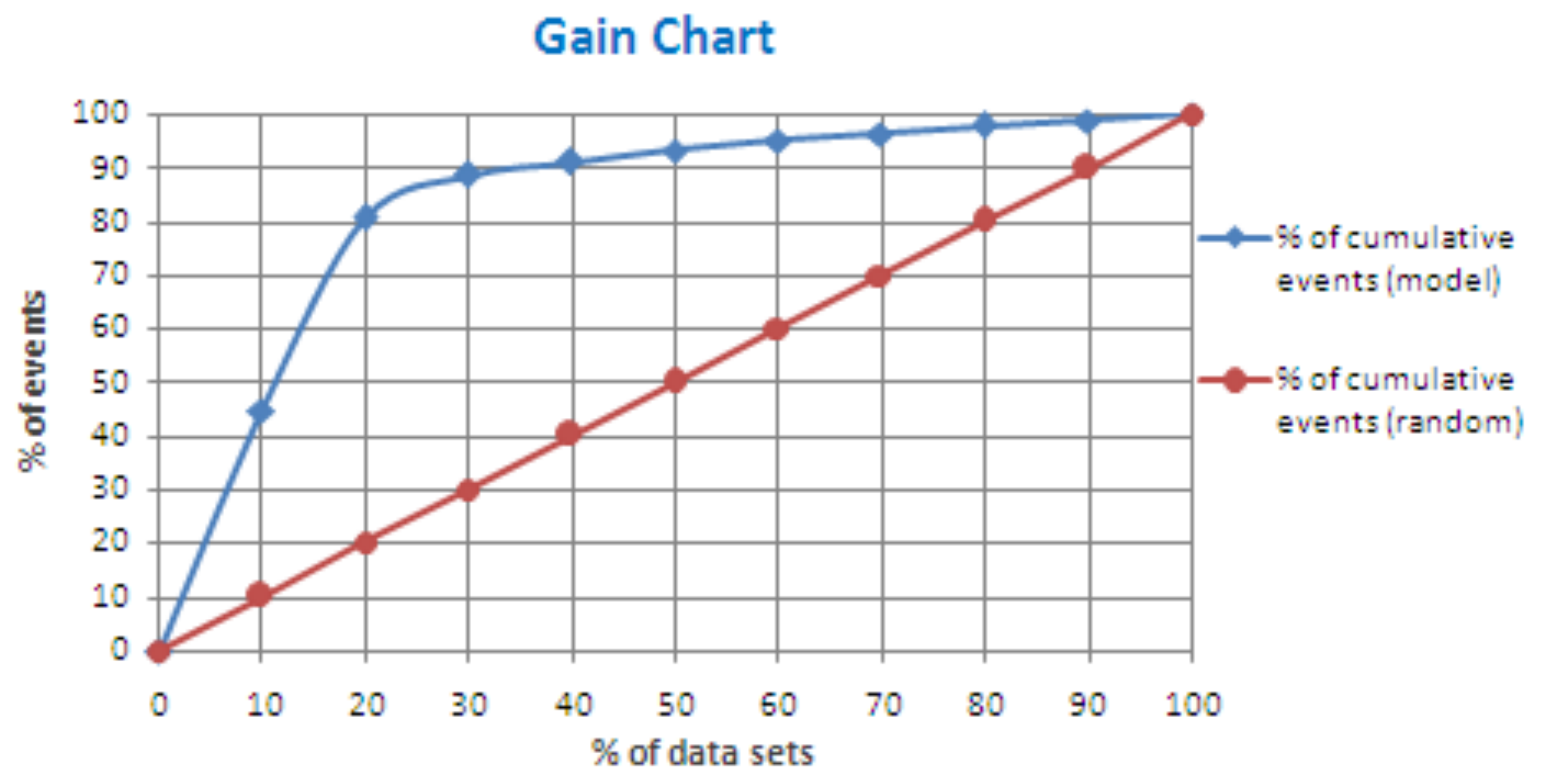



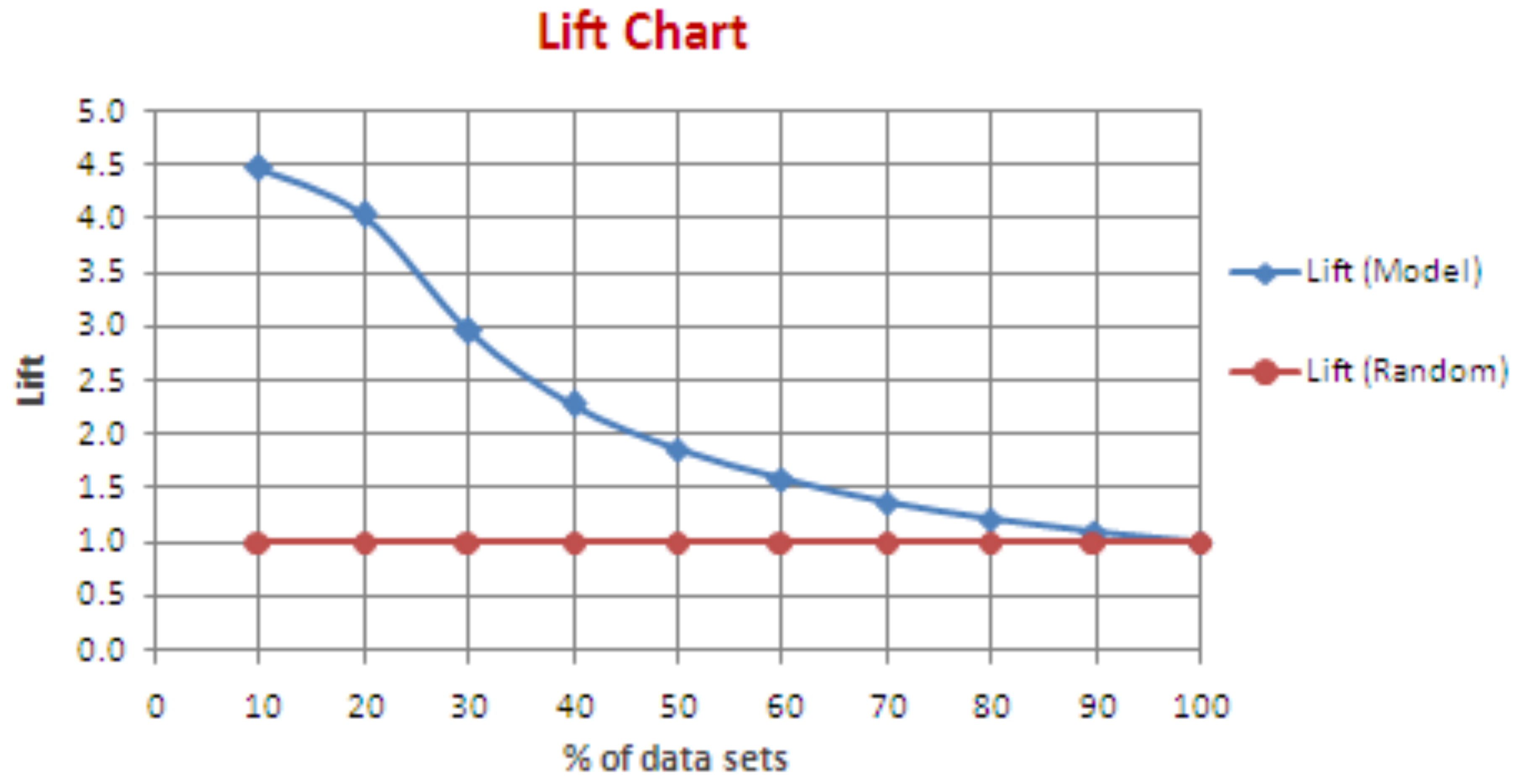

Cumulative response curves plot the percentage of positives correctly classified (tpr = TP/P) , as a function of the percentage of the instances in the data

Lift of a classifier represents the advantage it provides over random guessing.

# GAIN AND LIFT CHARTS

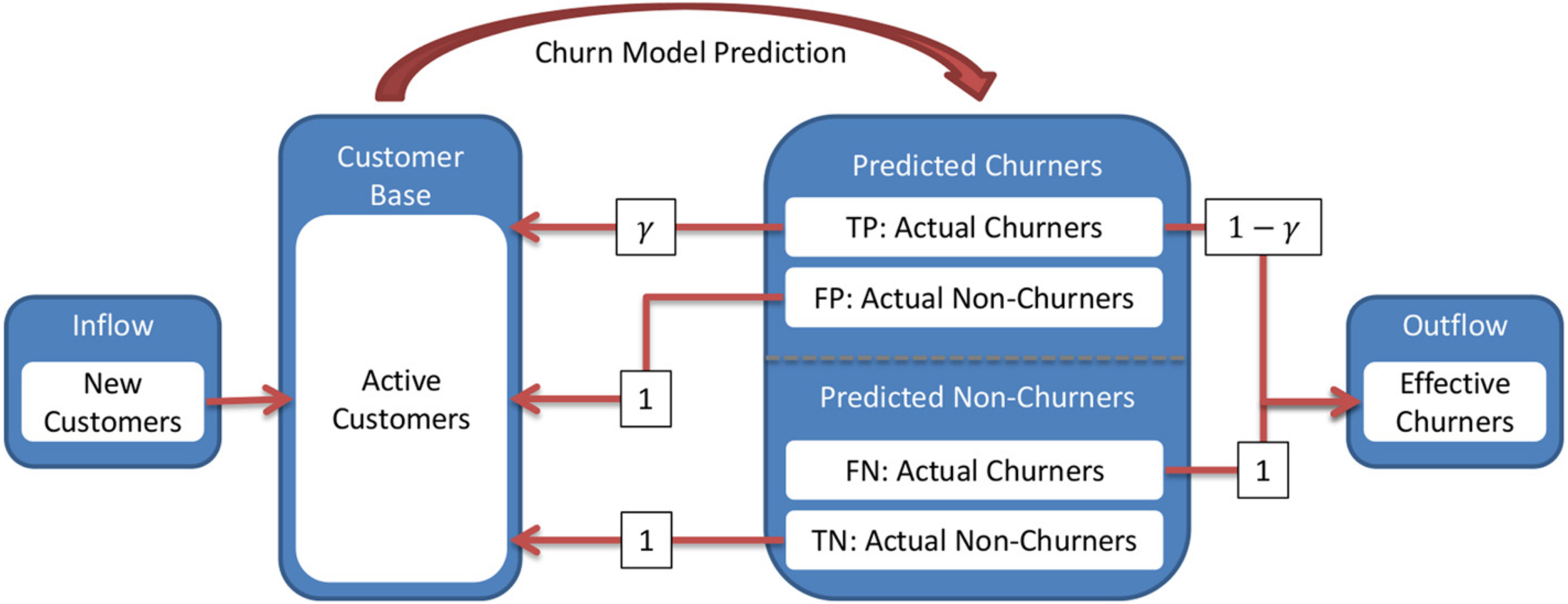
| Decile | Number of Cases (Input Values) | Number of Responses (Input Values) | Cumulative Responses | % of events | Gain | Cumulative Lift |
|---|---|---|---|---|---|---|
| 1 | 2500 | 2179 | 2179 | 44.71 | 44.71 | 4.47 |
| 2 | 2500 | 1753 | 3932 | 35.97 | 80.67 | 4.03 |
| 3 | 2500 | 396 | 4328 | 8.12 | 88.80 | 2.96 |
| 4 | 2500 | 111 | 4439 | 2.28 | 91.08 | 2.28 |
| 5 | 2500 | 110 | 4549 | 2.26 | 93.33 | 1.87 |
| 6 | 2500 | 85 | 4634 | 1.74 | 95.08 | 1.58 |
| 7 | 2500 | 67 | 4701 | 1.37 | 96.45 | 1.38 |
| 8 | 2500 | 69 | 4770 | 1.42 | 97.87 | 1.22 |
| 9 | 2500 | 49 | 4819 | 1.01 | 98.87 | 1.10 |
| 10 | 2500 | 55 | 4874 | 1.13 | 100.00 | 1.00 |
| | 25000 | 4874 | | | | |



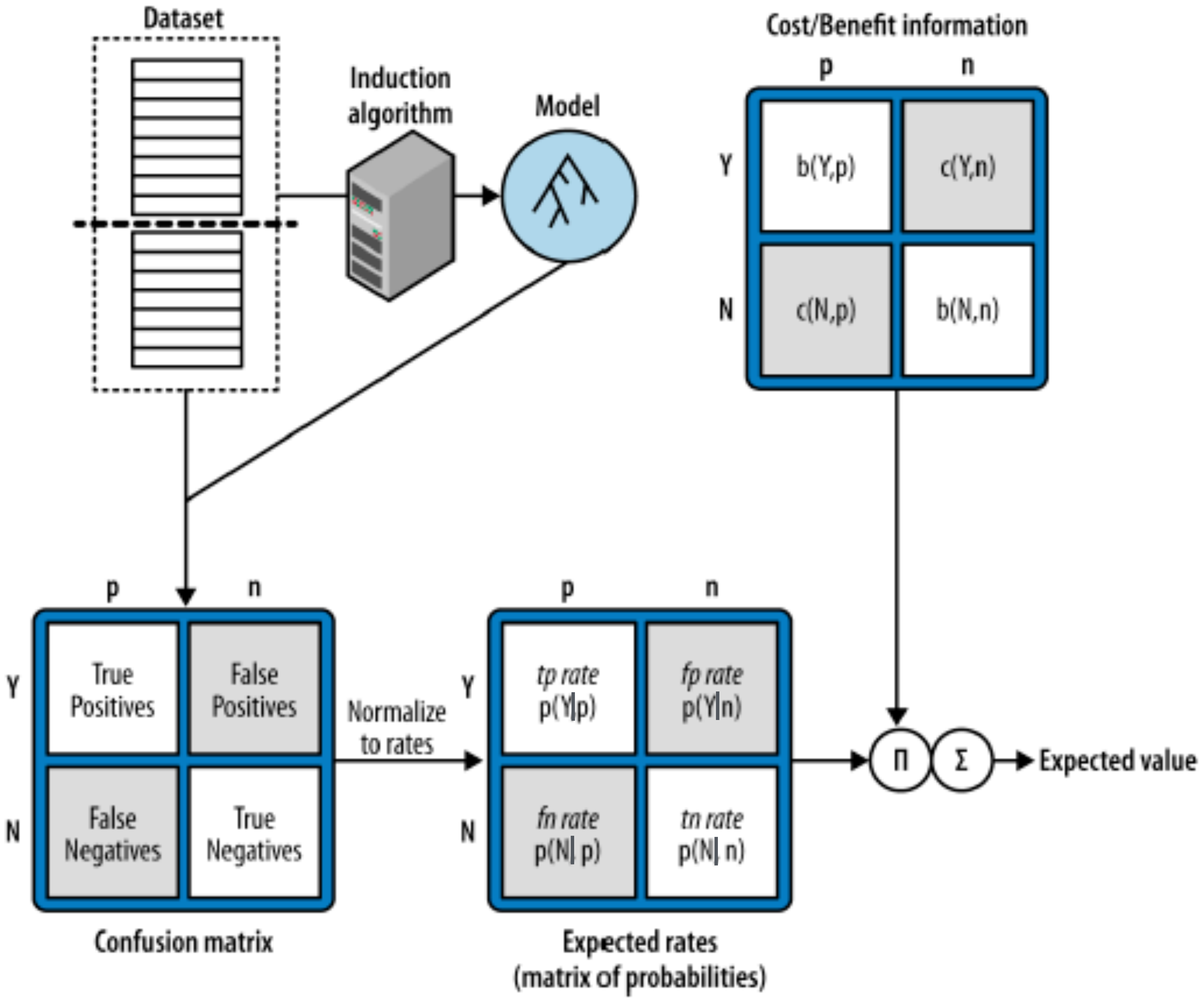

- Gain = cumulative response

# CUSTOMER CHURN MODELING



# COST-SENSITIVE LEARNING

## Cost-benefit analysis




$$EV = p(o_1) \cdot v(o_1) + p(o_2) \cdot v(o_2) + p(o_3) \cdot v(o_3) \ldots$$

$$\text{Expected profit} = p(\mathbf{p}) \cdot [p(\mathbf{Y} \mid \mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot c(\mathbf{N}, \mathbf{p})] + p(\mathbf{n}) \cdot [p(\mathbf{N} \mid \mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y} \mid \mathbf{n}) \cdot c(\mathbf{Y}, \mathbf{n})]$$

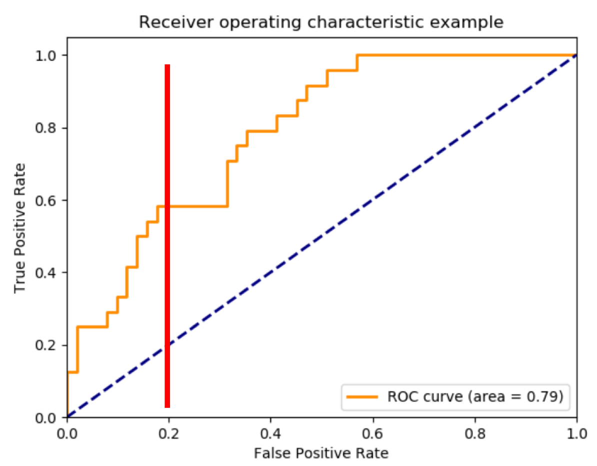
$p(Y|p) = tpr$
$p(Y|n) = fpr$
$p(N|p) = fnr$
$p(N|n) = tnr$

$p(p) = P/(P+N)$
$p(n) = N/(P+N)$

School of Data Analysis and Artificial Intelligence
Department of Computer Science

# PROFIT CURVES

## Profit from churn management campaign

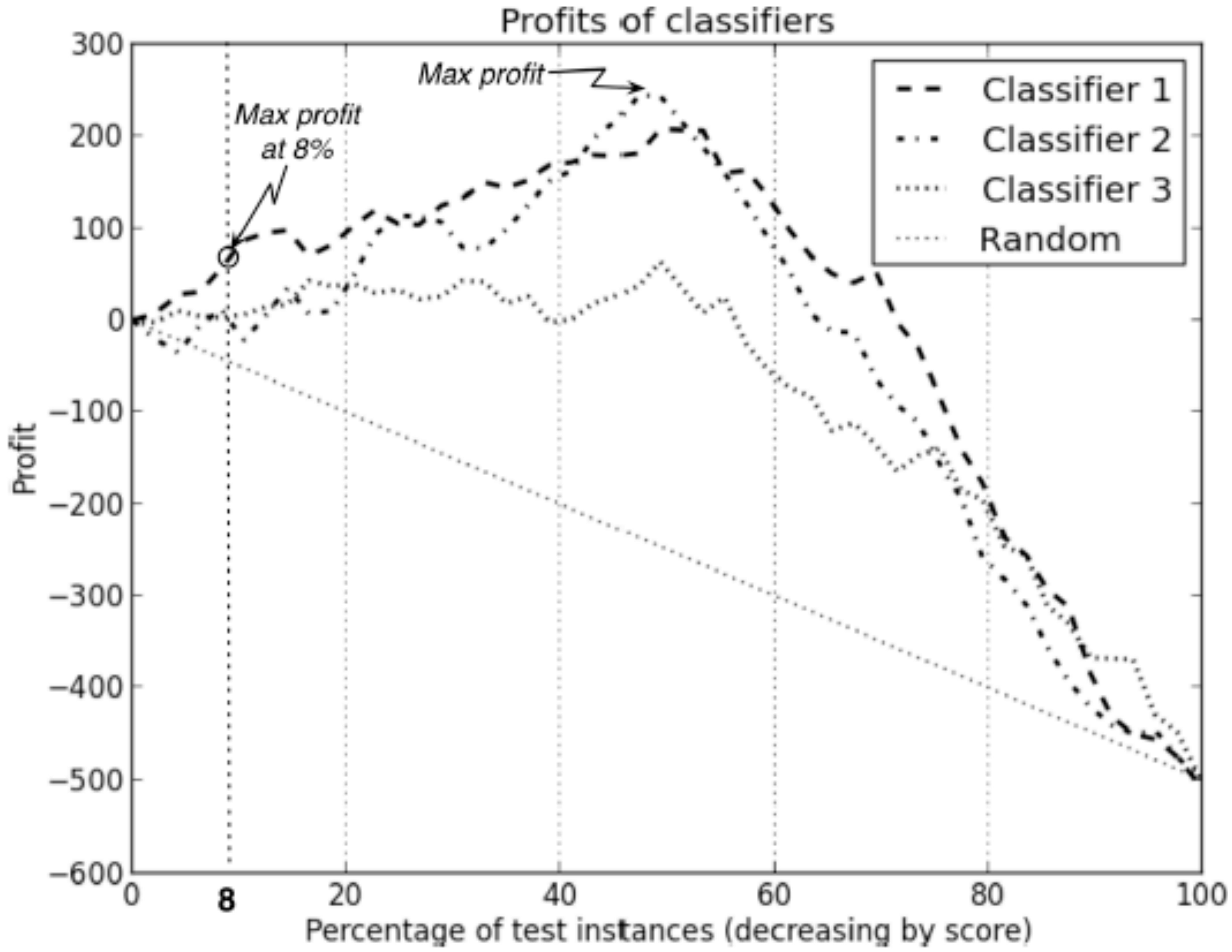



| | P | N |
|---|---|---|
| Y | 0.92 | 0.14 |
| N | 0.08 | 0.86 |

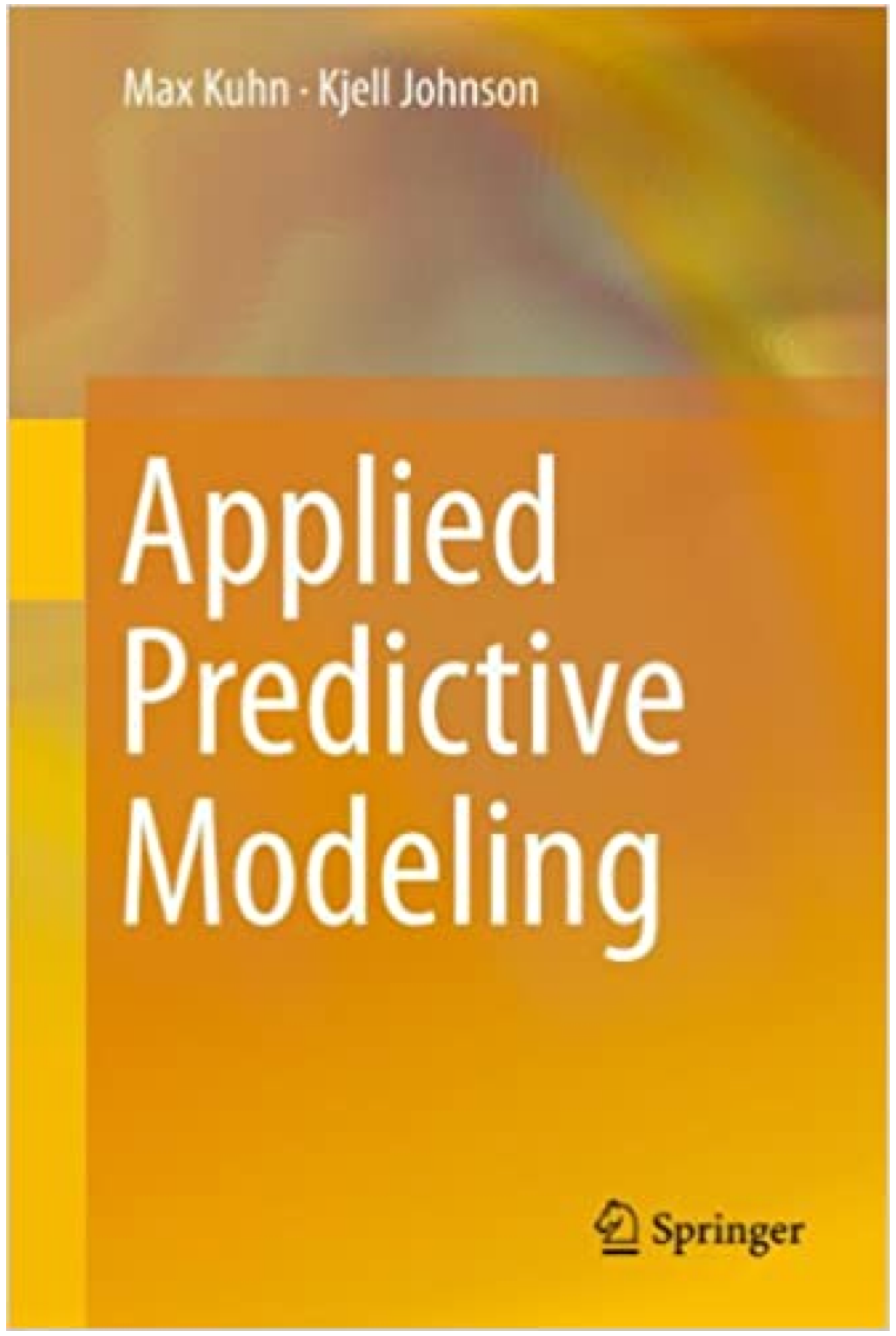
| | P | N |
|---|---|---|
| Y | $90 | -$10 |
| N | $0 | $0 |

$\Sigma$

$44.7



# ONE MORE BOOK



NATIONAL RESEARCH
UNIVERSITY