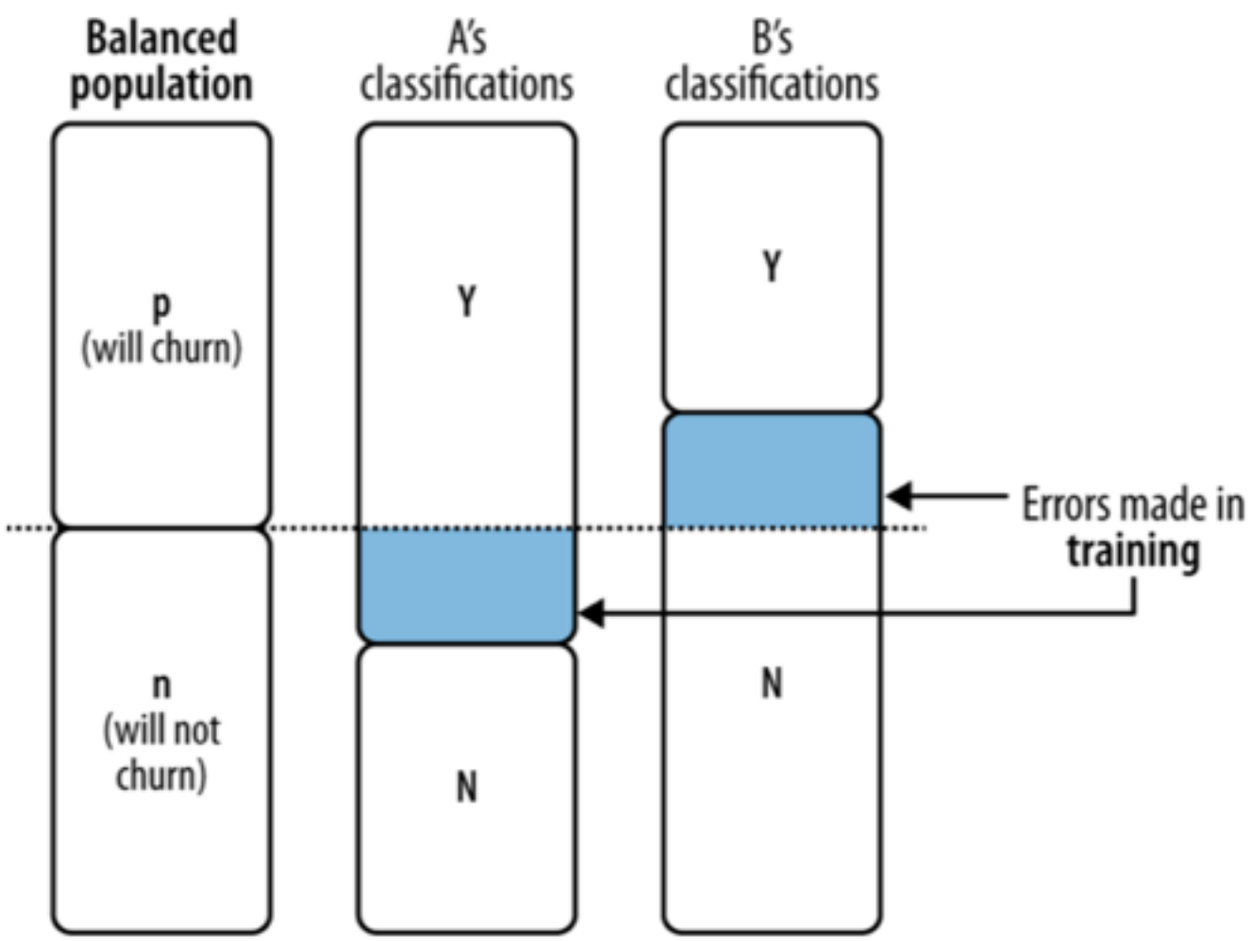
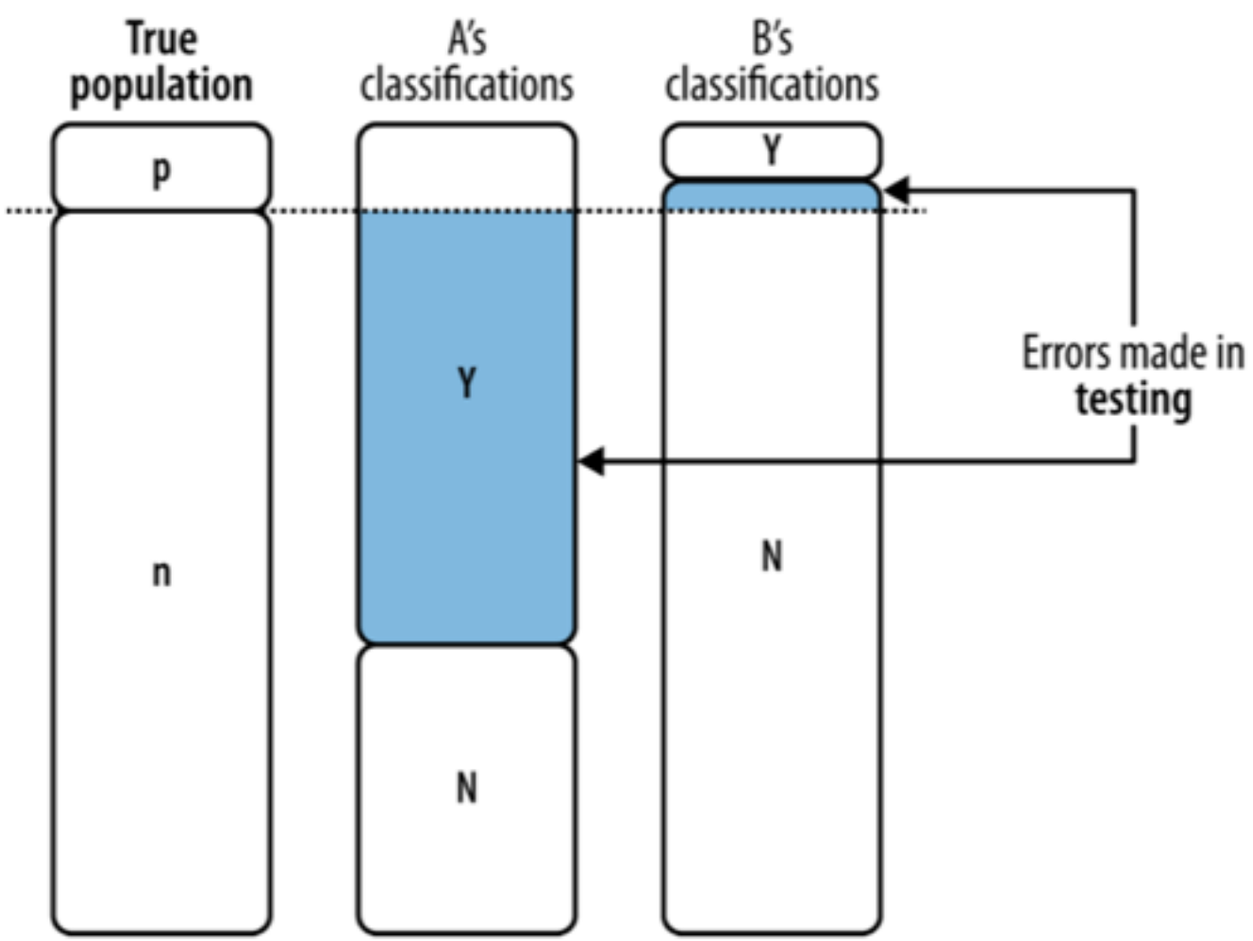
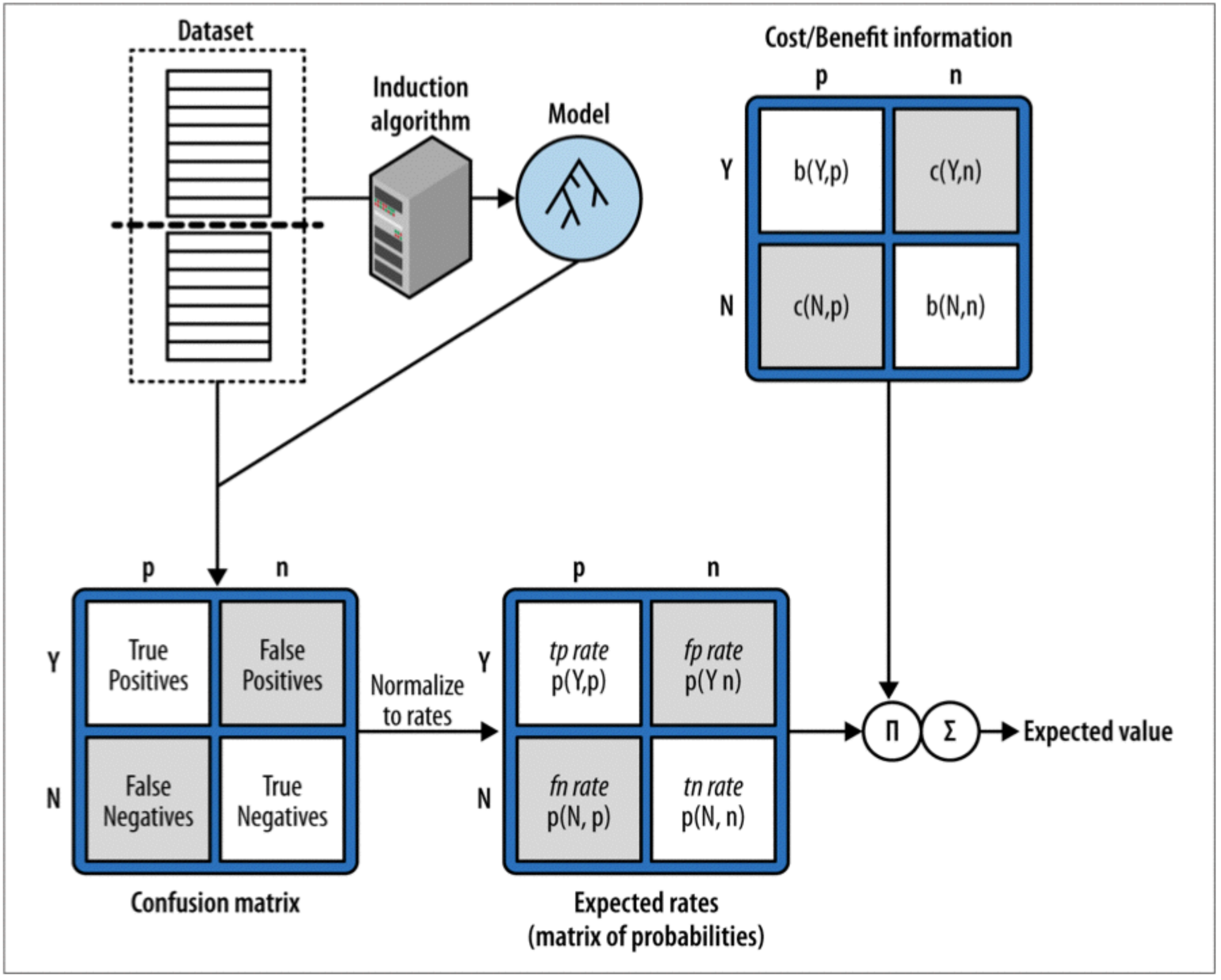
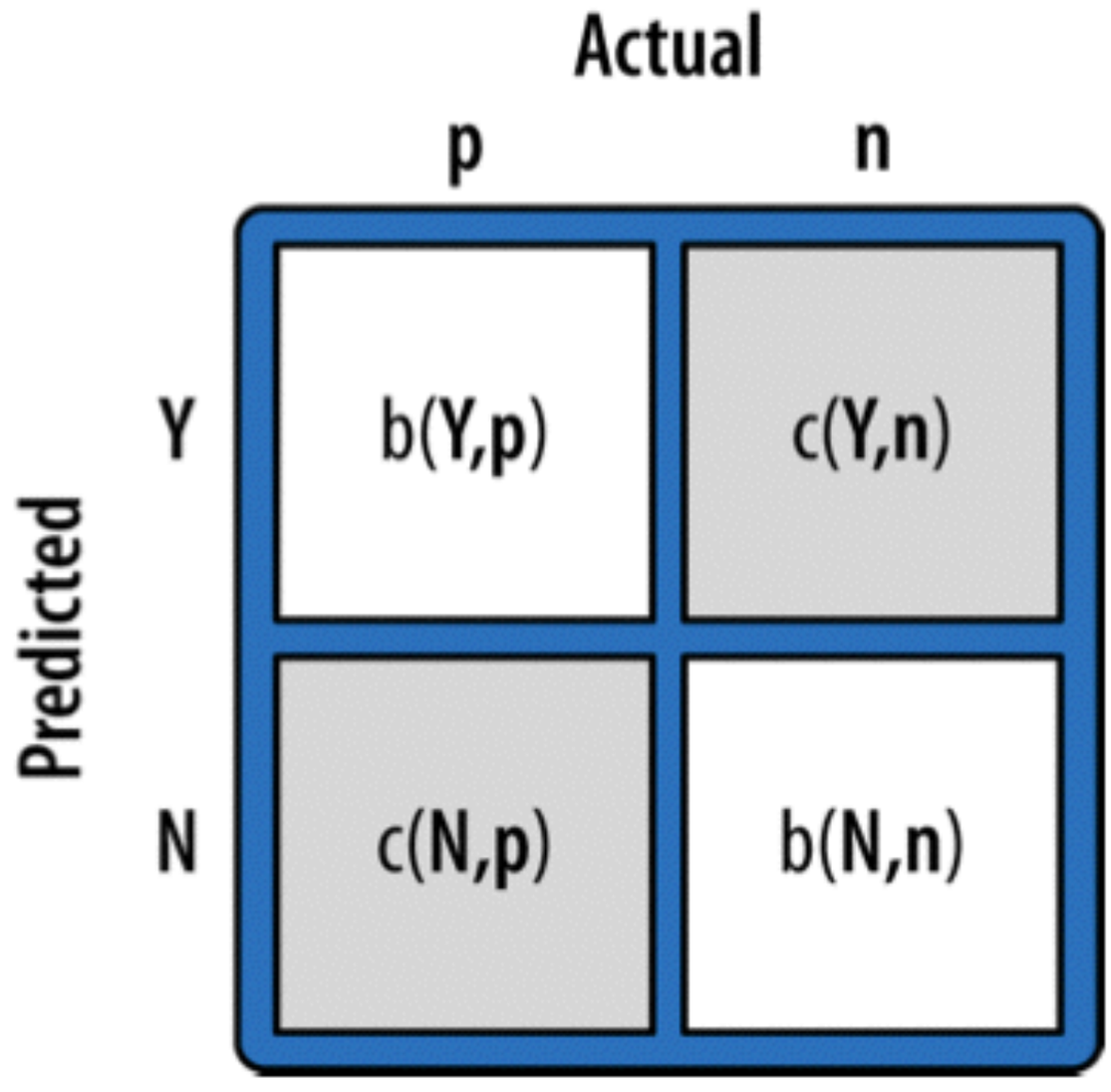

Balanced population
A's classifications
B's classifications
p
(will churn)
n
(will not churn)
Y
N
Y
N
Errors made in training

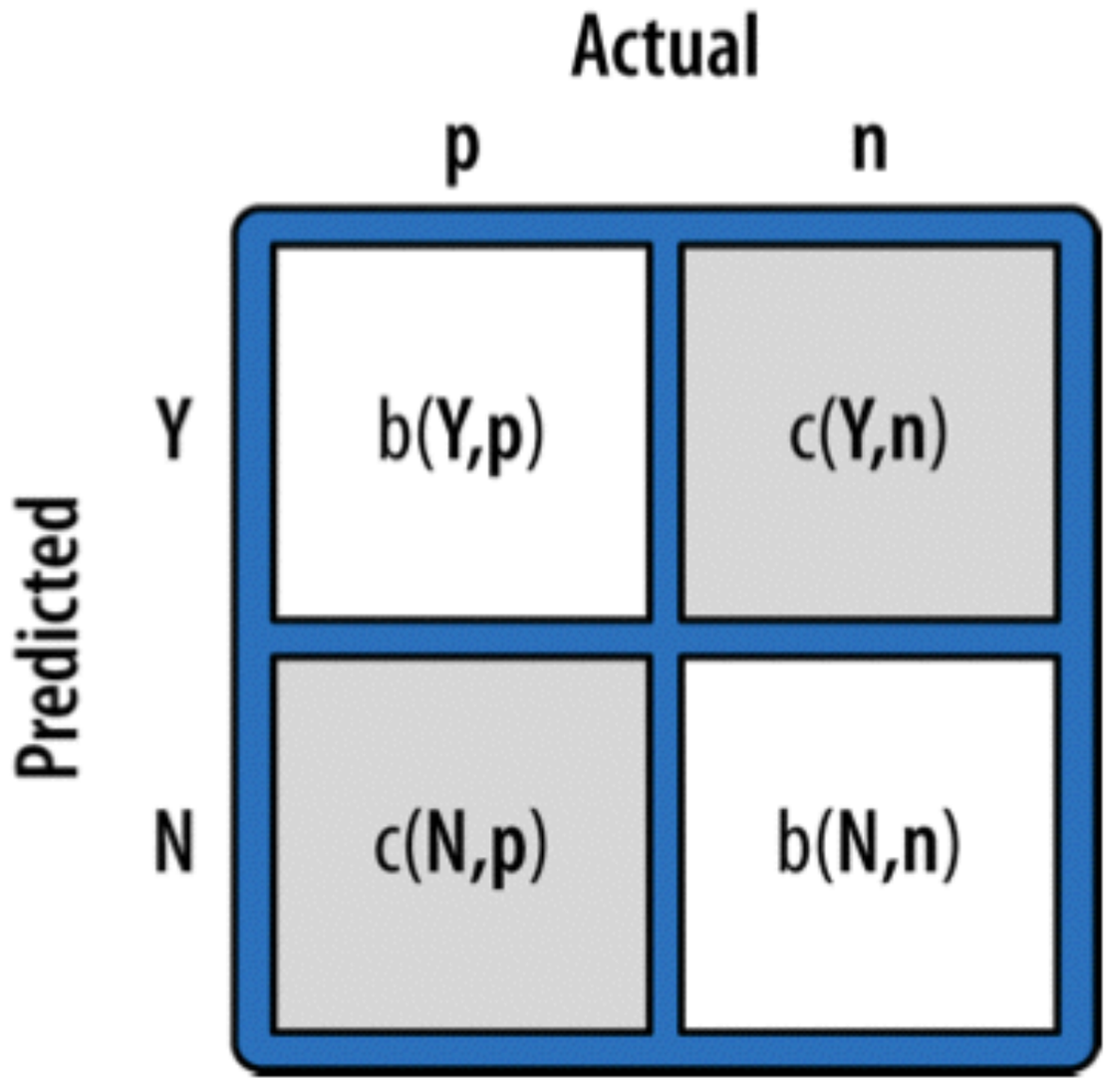

True population
A's classifications
B's classifications
p
n
Y
N
Y
N
Errors made in testing


Dataset
Induction algorithm
Model
Cost/Benefit information
p
n
Y
N
b(Y,p)
c(Y,n)
c(N,p)
b(N,n)
True Positives
False Positives
False Negatives
True Negatives
Normalize to rates
tp rate p(Y,p)
fp rate p(Y n)
fn rate p(N, p)
tn rate p(N, n)
Π
Σ
Expected value
Confusion matrix
Expected rates (matrix of probabilities)

Cost-Benefit Matrix



Cost-Benefit Matrix



Actual

Predicted		p	n
	Y	99	-1
	N	0	0

$$\textit{Expected profit} = p(\mathbf{Y}, \mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N}, \mathbf{p}) \cdot b(\mathbf{N}, \mathbf{p}) + p(\mathbf{N}, \mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y}, \mathbf{n}) \cdot b(\mathbf{Y}, \mathbf{n})$$

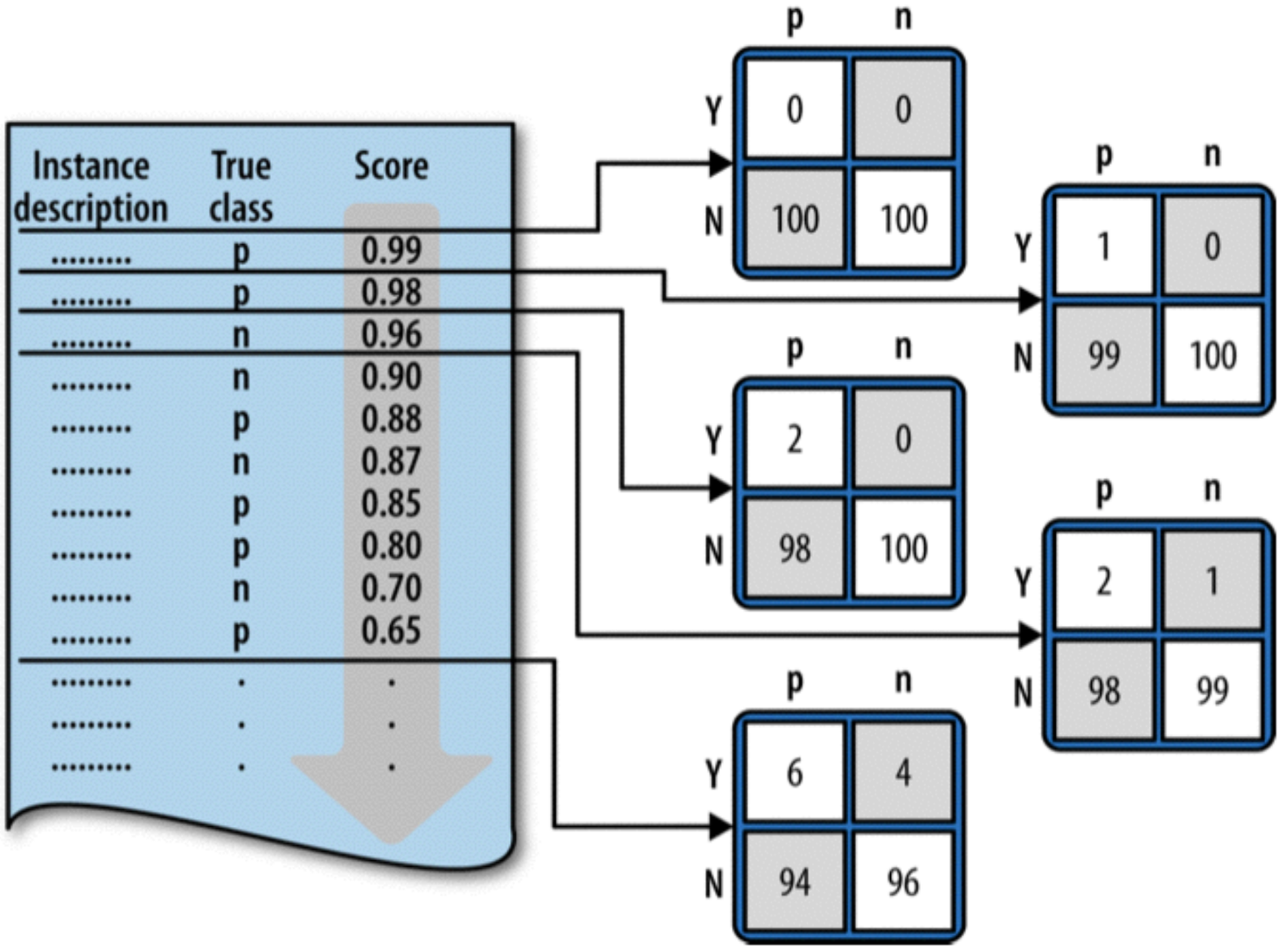
$$p(x, y) = p(y) \cdot p(x \mid y)$$

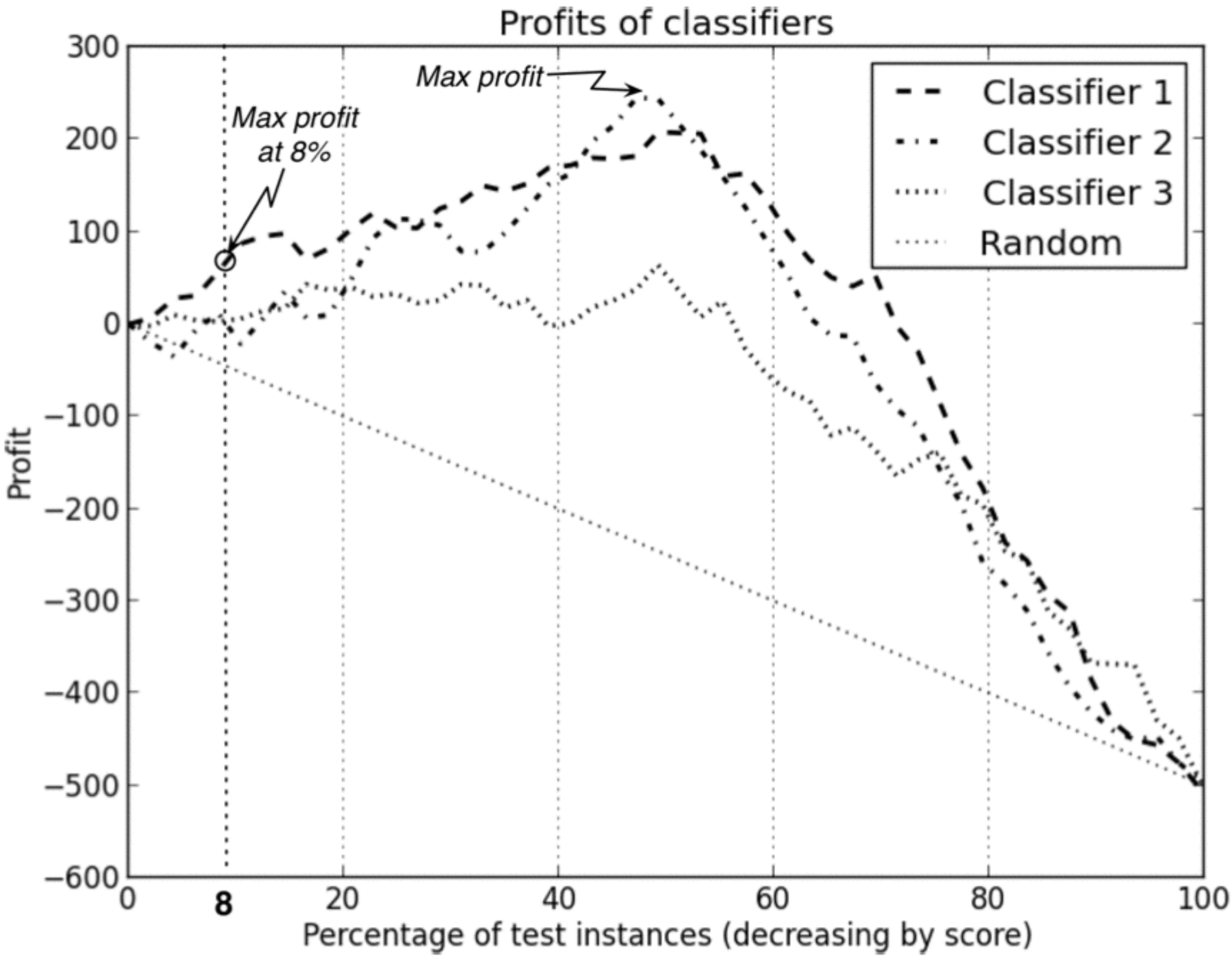
$$\begin{aligned} \text{Expected profit} = {} & p(\mathbf{Y} \mid \mathbf{p}) \cdot p(\mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot p(\mathbf{p}) \cdot b(\mathbf{N}, \mathbf{p}) + \\ & p(\mathbf{N} \mid \mathbf{n}) \cdot p(\mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y} \mid \mathbf{n}) \cdot p(\mathbf{n}) \cdot b(\mathbf{Y}, \mathbf{n}) \end{aligned}$$

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

$$\begin{array}{ccc} & & \text{Actual} \\ & & \begin{array}{cc} \mathbf{p} & \mathbf{n} \end{array} \\ \text{Predicted} & \begin{array}{c} \mathbf{Y} \\ \mathbf{N} \end{array} & \begin{pmatrix} 99 & -1 \\ 0 & 0 \end{pmatrix} \end{array}$$

$$\begin{aligned} \text{expected profit} \quad = \quad & p(\mathbf{p}) \cdot [p(\mathbf{Y} \mid \mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot c(\mathbf{N}, \mathbf{p})] + \\ & p(\mathbf{n}) \cdot [p(\mathbf{N} \mid \mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y} \mid \mathbf{p}) \cdot c(\mathbf{Y}, \mathbf{n})] \\ = \quad & 0.55 \cdot [0.92 \cdot b(\mathbf{Y}, \mathbf{p}) + 0.08 \cdot b(\mathbf{N}, \mathbf{p})] + \\ & 0.45 \cdot [0.86 \cdot b(\mathbf{N}, \mathbf{n}) + 0.14 \cdot p(\mathbf{Y}, \mathbf{n})] \\ = \quad & 0.55 \cdot [0.92 \cdot 99 + 0.08 \cdot 0] + \\ & 0.45 \cdot [0.86 \cdot 0 + 0.14 \cdot \ -1] \\ = \quad & 50.1 - 0.063 \\ \approx \quad & \mathbf{\$50.04} \end{aligned}$$


Instance description
True class
Score
..... p 0.99
..... p 0.98
..... n 0.96
..... n 0.90
..... p 0.88
..... n 0.87
..... p 0.85
..... p 0.80
..... n 0.70
..... p 0.65
p n
Y 0 0
N 100 100
p n
Y 1 0
N 99 100
p n
Y 2 0
N 98 100
p n
Y 2 1
N 98 99
p n
Y 6 4
N 94 96

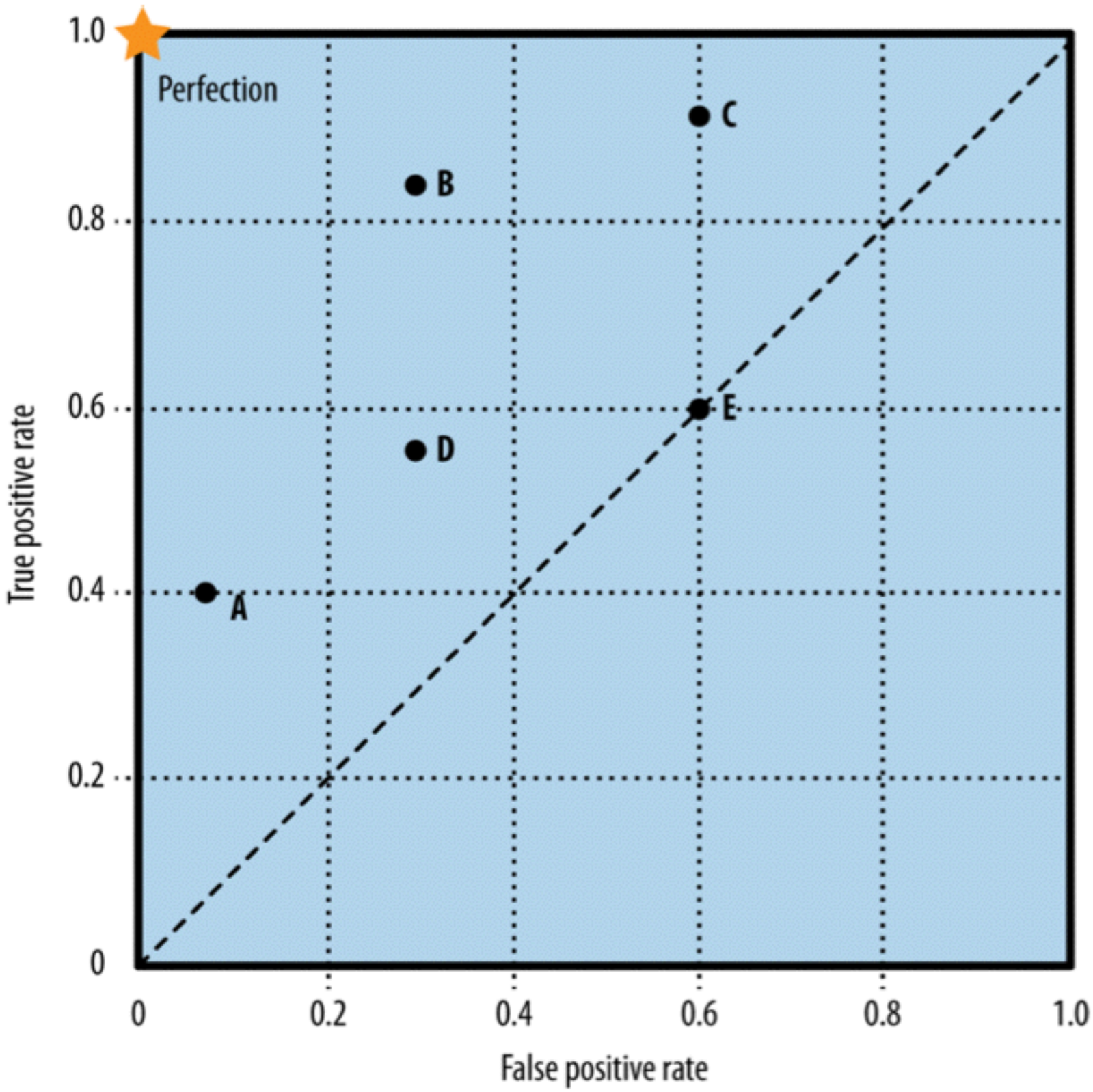

Profits of classifiers
Max profit
Max profit
at 8%
Classifier 1
Classifier 2
Classifier 3
Random
Profit
300
200
100
0
−100
−200
−300
−400
−500
−600
0
8
20
40
60
80
100
Percentage of test instances (decreasing by score)

ROC Graphs & Curves

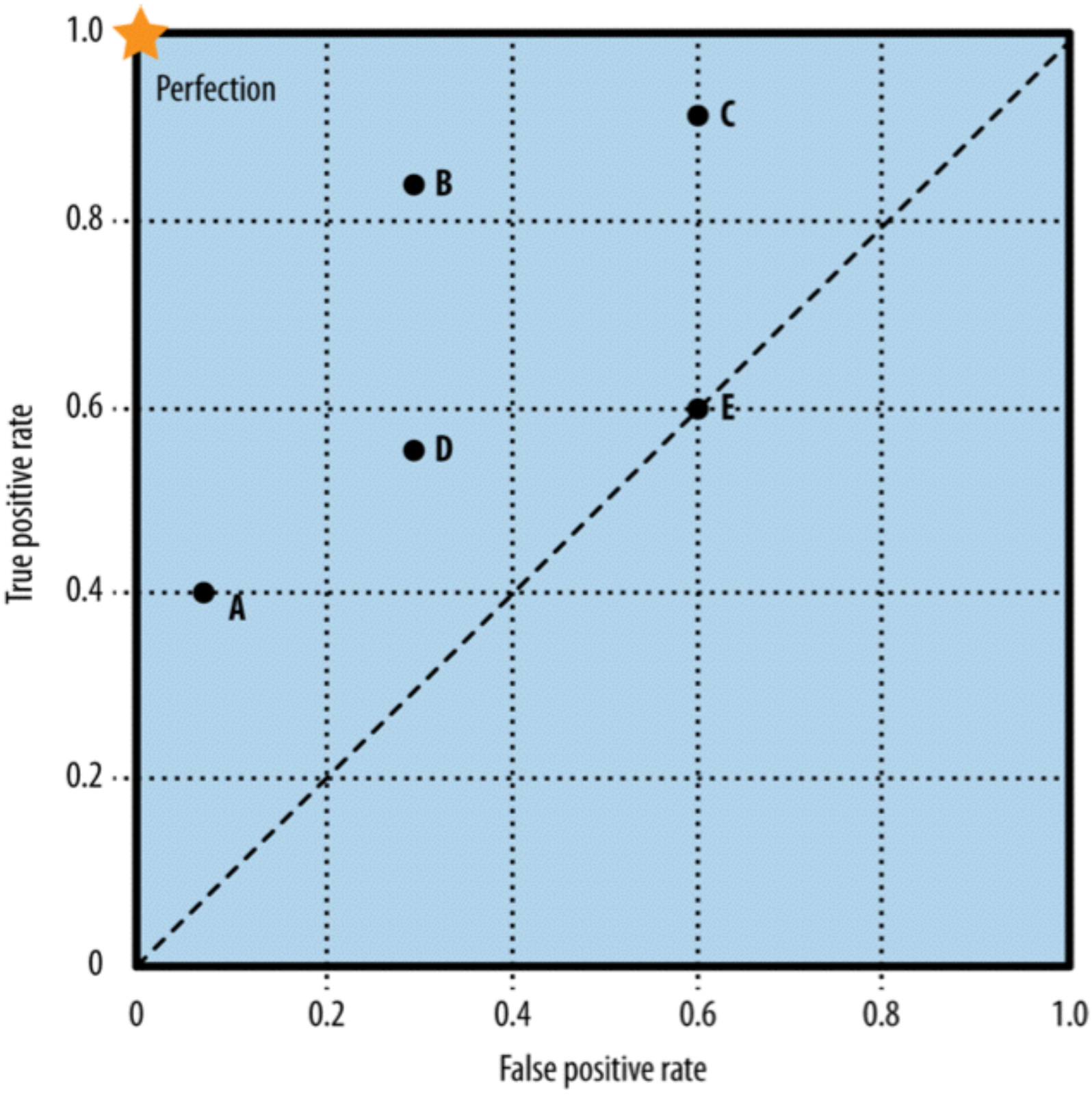
There are 2 critical conditions underlying the profit calculation

- The *class priors*; that is, the proportion of positive and negative instances in the target population, also know as the *base rate* (usually referring to the proportion of positives)

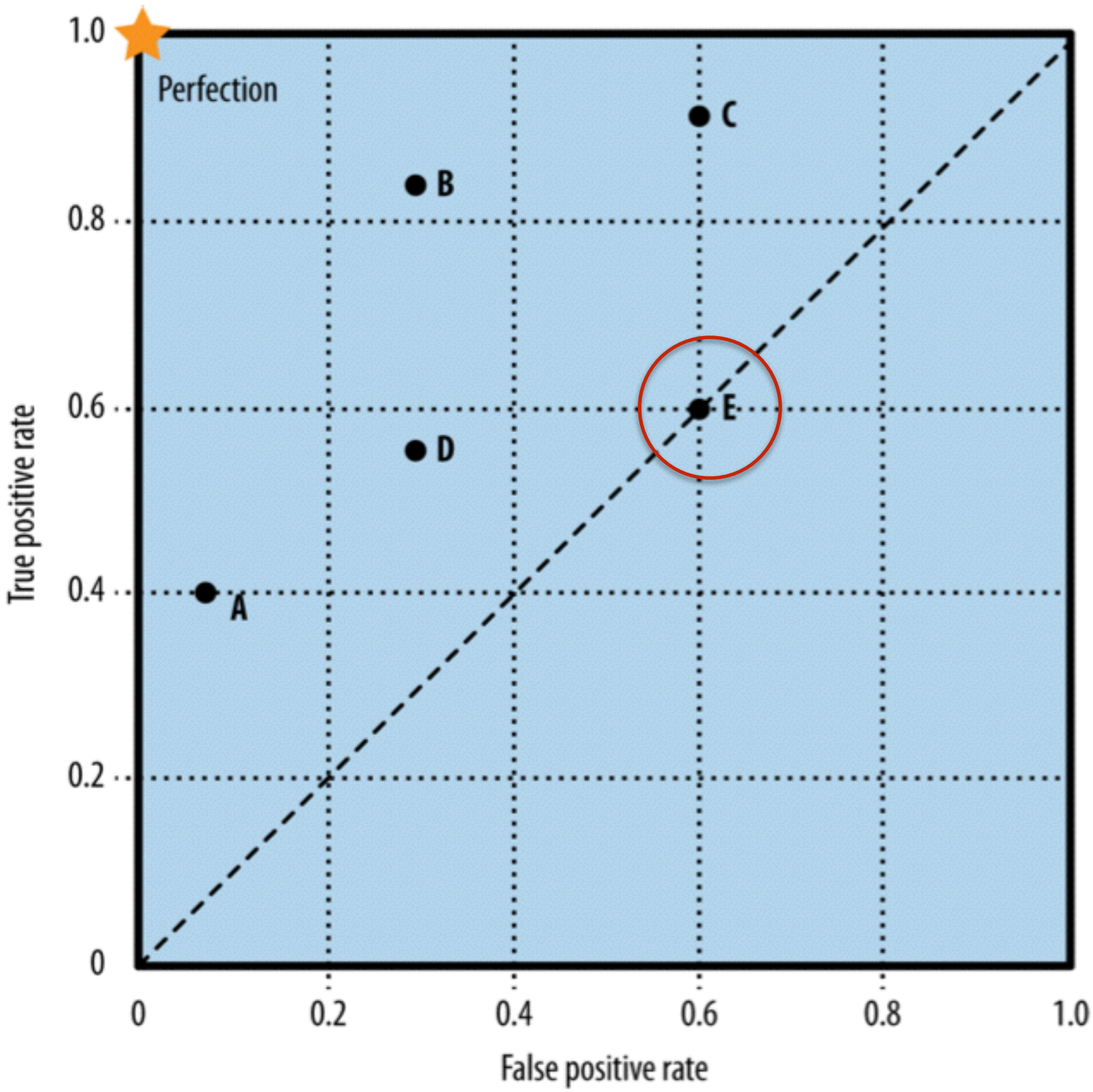
- The *costs and benefits*. The expected profit is specifically sensitive to the relative levels of costs and benefits for the different cells of the cost-benefit matrix.

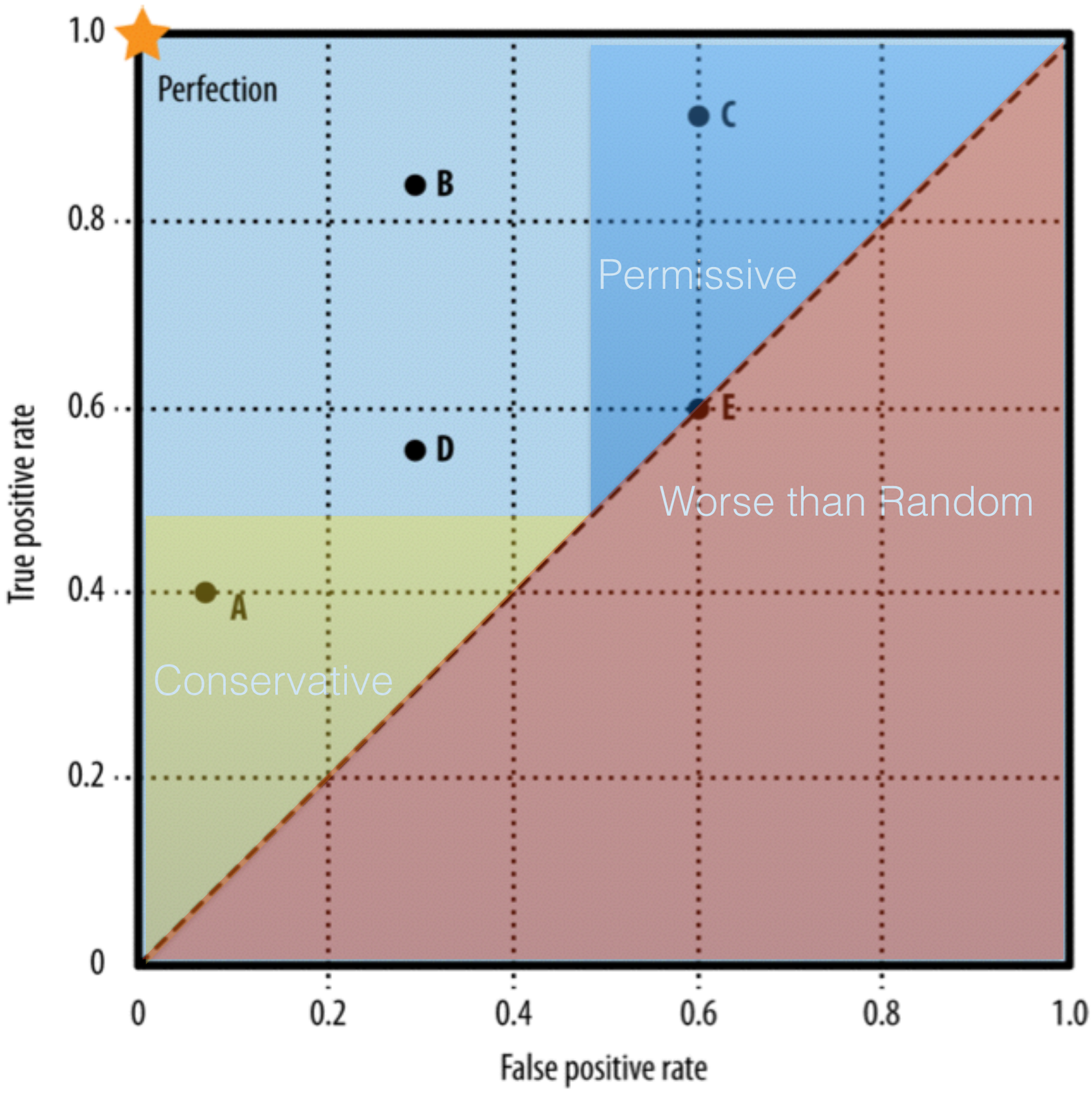

1.0
Perfection
C
B
0.8
0.6
E
D
0.4
A
0.2
0
True positive rate
0
0.2
0.4
0.6
0.8
1.0
False positive rate

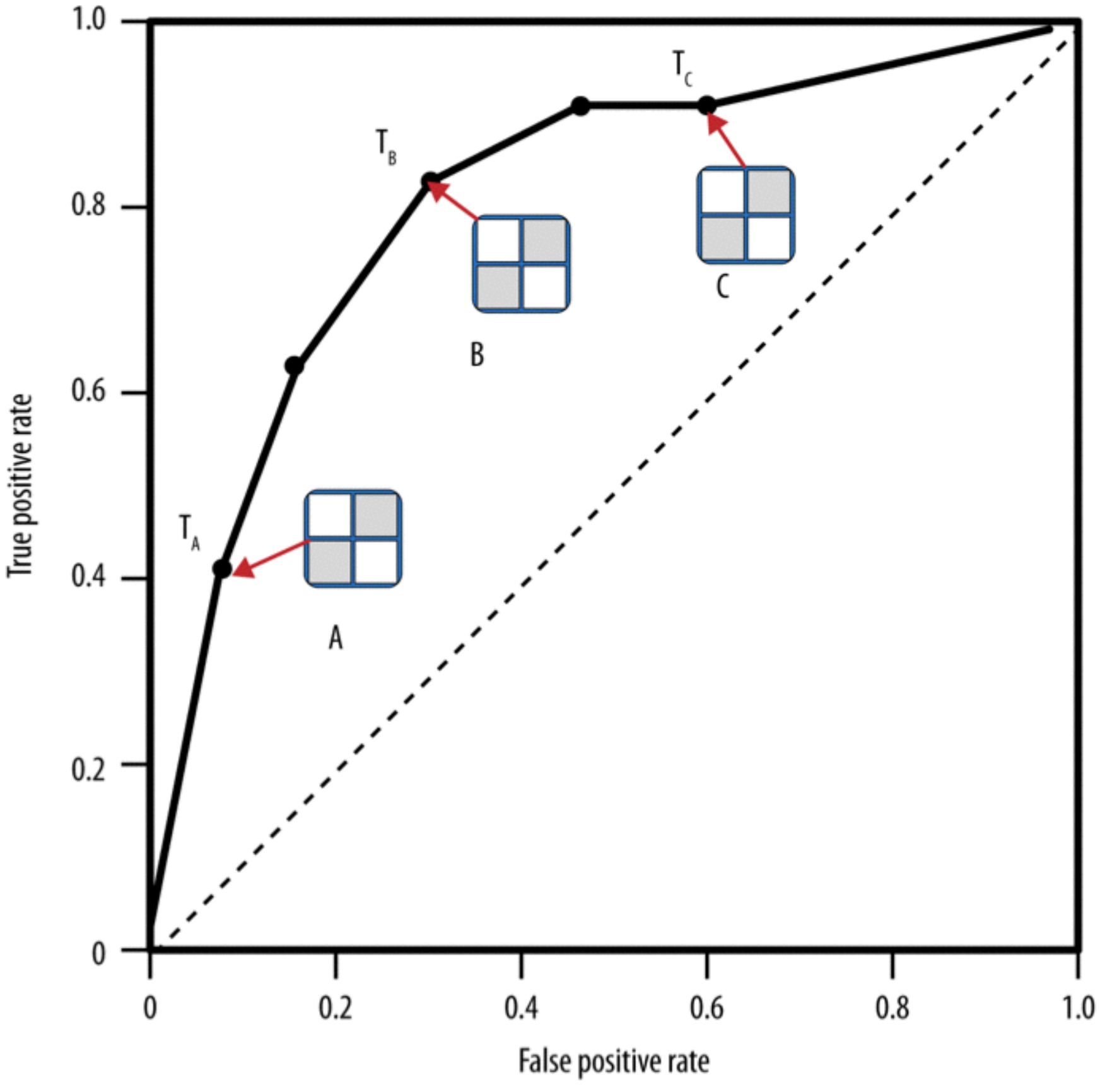
“Hit Rate”

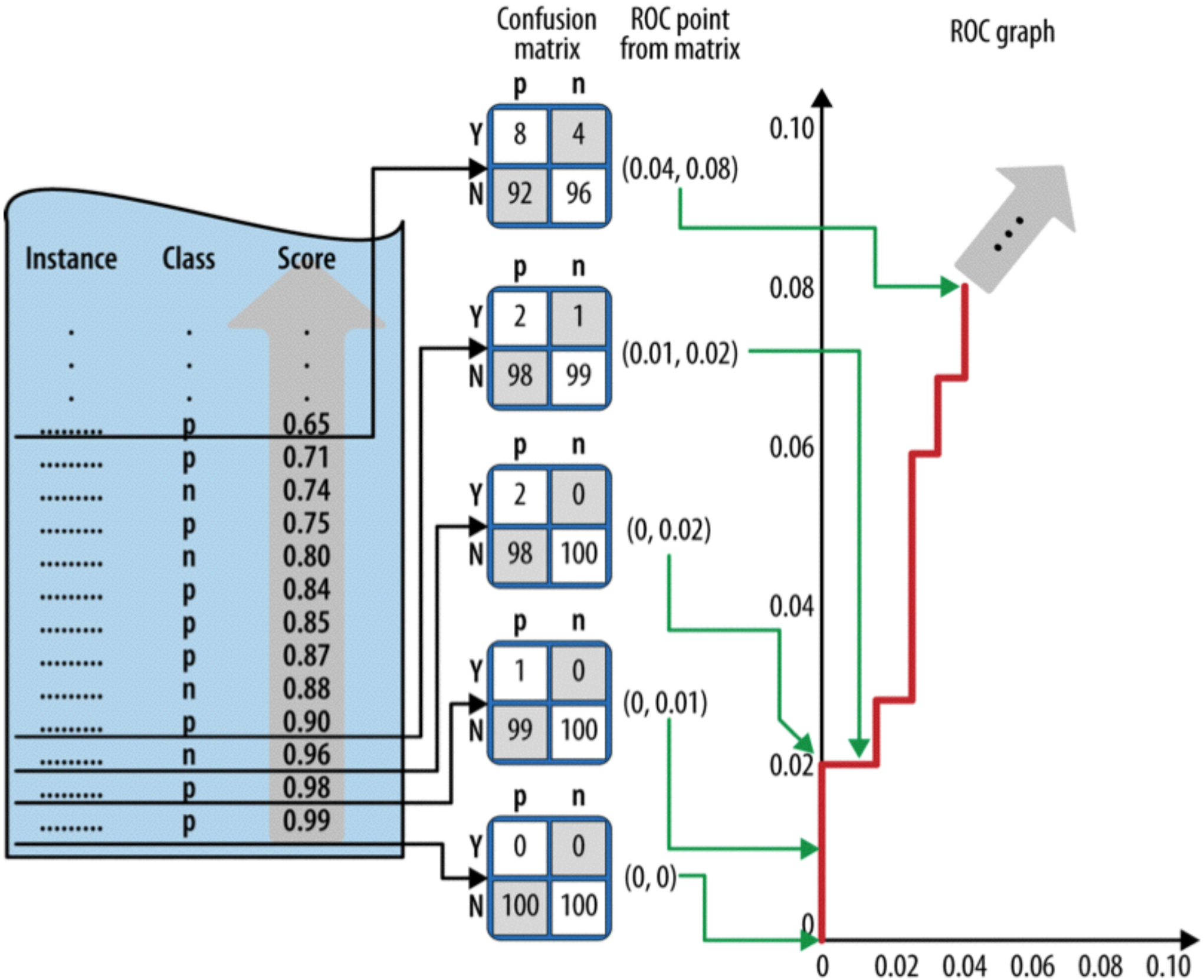


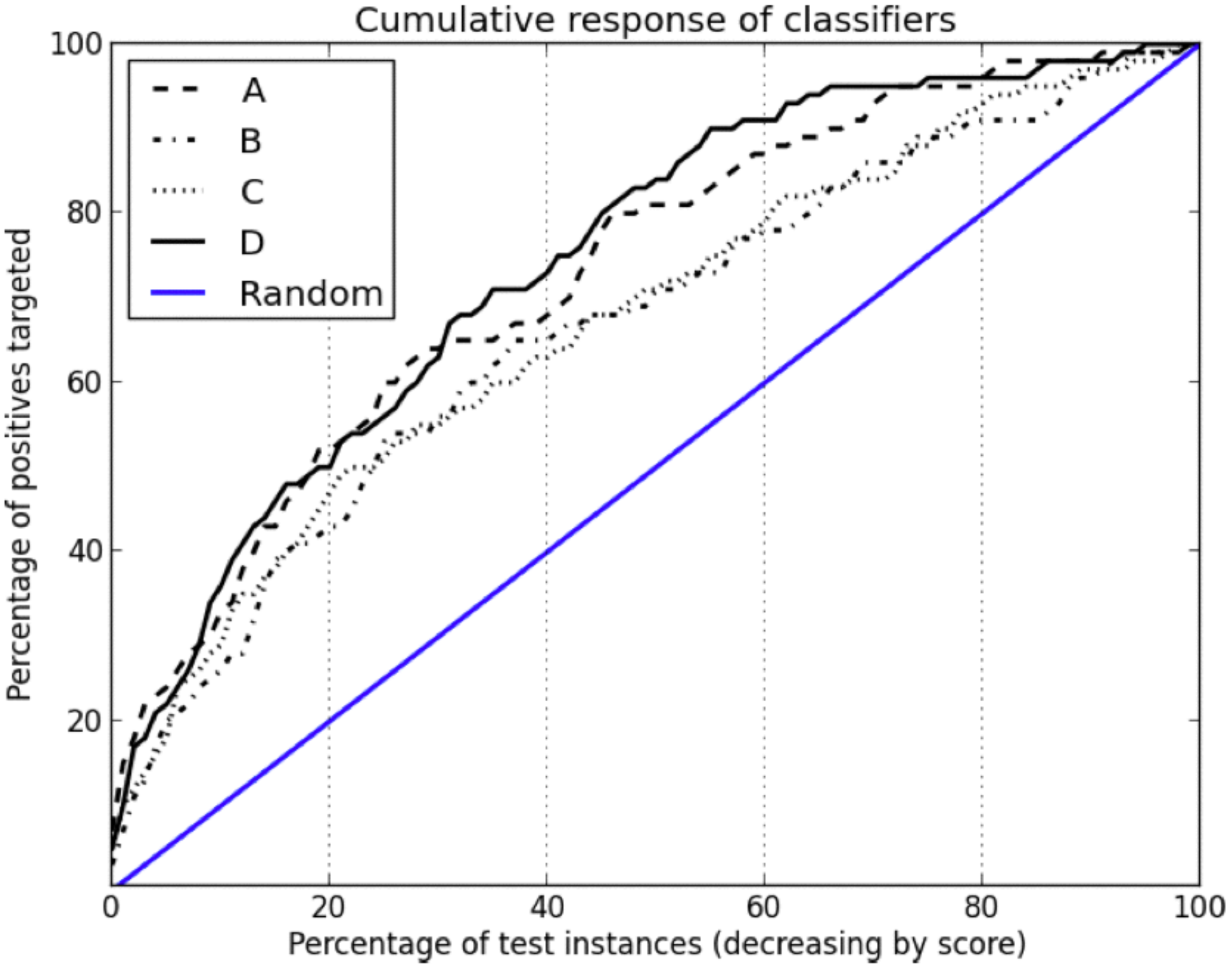
“False Alarm Rate”


1.0
Perfection
C
B
0.8
0.6
E
D
A
0.4
0.2
0
True positive rate
0
0.2
0.4
0.6
0.8
1.0
False positive rate


1.0
0.8
0.6
0.4
0.2
0
True positive rate
Perfection
B
C
D
E
A
Permissive
Worse than Random
Conservative
0
0.2
0.4
0.6
0.8
1.0
False positive rate


1.0
0.8
0.6
0.4
0.2
0
True positive rate
T_A
T_B
T_C
A
B
C
0
0.2
0.4
0.6
0.8
1.0
False positive rate


Confusion matrix
ROC point from matrix
ROC graph
Instance
Class
Score
p
0.65
p
0.71
n
0.74
p
0.75
n
0.80
p
0.84
p
0.85
p
0.87
n
0.88
p
0.90
n
0.96
p
0.98
p
0.99
p n
Y 8 4
N 92 96
(0.04, 0.08)
p n
Y 2 1
N 98 99
(0.01, 0.02)
p n
Y 2 0
N 98 100
(0, 0.02)
p n
Y 1 0
N 99 100
(0, 0.01)
p n
Y 0 0
N 100 100
(0, 0)
0.10
0.08
0.06
0.04
0.02
0
0
0.02
0.04
0.06
0.08
0.10


Cumulative response of classifiers
A
B
C
D
Random
Percentage of positives targeted
Percentage of test instances (decreasing by score)
0
20
40
60
80
100

Model	Accuracy
Classification tree	95%
Logistic regression	93%
k-Nearest Neighbor	100%
Naive Bayes	76%

Model	Accuracy (%)	AUC
Classification Tree	91.8 ± 0.0	0.614 ± 0.014
Logistic Regression	93.0 ± 0.1	0.574 ± 0.023
k-Nearest Neighbor	93.0 ± 0.0	0.537 ± 0.015
Naive Bayes	76.5 ± 0.6	0.632 ± 0.019

	p	n
Y	127 (3%)	848 (18%)
N	200 (4%)	3518 (75%)

Here is the *k*-Nearest Neighbors confusion matrix on the same test data:

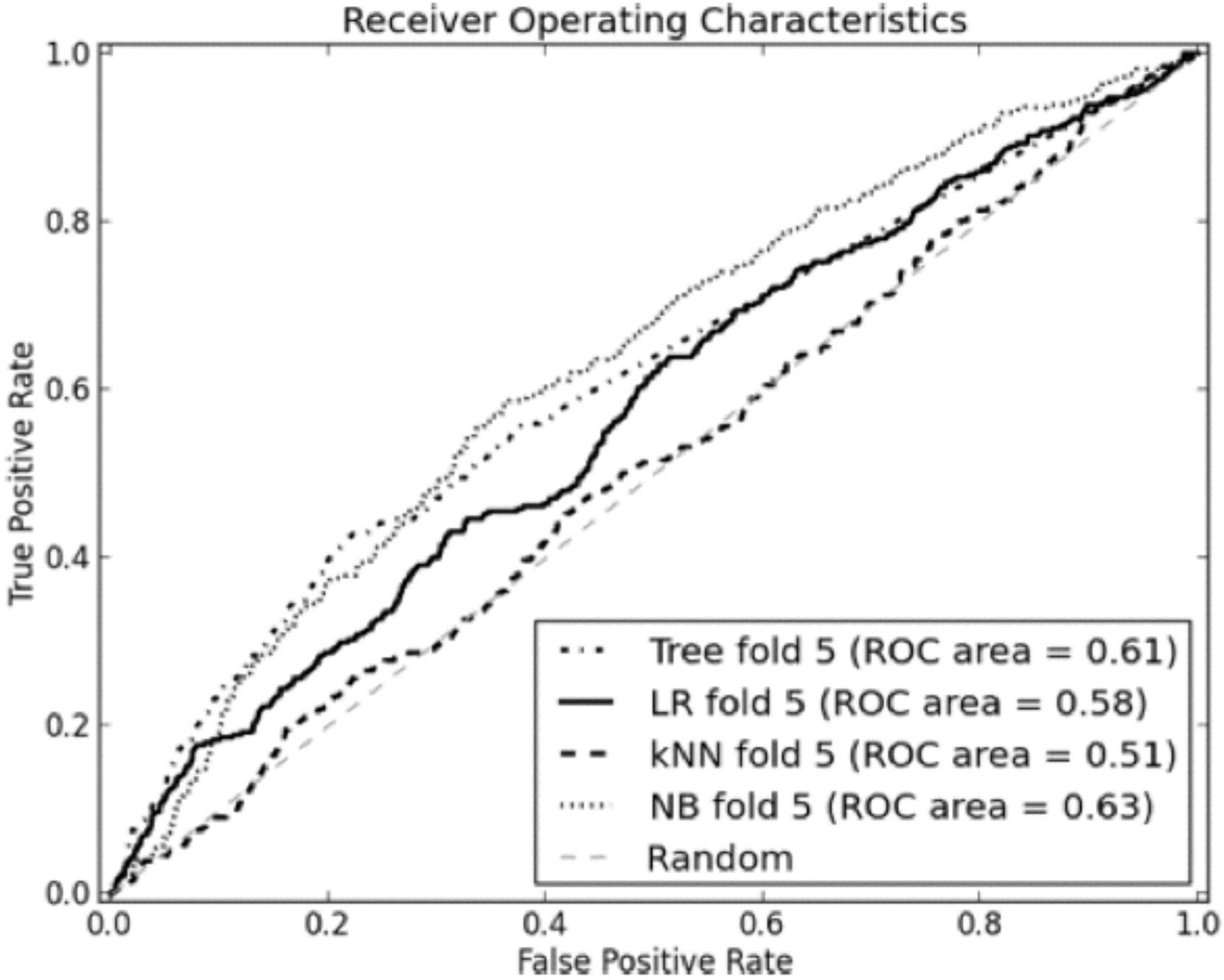
Model	Accuracy
Classification tree	95%
Logistic regression	93%
k-Nearest Neighbor	100%
Naive Bayes	76%

Model	Accuracy (%)	AUC
Classification Tree	91.8 ± 0.0	0.614 ± 0.014
Logistic Regression	93.0 ± 0.1	0.574 ± 0.023
k-Nearest Neighbor	93.0 ± 0.0	0.537 ± 0.015
Naive Bayes	76.5 ± 0.6	0.632 ± 0.019

	p	n
Y	127 (3%)	848 (18%)
N	200 (4%)	3518 (75%)

Here is the *k*-Nearest Neighbors confusion matrix on the same test data:

	p	n
Y	3 (0%)	15 (0%)
N	324 (7%)	4351 (93%)



ROC curves of the classifiers on one fold of cross-validation for the churn problem