Balance Precisión-Memoria (*Precision- Recall trade-Off*)

- **Memoria-Sensitividad (**Recall. Sensitivity)**: Recall = VP / (VP + FN)**
  Porcentaje de **REALMENTE** POSITIVOS (VP+FN) que el modelo clasificó correctamente como POSITIVOS (V**P**).

  OBJETIVO: mide qué tan bien se controlan los FALSOS NEGATIVOS (ej: cáncer)

  PROBLEMAS: El modelo puede engañarnos y maximizar esta Métrica si predice siempre positivo, o sea bajando el umbral (treshold).

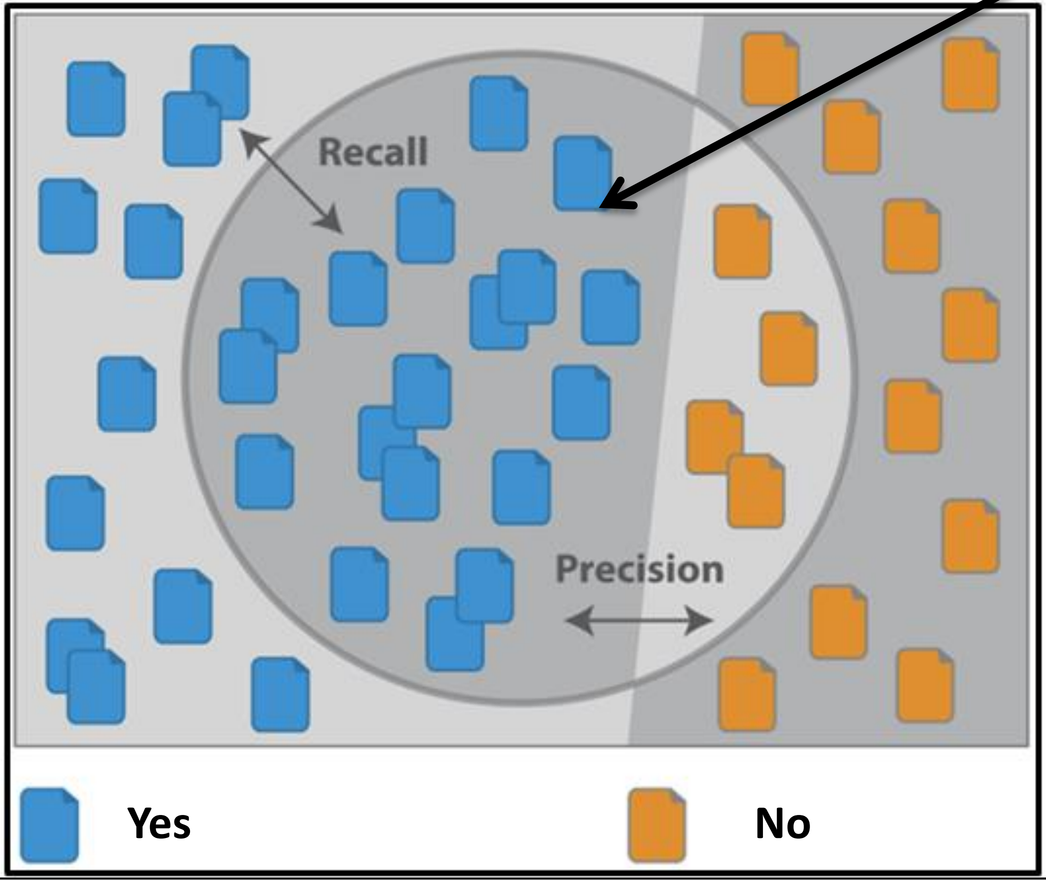
- **Precisión:** porcentaje de los **CLASIFICADOS** POSITIVOS (VP+FP)que realmente son POSITIVOS (**V**P). **Precision = VP / (VP + FP)**

  OBJETIVO: mide qué tan bien se controlan los FALSOS POSITIVOS (ej: Justicia)

  PROBLEMAS: El modelo puede engañarnos si sólo clasifica como positivos a los que tiene mucha confianza, o sea subiendo el umbral.

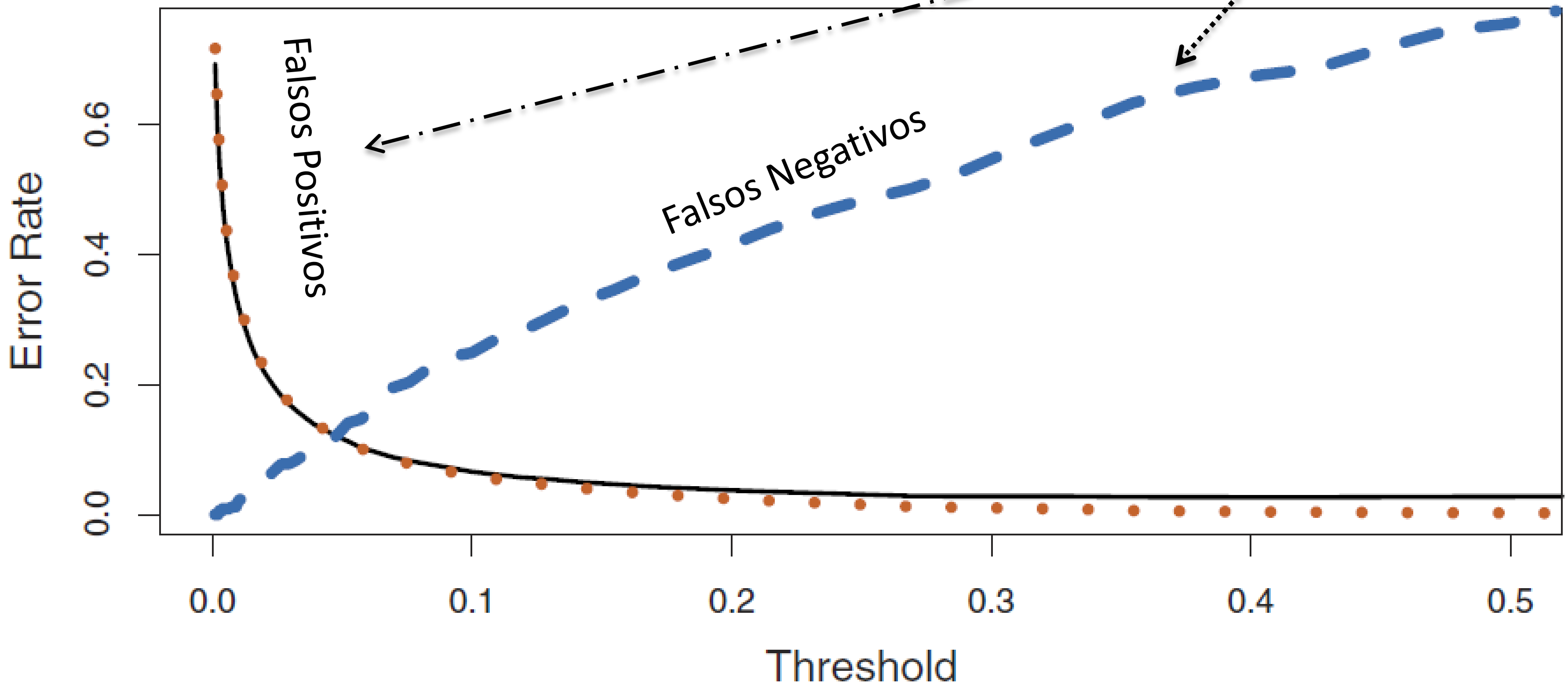
Matriz de Confusión

		Real		
		No	Yes	Total
Clasificado	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000



Matriz de Confusión

		Real		
		No	Yes	Total
Clasificado	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000



Comparación de Modelos: Costo-Beneficio

Matriz de Confusión

Datos reales

Predicciones del Modelo

		True default status		
		No	Yes	Total
	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000

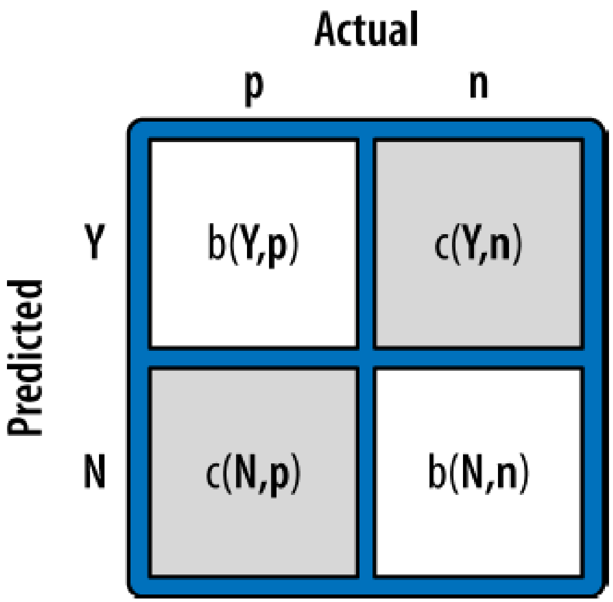
Valor Esperado

$$E(f) = \sum_i p(x_i) \cdot f(x_i)$$

Si una decisión pudiera presentar *t* situaciones o salidas diferentes, con un resultado (positivo o negativo) asociado y una probabilidad asociado a cada uno de ellas, el valor esperado de esa decisión sería :

$$\boxed{\text{Valor esperado} = p(s_1)\, v(s_1) + p(s_2) v(s_2) + \ldots + p(s_t) v(s_t)}$$

Veamos un ejemplo: Realizaremos una campaña de marketing en base a un dataset con información de nuestros clientes y campañas previas. El beneficio que obtendremos si un cliente responde es de $\mathbf{v}_R$ = \$99 pues compra el producto que le enviamos en la promoción. El costo si no responde a la campaña es 1, o sea $\mathbf{v}_{NR}$ = -\$1



Matriz de confusión

	p	n
Y	56	7
N	5	42

Matriz de Beneficio

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

$$\textit{Expected profit} = \overbrace{p(\mathbf{Y},\mathbf{p}) \cdot b(\mathbf{Y},\mathbf{p})}^{\text{VP}} + \overbrace{p(\mathbf{N},\mathbf{p}) \cdot b(\mathbf{N},\mathbf{p})}^{\text{FN}} + \overbrace{p(\mathbf{N},\mathbf{n}) \cdot b(\mathbf{N},\mathbf{n})}^{\text{VN}} + \overbrace{p(\mathbf{Y},\mathbf{n}) \cdot b(\mathbf{Y},\mathbf{n})}^{\text{FP}} \quad \textbf{(1)}$$

Por teoría de la probabilidad sabemos que: $p(x, y) = p(y) \cdot p(x \mid y)$

Por lo que podemos expresar a $p(\mathbf{Y},\mathbf{p})$ como $p(\mathbf{Y},\mathbf{p}) = p(\mathbf{Y} \mid \mathbf{p}) \cdot p(\mathbf{p})$
Quedándonos el beneficio esperado:

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

$$\boxed{\begin{aligned} \text{Expected profit} = \; & p(\mathbf{p}) \cdot [p(\mathbf{Y} \mid \mathbf{p}) \cdot b(\mathbf{Y},\mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot c(\mathbf{N},\mathbf{p})] + \\ & p(\mathbf{n}) \cdot [p(\mathbf{N} \mid \mathbf{n}) \cdot b(\mathbf{N},\mathbf{n}) + p(\mathbf{Y} \mid \mathbf{n}) \cdot c(\mathbf{Y},\mathbf{n})] \end{aligned}} \quad \textbf{(2)}$$

Matriz de confusión

	p	n
Y	56	7
N	5	42

$T = 110$

$P = 61$ $\quad N = 49$

$p(\mathbf{p}) = 0.55$ $\quad p(\mathbf{n}) = 0.45$

$tp\ rate = 56/61 = 0.92$ $\quad fp\ rate = 7/49 = 0.14$

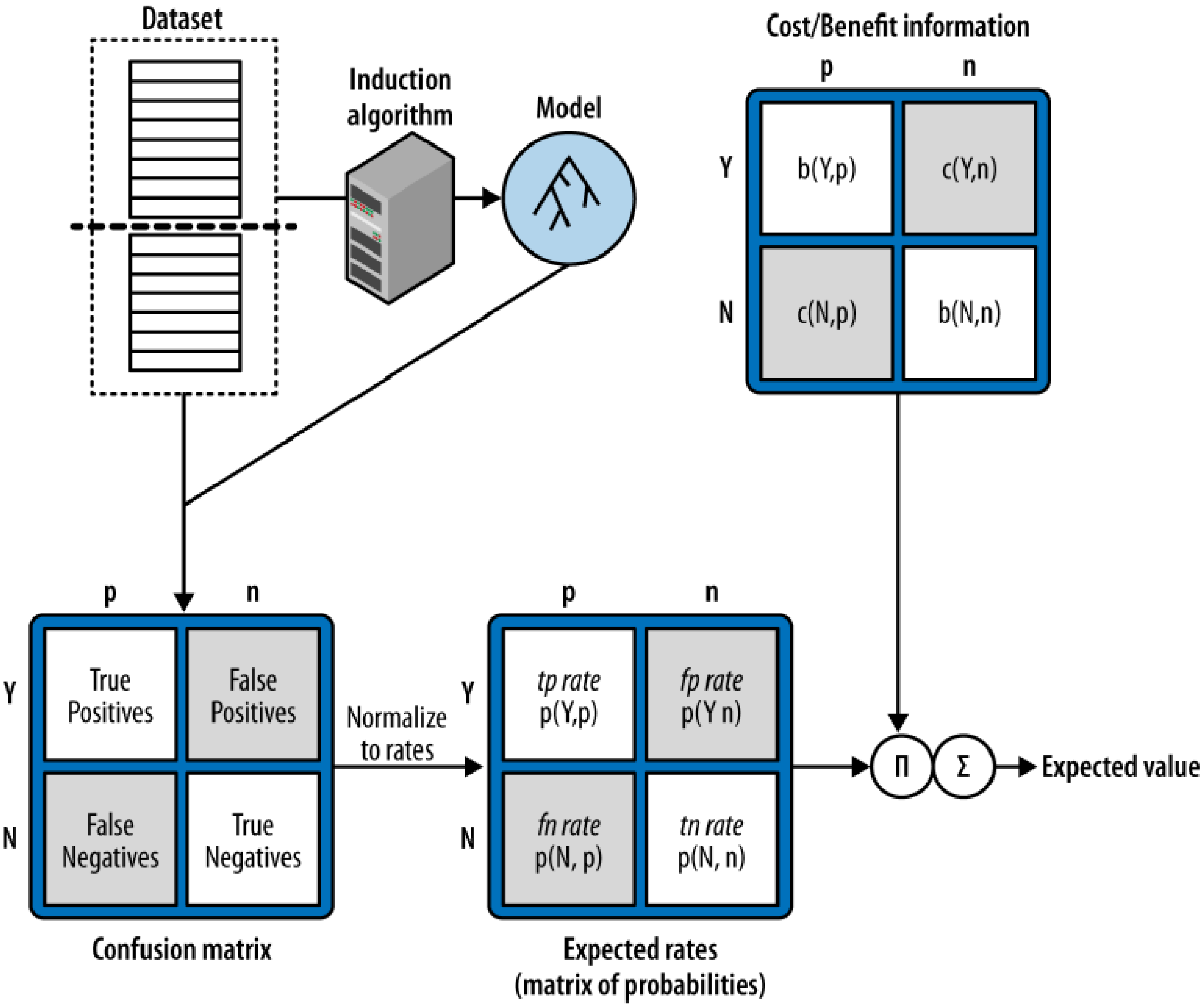
$fn\ rate = 5/61 = 0.08$ $\quad tn\ rate = 42/49 = 0.86$

Matriz de Costo/Beneficio

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

Reemplazando estos valores en (2), obtenemos:

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

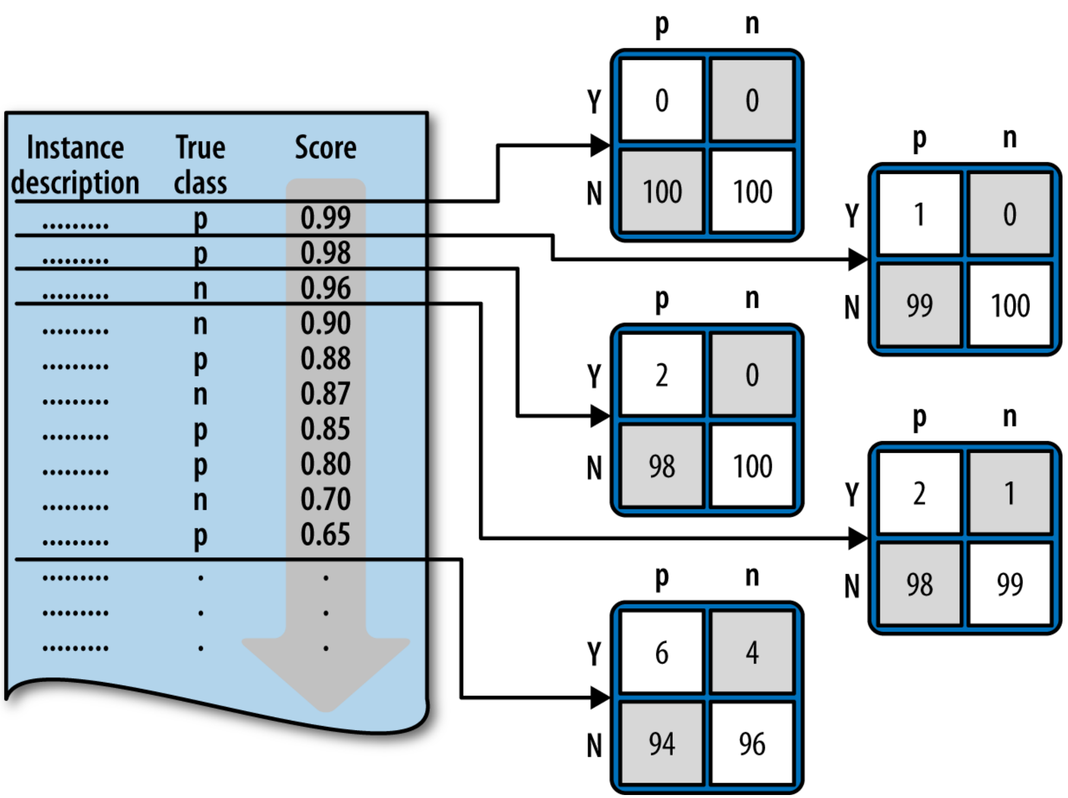


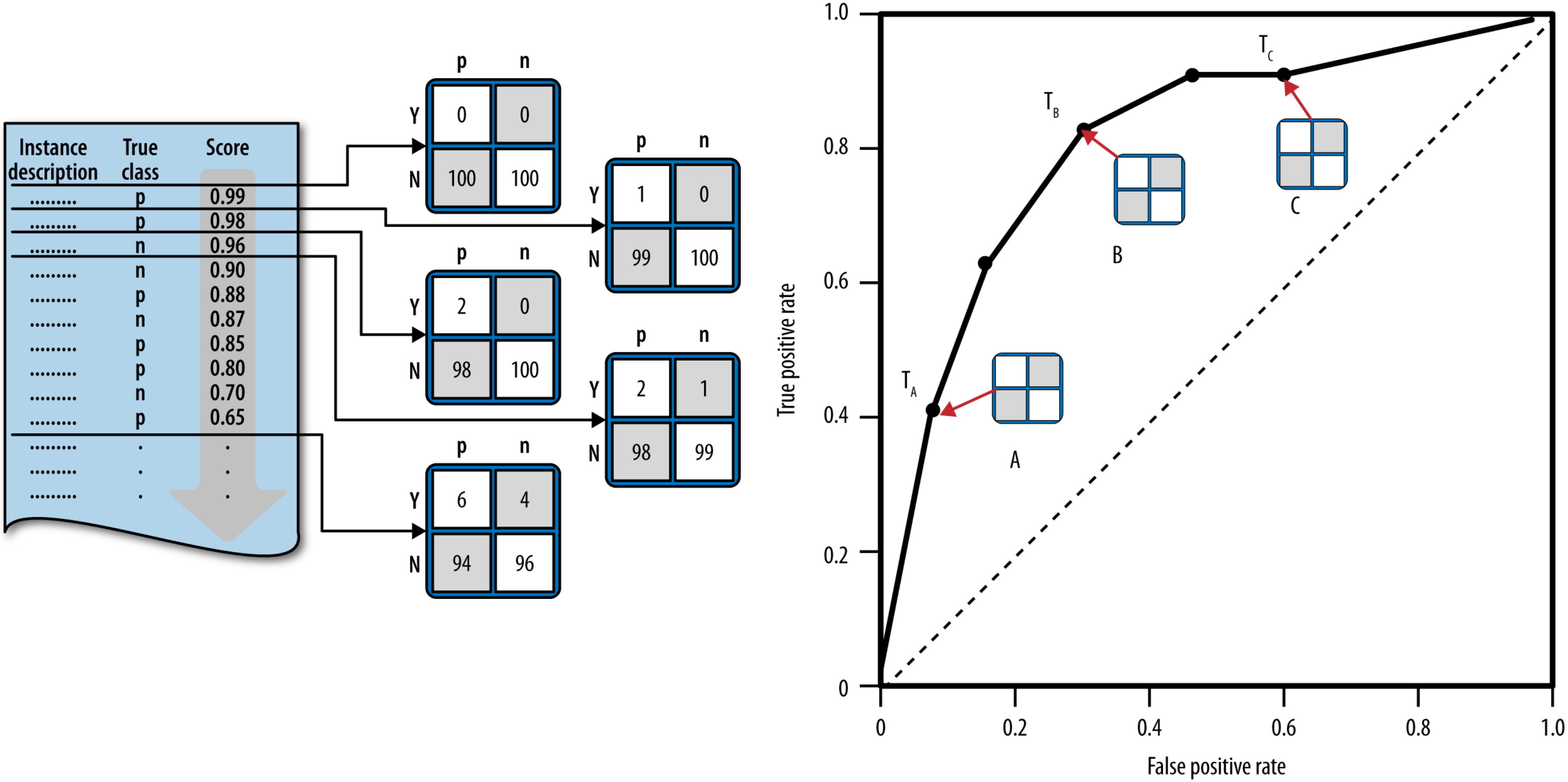
FACULTAD DE INGENIERIA
Universidad de Buenos Aires

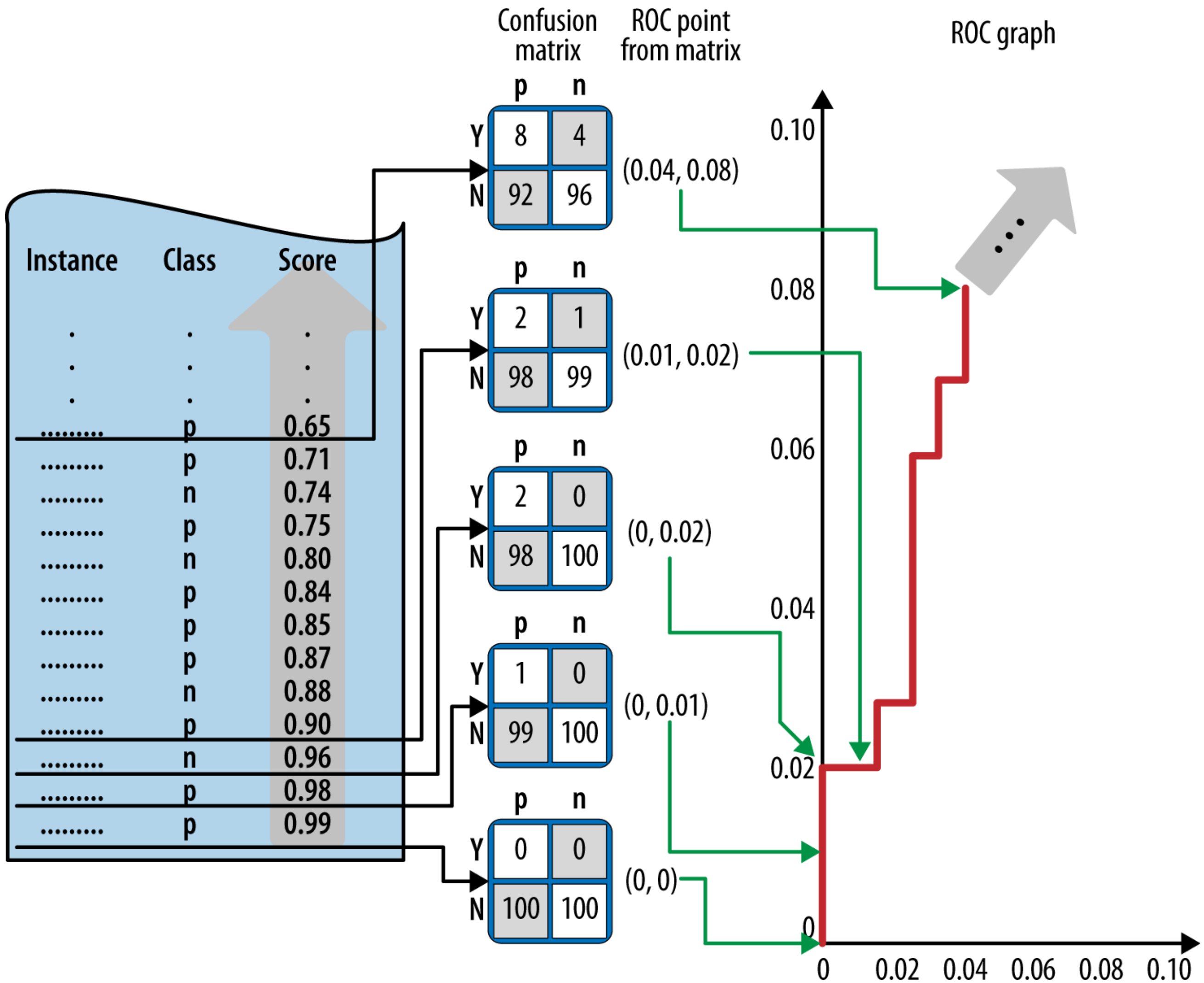
Curva ROC-AUC

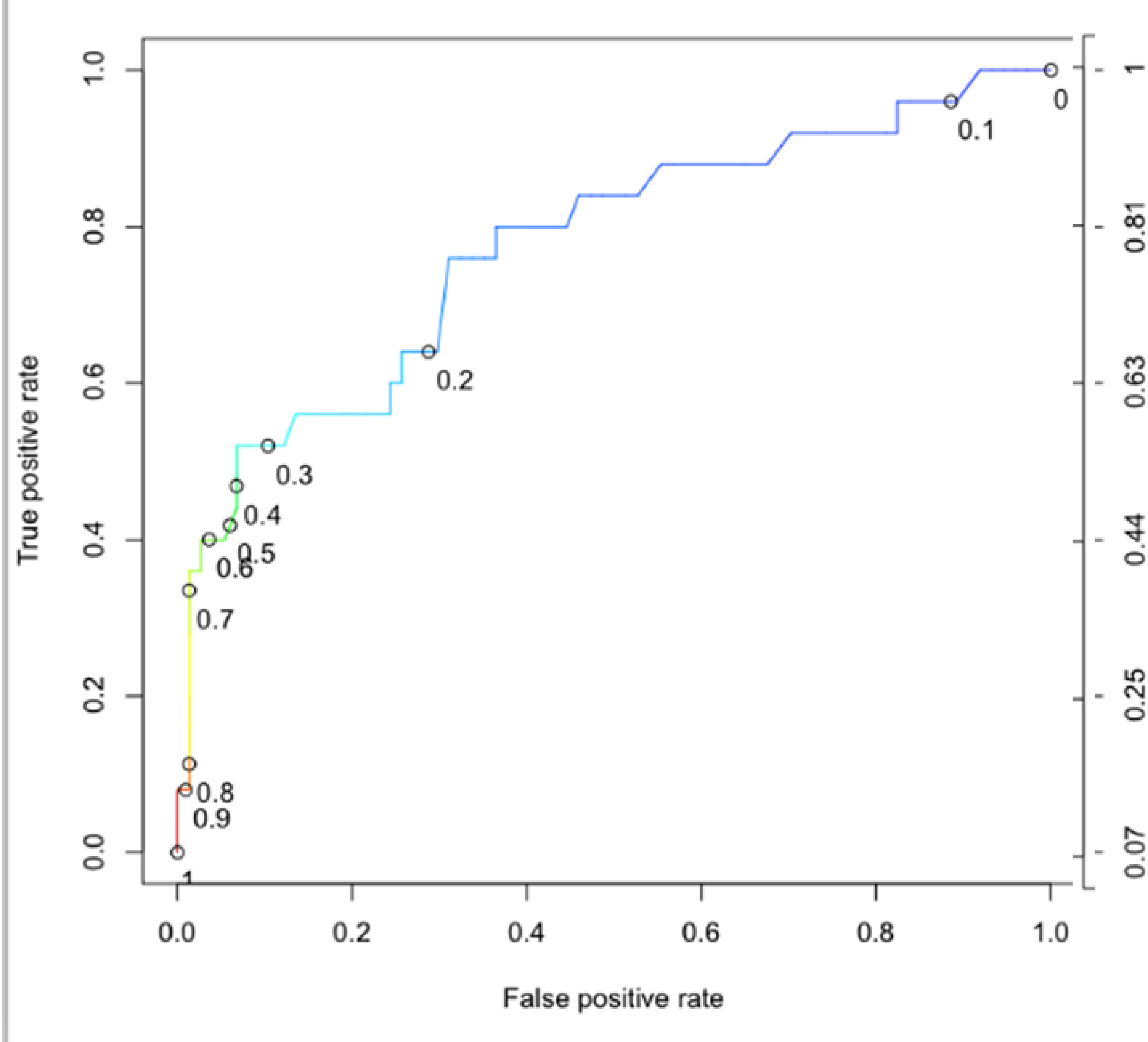
Método que puede mostrar la performance para todos los umbrales de clasificación posibles.

Curva de **Características Operativas del Receptor** (ROC). Un gráfico ROC es un diagrama bidimensional de un clasificador con la tasa de falsos positivos en el eje x contra la tasa de verdaderos positivos en el eje y. Como tal, un gráfico ROC representa las compensaciones relativas que hace un clasificador entre los ganancias (verdaderos positivos) y los costos (falsos positivos).

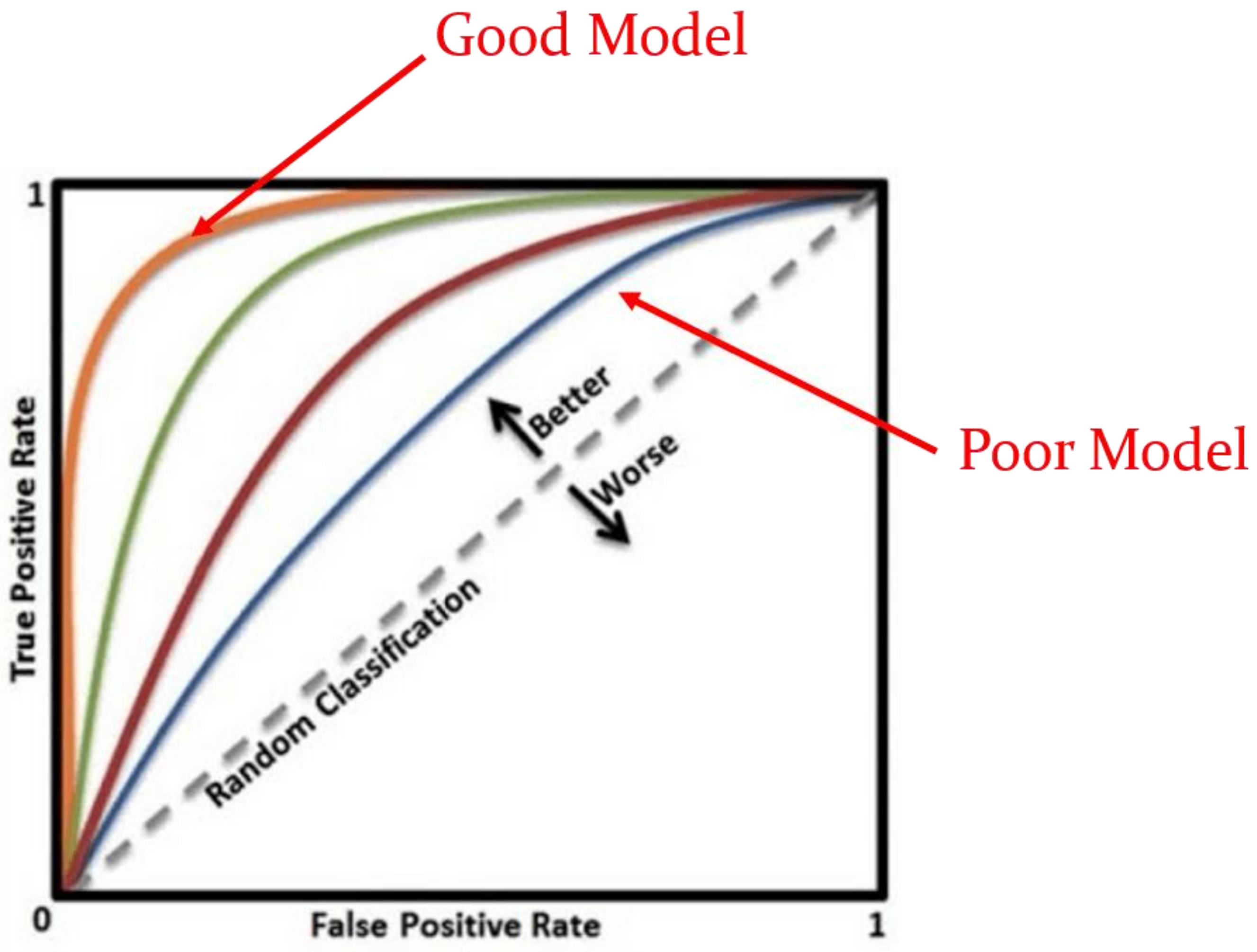


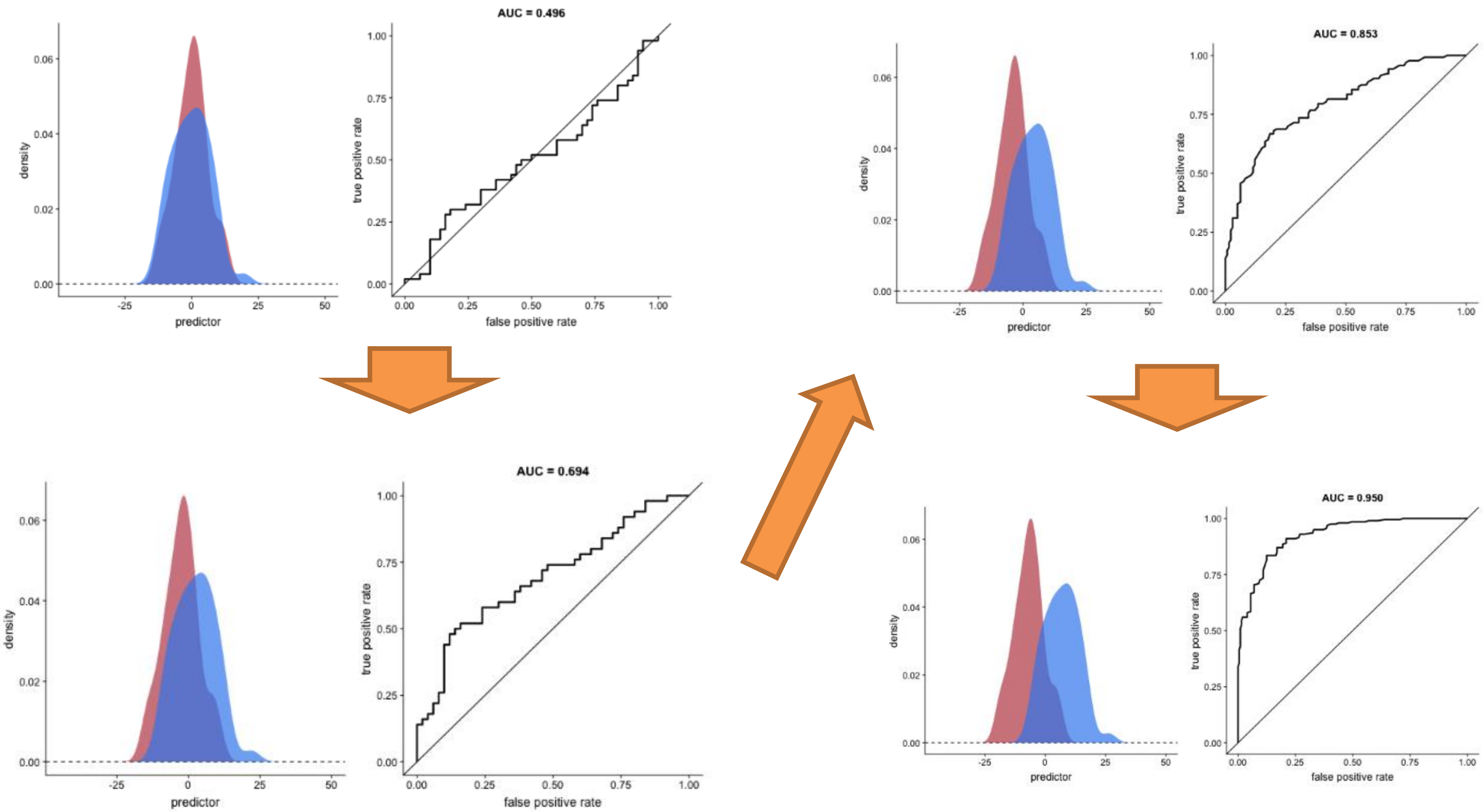

Instance description
True class
Score
p 0.99
p 0.98
n 0.96
n 0.90
p 0.88
n 0.87
p 0.85
p 0.80
n 0.70
p 0.65
p n
Y 0 0
N 100 100
Y 1 0
N 99 100
Y 2 0
N 98 100
Y 2 1
N 98 99
Y 6 4
N 94 96
True positive rate
False positive rate
0
0.2
0.4
0.6
0.8
1.0
A
B
C

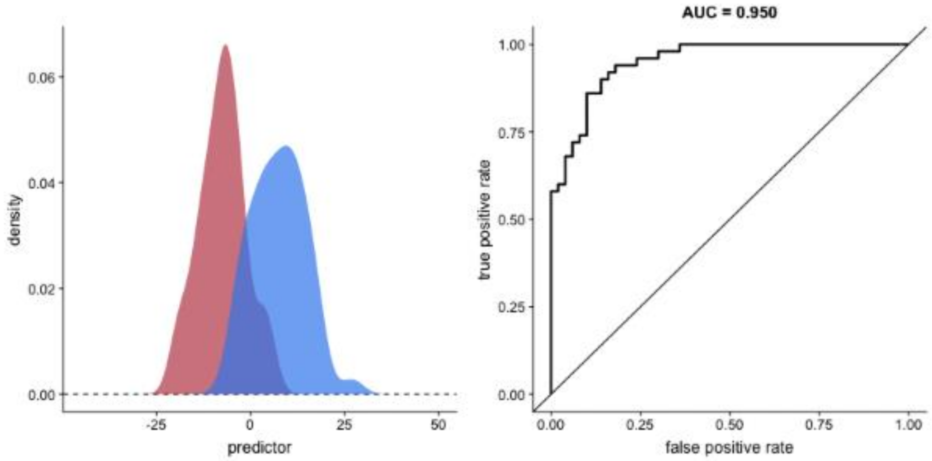

Confusion matrix
ROC point from matrix
ROC graph
Instance
Class
Score
.... p 0.65
.... p 0.71
.... n 0.74
.... p 0.75
.... n 0.80
.... p 0.84
.... p 0.85
.... p 0.87
.... n 0.88
.... p 0.90
.... n 0.96
.... p 0.98
.... p 0.99
p n
Y 8 4
N 92 96
(0.04, 0.08)
p n
Y 2 1
N 98 99
(0.01, 0.02)
p n
Y 2 0
N 98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

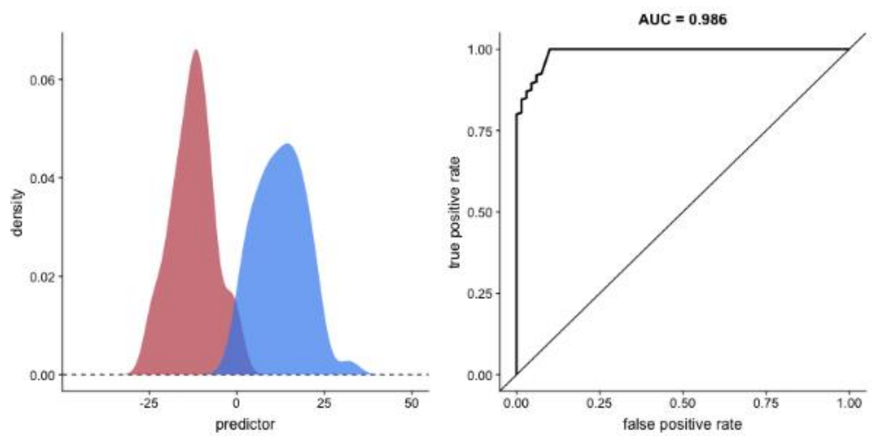


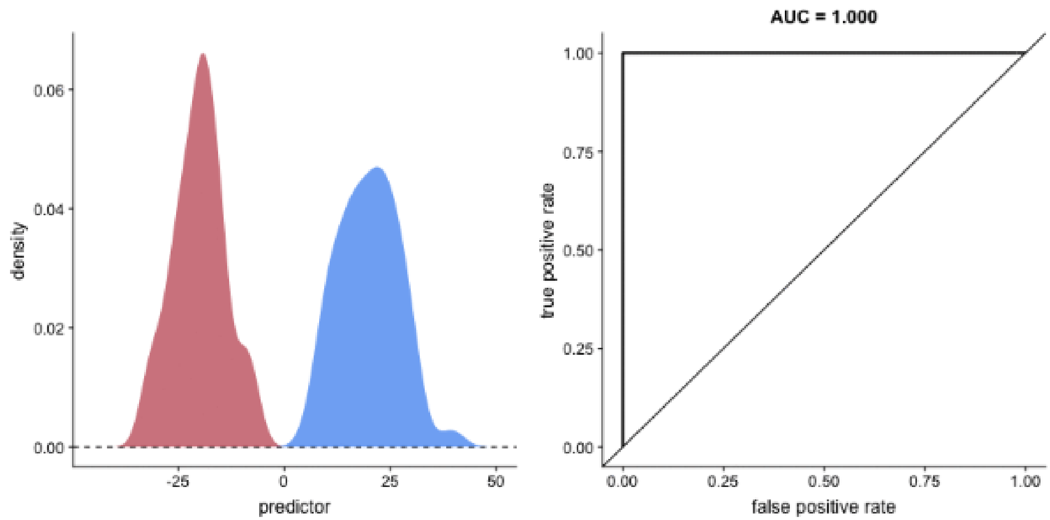
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.07
0.25
0.44
0.63
0.81
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


Good Model
Poor Model
1
True Positive Rate
Better
Worse
Random Classification
0
False Positive Rate
1


AUC = 0.496
AUC = 0.694
AUC = 0.853
AUC = 0.950
density
predictor
true positive rate
false positive rate


AUC = 0.950
density
predictor
true positive rate
false positive rate
0.00
0.02
0.04
0.06
-25
0
25
50
0.25
0.50
0.75
1.00


AUC = 0.986
density
predictor
true positive rate
false positive rate
0.00
0.02
0.04
0.06
-25
0
25
50
0.25
0.50
0.75
1.00


AUC = 1.000
density
0.06
0.04
0.02
0.00
-25
0
25
50
predictor
true positive rate
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
false positive rate