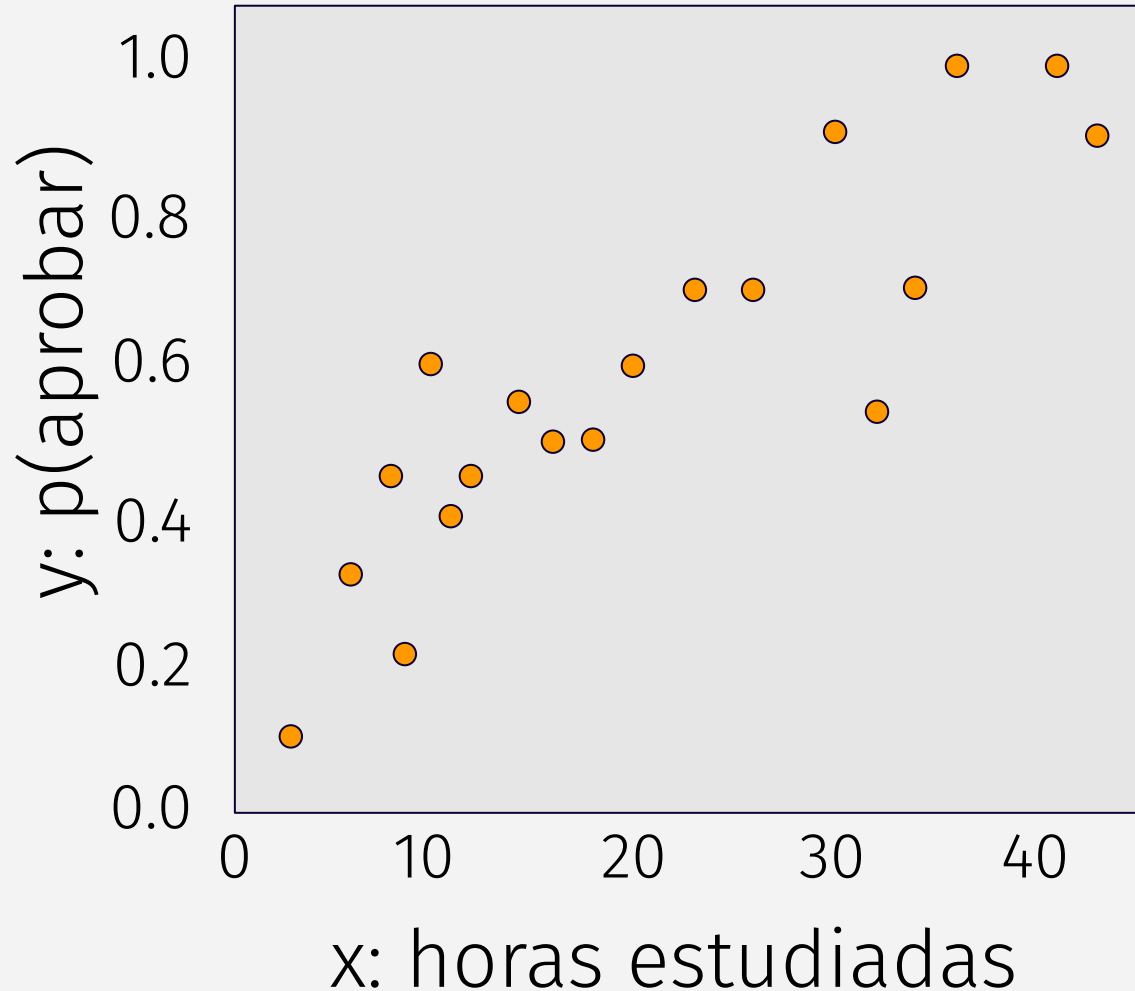


# Clasificación binaria y predicción de probabilidades

---

# Predicción de probabilidad de aprobar

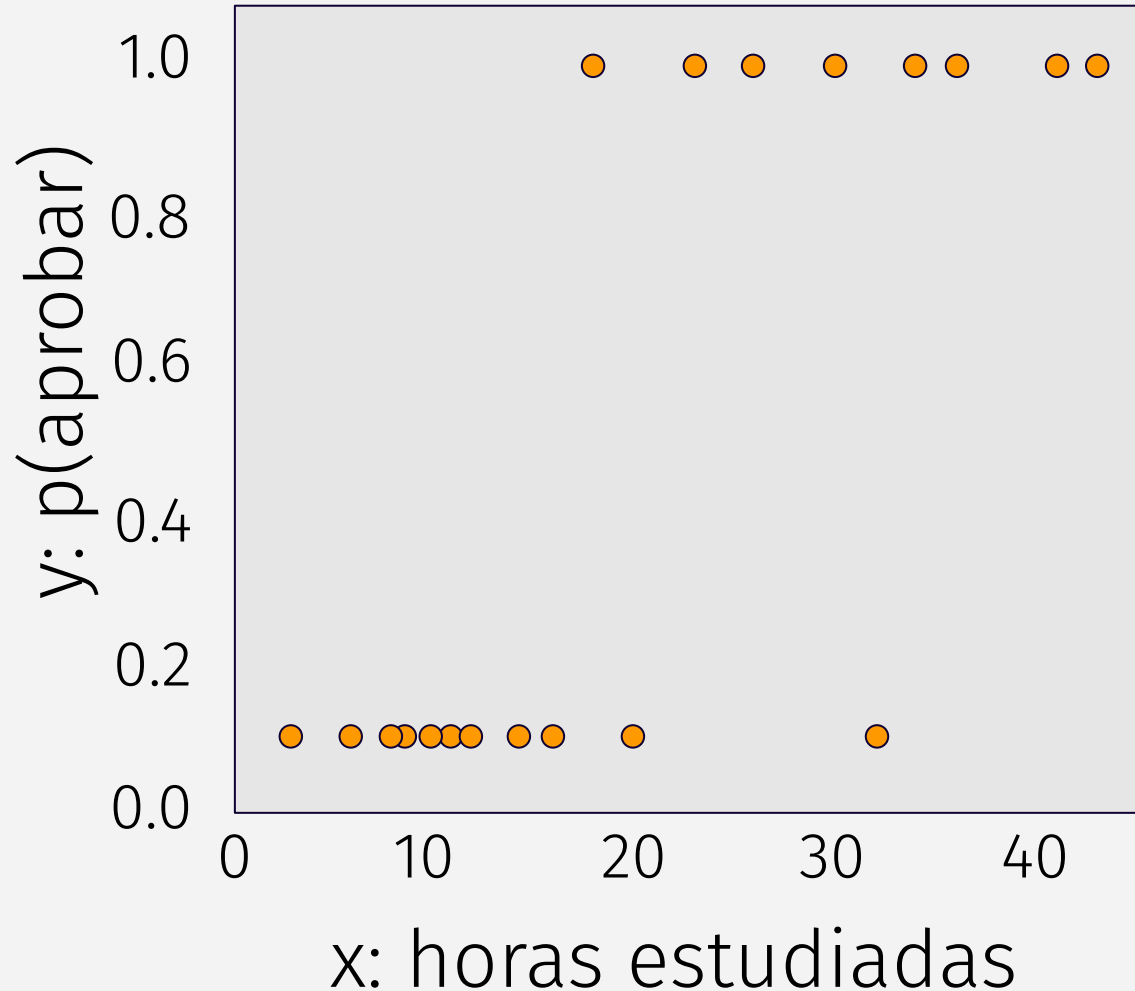
- Entrada
  - Horas estudiadas
- Salida
  - Probabilidad de aprobar
- Datos de probabilidades
  - No son comunes



Horas	P(aprobar)
2	0.1
5	0.32
7	0.45
9	0.6
10	0.4
11	0.45
13.4	0.55
14	0.3
15	0.5

# Clasificación en aprobado/desaprobado





- Entrada
  - Horas estudiadas
- Salida
  - 1: Aprobado
  - 0: Desaprobado
- Datos de aprobado/desaprobado
  - Fáciles de obtener

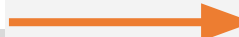
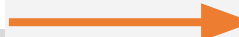
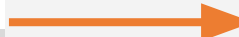
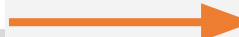


Horas	Aprobado
2	0
5	0
7	0
9	0
10	0
11	0
13.4	1
14	0
15	1

# Convertir probabilidades en clases

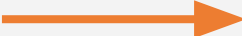
- Umbral  $u$ 
  - Método más común
  - $u$ : valor entre 0 y 1
- Si  $p(\text{aprobar}) > u$ 
  - Aprobado
- Sino
  - Desaprobado

Horas	P(aprobar)	$u=0.35$	Horas	Aprobado
2	0.1		2	0
5	0.32		5	0
7	0.45		7	1
9	0.6		9	1
10	0.4		10	1
11	0.45		11	1
13.4	0.55		13.4	1
14	0.3		14	0
15	0.5		15	1

Horas	P(aprobar)	$u=0.5$	Horas	Aprobado
2	0.1		2	0
5	0.32		5	0
7	0.45		7	0
9	0.6		9	1
10	0.4		10	0
11	0.45		11	0
13.4	0.55		13.4	1
14	0.3		14	0
15	0.5		15	1

# Convertir clases en probabilidades

- 0 codifica desaprobado
- 1 codifica aprobado
  - Ya son probabilidades!



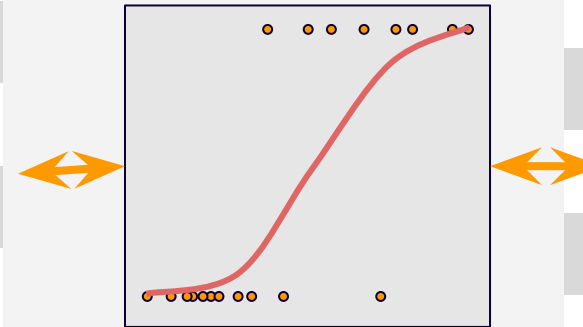
Horas	Aprobado
2	0
5	0
7	0
9	1
10	0
11	0
13.4	1
14	0
15	1

Horas	P(aprobar)
2	0
5	0
7	0
9	1
10	0
11	0
13.4	1
14	0
15	1

# Predecir $p(\text{aprobar})$ vs clasificar aprobado/desap

- **Problemas distintos**
- Clasificación
  - 0 o 1
  - No hay intermedios
  - Caso especial de probabilidad
- Probabilidad
  - Valores entre 0 y 1
  - Más detalle
- Pero **relacionados**
  - Mismo modelo

Horas	P(aprobar)
2	0.1
5	0.32
7	0.45
9	0.6
10	0.4
11	0.45
13.4	0.55
14	0.3
15	0.5



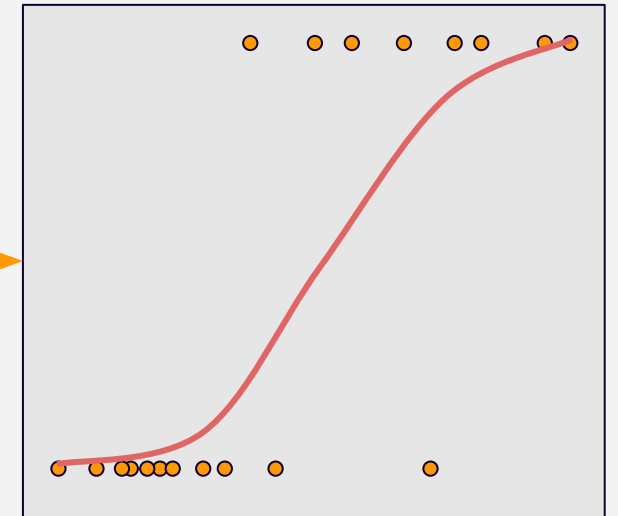
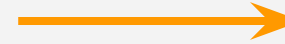
Horas	Aprobado
2	0
5	0
7	0
9	0
10	0
11	0
13.4	1
14	0
15	1

# Modelo para probabilidades/clases

- **Entrenar**

- Datos 0/1
  - Más comunes
  - Pueden considerarse probabilidades

Horas	Aprobado
2	0
5	0
7	0
9	0
10	0
11	0
13.4	1
14	0
15	1



# Modelo para probabilidades/clases

- **Predecir**

- Genera probabilidades
- Opcional: convertir a clases
  - Umbral
    - $u=0.5$  lo más común

## Salida modelo

Horas	P(aprobar)
2	0.1
5	0.32
7	0.45
9	0.6
10	0.4
11	0.45
13.4	0.55
14	0.3
15	0.5

u

## Opcional

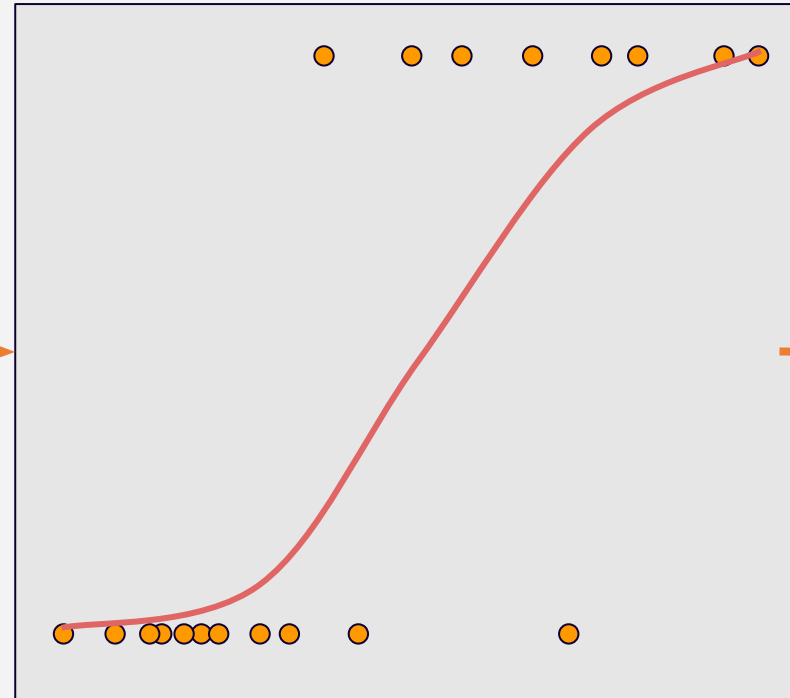
Horas	Aprobado
2	0
5	0
7	0
9	0
10	0
11	0
13.4	1
14	0
15	1



# Resumen

- Entrada
  - Horas estudiadas
- Salida
  - Probabilidad de aprobar
- Entrenado con
  - Datos de aprobado/desaprobado
  - Interpretados como probabilidades

Horas	Aprobado
2	0
5	0
7	0
9	0
10	0
11	0
13.4	1
14	0
15	1



Horas	P(aprobar)
2	0.1
5	0.32
7	0.45
9	0.6
10	0.4
11	0.45
13.4	0.55
14	0.3
15	0.5