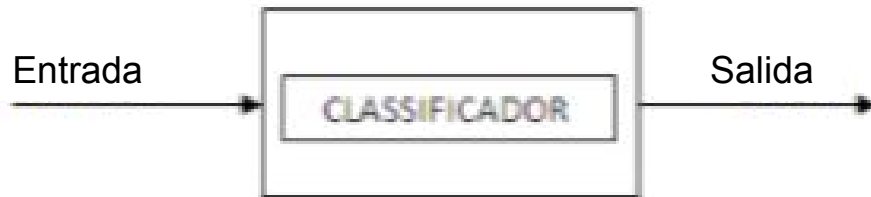


Métodos de evaluación de clasificadores I

Predicción: Para dos clases

Dos clases: mapeamiento de pertenencia al conjunto $\{p, n\}$

- Cada caso nuevo de variables X es mapeada para una de las clases utilizando la hipótesis del clasificador entrenado.
- - p = positivo ----- > 1
 - n = negativo ---- > 0



Para k clases

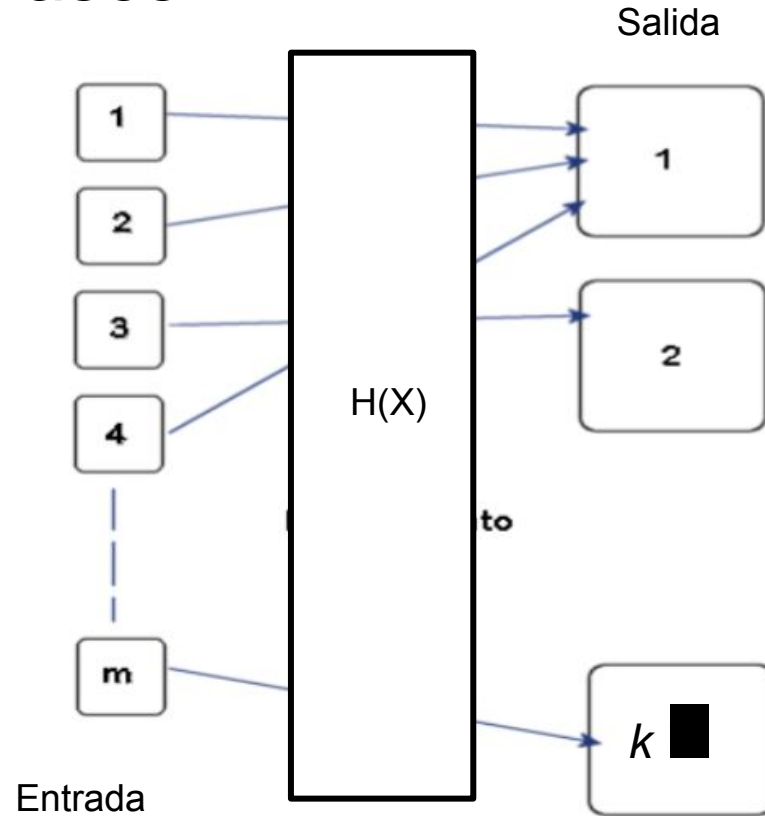


Tabla de confusión (para 2 clases)

Saída \ Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo	<u>VP(TP)</u>	FP	TSP (B)
Negativo	FN	<u>VN(TN)</u>	TSN
	TG (A)	TI	TS

- **Salida Positivo:** El clasificador clasifica como positivo el caso nuevo.
- **Salida Negativo:** El clasificador clasifica como negativo el caso nuevo.

Tabla de confusión (para 2 clases)

Saída \ Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo	<u>VP(TP)</u>	FP	TSP (B)
Negativo	FN	<u>VN(TN)</u>	TSN
	TG (A)	TI	TS

- **Verdadero Positivo (VP):** El caso nuevo es Genuíno (Positivo) y el clasificador lo clasifica como positivo
- **Verdadero Negativo (VN):** El caso nuevo es Impostor (Negativo) y el clasificador lo clasifica como negativo.

Tabla de confusión (para 2 clases)

Saída \ Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo	<u>VP(TP)</u>	FP	TSP (B)
Negativo	FN	<u>VN(TN)</u>	TSN
	TG (A)	TI	TS

- **Falso Positivo (FP):** El caso nuevo es Impostor (Negativo) y el clasificador lo clasifica como positivo
- **Falso Negativo (FN):** El caso nuevo es Genuíno (Positivo) y el clasificador lo clasifica como negativo.

Tabla de confusión (para 2 clases)

Saída \ Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo	<u>VP(TP)</u>	FP	TSP (B)
Negativo	FN	<u>VN(TN)</u>	TSN
	TG (A)	TI	TS

- **Total Genuínos (TG):** Todos los casos genuínos de la clase (VP + FN)
- **Total Impostores (TI):** Todos los casos impostores de la clase (FP + VN)

Tabla de confusión (para 2 clases)

Saída \ Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo	<u>VP(TP)</u>	FP	TSP (B)
Negativo	FN	<u>VN(TN)</u>	TSN
	TG (A)	TI	TS

- **Total Salidas Positivas (TG):** $(VP + FP)$
- **Total Salidas Negativas (TI):** $(FN + VN)$

Tabla de confusión (para 2 clases)

Saída \ Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo	<u>VP(TP)</u>	FP	TSP (B)
Negativo	FN	<u>VN(TN)</u>	TSN
	TG (A)	TI	TS

- **Total Clasificaciones Correctas (TCC):** $(VP + VN)$
- **Total Clasificaciones Incorrectas (TCI):** $(FN + FP)$
- **Total de salidas (TS):** $TG + TI = TSP + TSN = VP + FP + VN + FN$

Tabla de confusión (para 2 clases): Métricas

- **Sensibilidad (True Positive Rate):** Proporción de Verdaderos Positivos Clasificados correctamente como genuinos.

$$TPR = R = \frac{TP}{TG} = \frac{TP}{TP+FN} \quad 0 \leq TPR \leq 1$$

Saída \ Entrada	Entrada	Genuínos (Positivo)	Impostores (Negativo)	
Positivo		<u>VP(TP)</u>	FP	TSP (B)
Negativo		FN	<u>VN(TN)</u>	TSN
		TG (A)	TI	TS

Tabla de confusión (para 2 clases): Métricas

- **Razón de falsos positivos (False Positive Rate):** Proporción de Falsos Positivos clasificados en relación al total de impostores.

$$FPR = \frac{FP}{TI} = \frac{FP}{TN+FP} \qquad 0 \leq FPR \leq 1$$

Saída \ Entrada	Entrada	Genuínos (Positivo)	Impostores (Negativo)	
	Positivo	<u>VP(TP)</u>	FP	TSP (B)
	Negativo	FN	<u>VN(TN)</u>	TSN
		TG (A)	TI	TS

Tabla de confusión (para 2 clases): Métricas

- **Razón de falsos negativos (False Negative Rate):** Proporción de Falsos Negativos en relación al total de genuinos.

$$FNR = \frac{FN}{TG} = \frac{FN}{TP+FN} \quad 0 \leq FNR \leq 1$$

Saída \ Entrada	Entrada	Genuínos (Positivo)	Impostores (Negativo)	
	Positivo	<u>VP(TP)</u>	FP	TSP (B)
	Negativo	FN	<u>VN(TN)</u>	TSN
		TG (A)	TI	TS

Tabla de confusión (para 2 clases): Métricas

- **Especificidad (True Negative Rate):** Proporción de Verdaderos Negativos clasificados correctamente como impostores.

$$TNR = E = \frac{TN}{TI} = \frac{TN}{TN+FP} = 1 - FPR \quad 0 \leq E \leq 1$$

Saída \ Entrada	Entrada	Genuínos (Positivo)	Impostores (Negativo)	
	Positivo	<u>VP(TP)</u>	FP	TSP (B)
	Negativo	FN	<u>VN(TN)</u>	TSN
		TG (A)	TI	TS

Tabla de confusión (para 2 clases): Métricas

- **Precisión (P):** Proporción de verdaderos positivos clasificados.

$$P = \frac{TP}{TSP} = \frac{TP}{TP+FP} \quad 0 \leq P \leq 1$$

Saída \ Entrada	Entrada	Genuínos (Positivo)	Impostores (Negativo)	
	Positivo	<u>VP(TP)</u>	FP	TSP (B)
	Negativo	FN	<u>VN(TN)</u>	TSN
		TG (A)	TI	TS

Tabla de confusión (para 2 clases): Métricas

- **Exactitud (A):** Proporción de clasificaciones correctas del total de salidas.

$$A = \frac{TCC}{TS} = \frac{TP+TN}{TS} = \frac{TP+TN}{TG+TI} = \frac{TP+TN}{TP+FN+FP+TN} \quad 0 \leq A \leq 1$$

- **Error (Err):** Proporción de clasificaciones incorrectas en relación al total de salidas.

$$Err = 1 - A$$

Consideraciones: Sensibilidad

- Si el clasificador tiene un alto valor para Falsos Negativos (FN). Significa que clasifica como correcto a quien no debería ser clasificado en muchos casos. Por lo tanto su sensibilidad es baja.

$$TPR = R = \frac{TP}{TG} = \frac{TP}{TP+FN}$$

$$0 \leq TPR \leq 1$$

Consideraciones: Especificidad y precisión

- Si el clasificador tiene un alto valor de falsos positivos (FP) entonces su *Especificidad* y *precisión* son bajos.

$$TNR = E = \frac{TN}{TI} = \frac{TN}{TN+FP} = 1 - FPR \quad 0 \leq E \leq 1$$

Matriz de confusión (múltiples clases)

		Entrada							FPR
		Classe 1	Classe 2	Classe 3	Classe 4	-----	-----	Classe n	
Saída	Classe 1	TP	FN2/FP1	FN3/FP1	FN4/FP1	---/FP1	---/FP1	<u>FNn/FP1</u>	FPR1
	Classe 2	FN1/FP2	TP	FN3/FP2	FN4/FP2	---/FP2	---/FP2	<u>FNn/FP2</u>	FPR2
	Classe 3	FN1/FP3	FN2/FP3	TP	FN4/FP3	---/FP3	---/FP3	<u>FNn/FP3</u>	FPR3
	Classe 4	FN1/FP4	FN2/FP4	FN3/FP4	TP	---/FP4	---/FP4	<u>FNn/FP4</u>	FPR4
	-----	FN1/---	FN2/---	FN3/---	FN4/---	TP	---/---	<u>FNn/---</u>	----
	-----	FN1/---	FN2/---	FN3/---	FN4/---	---/---	TP	<u>FNn/---</u>	----
	Classe n	<u>FN1/FPn</u>	<u>FN2/FPn</u>	<u>FN3/FPn</u>	<u>FN4/FPn</u>	<u>---/FPn</u>	<u>---/FPn</u>	TP	<u>FPRn</u>
FNR		FNR1	FNR2	FNR3	FNR4	----	----	<u>FNRn</u>	TPR

$$P = \frac{TP}{TSP} = \frac{TP}{TP+FP} \rightarrow \text{Precisão}$$

$$FNR = \frac{FN}{TG} = \frac{FN}{TP+FN} \rightarrow \text{Falsa Rejeição}$$

$$TPR = R = \frac{TP}{TG} = \frac{TP}{TP+FN} \rightarrow \text{Sensibilidade}$$

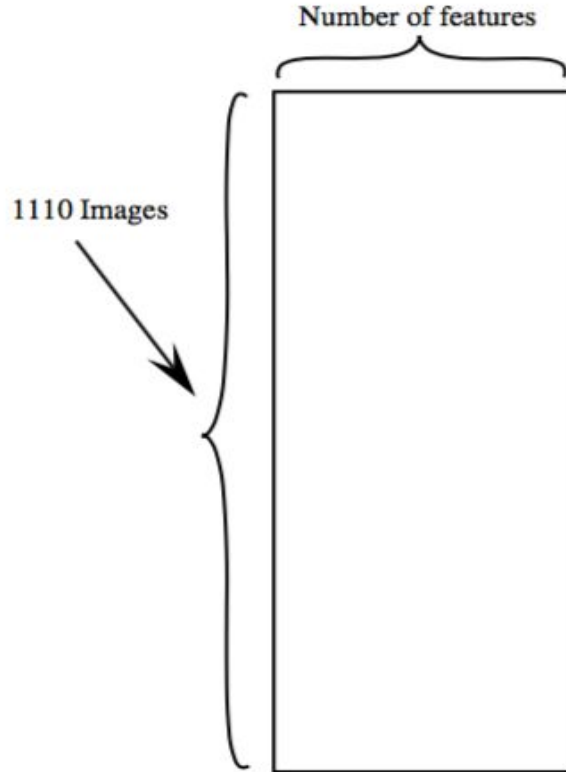
$$FPR = \frac{FP}{TI} = \frac{FP}{TN+FP} \rightarrow \text{Falsa Aceitação}$$

Matriz de confusión: Ejemplo:

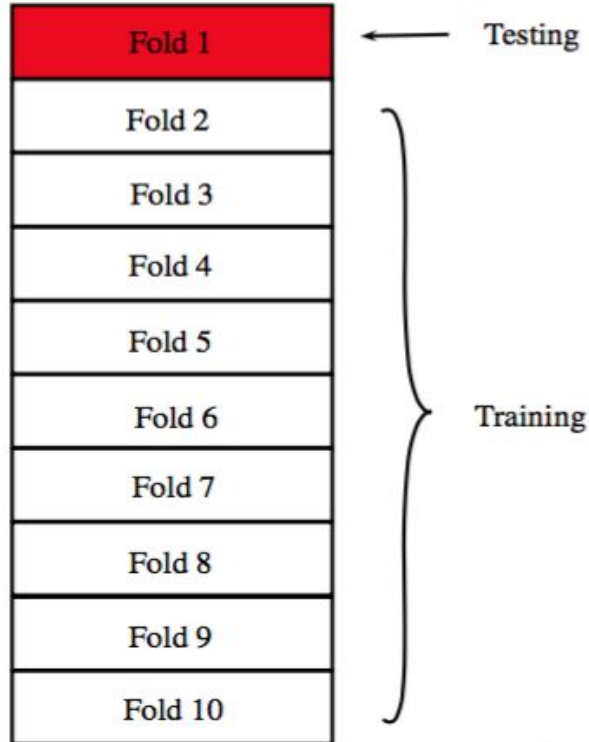
- Clase 1 = 100 cachorros
- Clase 2 = 50 conejos
- Clase 3: 80 gatos
- Clase 4: 120 lobos

		Entradas				FPR
		Cachorros	Coelhos	Gatos	Lobos	
Saída	Cachorros	80	0	2	10	0,046
	Coelhos	10	50	8	5	0,071
	Gatos	0	0	70	5	0,018
	Lobos	10	0	0	100	0,042
FNR		0,2	0	0,125	0,167	0,857

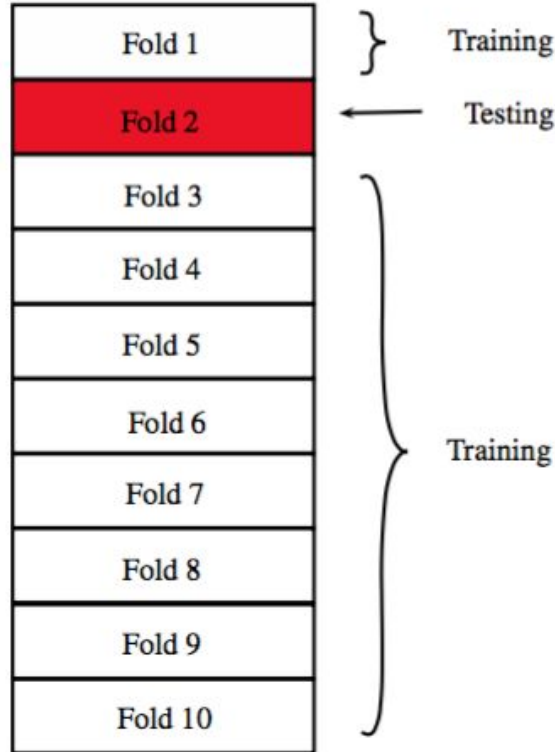
Validación Cruzada: Datos de entrada



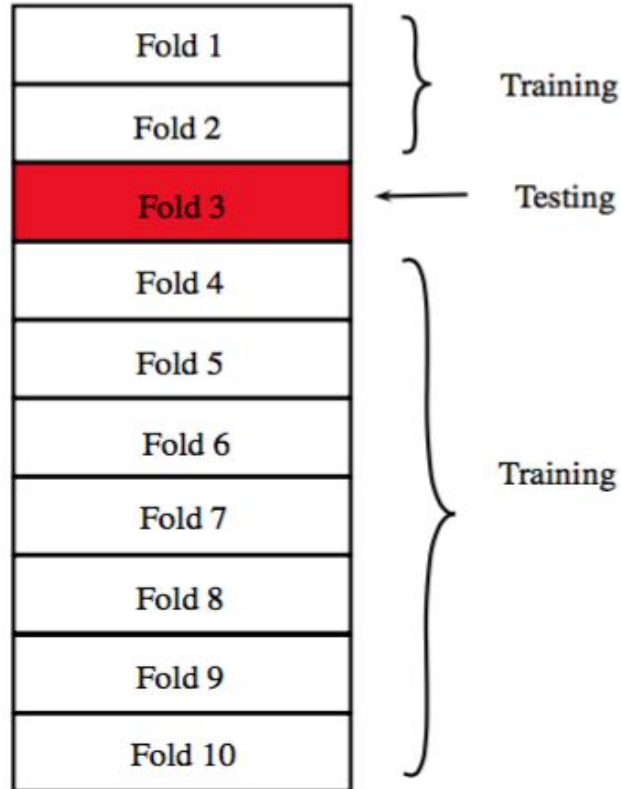
Validación Cruzada: para k-folds, $k = 10$.



Validación Cruzada: para k-folds, $k = 10$.



Validación Cruzada: para k-folds, $k = 10$.



Validación Cruzada: para k-folds, $k = 10$.

