

Práctica 5. Reducción de la dimensión.

Pedro Sánchez García.

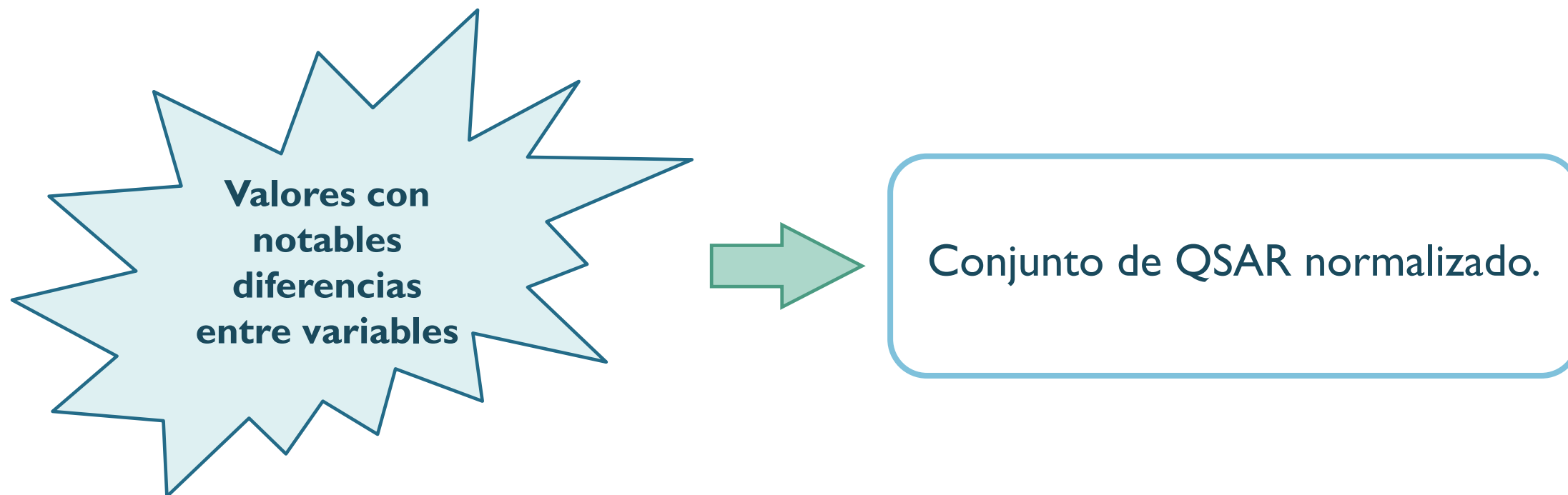
Inteligencia Computacional para Bioinformática.

MUBICS | Universidade da Coruña.



UNIVERSIDADE DA CORUÑA

PREPROCESADO DE DATOS.



APRENDIZAJE.

- Extracción de características: **Análisis de Componentes Principales.**
- Selección de características: **Relief.**
 - Número de vecinos: 10 | Número de características: 20.
- **Red neuronal:**
 - Función de transferencia: lineal | 84 unidades en la capa oculta.
 - Entrenamiento con las entradas transformadas y *targets* correspondientes.
- Método de estimación del error: **Validación cruzada.**
 - k: 5

APRENDIZAJE. Medidas de rendimiento.

ANÁLISIS DE COMPONENTES PRINCIPALES.

	Entrenamiento	Test
Sensibilidad (Recall)	$0,8267 \pm 0,1611$	$0,8157 \pm 0,1533$
Especificidad	$0,8267 \pm 0,1611$	$0,8157 \pm 0,1533$
Precisión	$0,8499 \pm 0,0633$	$0,8314 \pm 0,0675$
Valor predictivo negativo	$0,8499 \pm 0,0633$	$0,8314 \pm 0,0675$
Exactitud (accuracy)	$0,8533 \pm 0,0489$	$0,8414 \pm 0,0473$
F1-Score	$0,8223 \pm 0,1499$	$0,8116 \pm 0,1371$

RELIEF.

	Entrenamiento	Test
Sensibilidad (Recall)	$0,7407 \pm 0,1374$	$0,7425 \pm 0,1437$
Especificidad	$0,7407 \pm 0,1374$	$0,7425 \pm 0,1437$
Precisión	$0,7554 \pm 0,0758$	$0,7593 \pm 0,0866$
Valor predictivo negativo	$0,7554 \pm 0,0758$	$0,7593 \pm 0,0866$
Exactitud (accuracy)	$0,7800 \pm 0,0345$	$0,7826 \pm 0,0493$
F1-Score	$0,7459 \pm 0,1067$	$0,7486 \pm 0,1138$

COMPARACIÓN DE MODELOS.

ANÁLISIS DE COMPONENTES PRINCIPALES Y TÉCNICAS ANTERIORES:

- **F1-Score** en test:

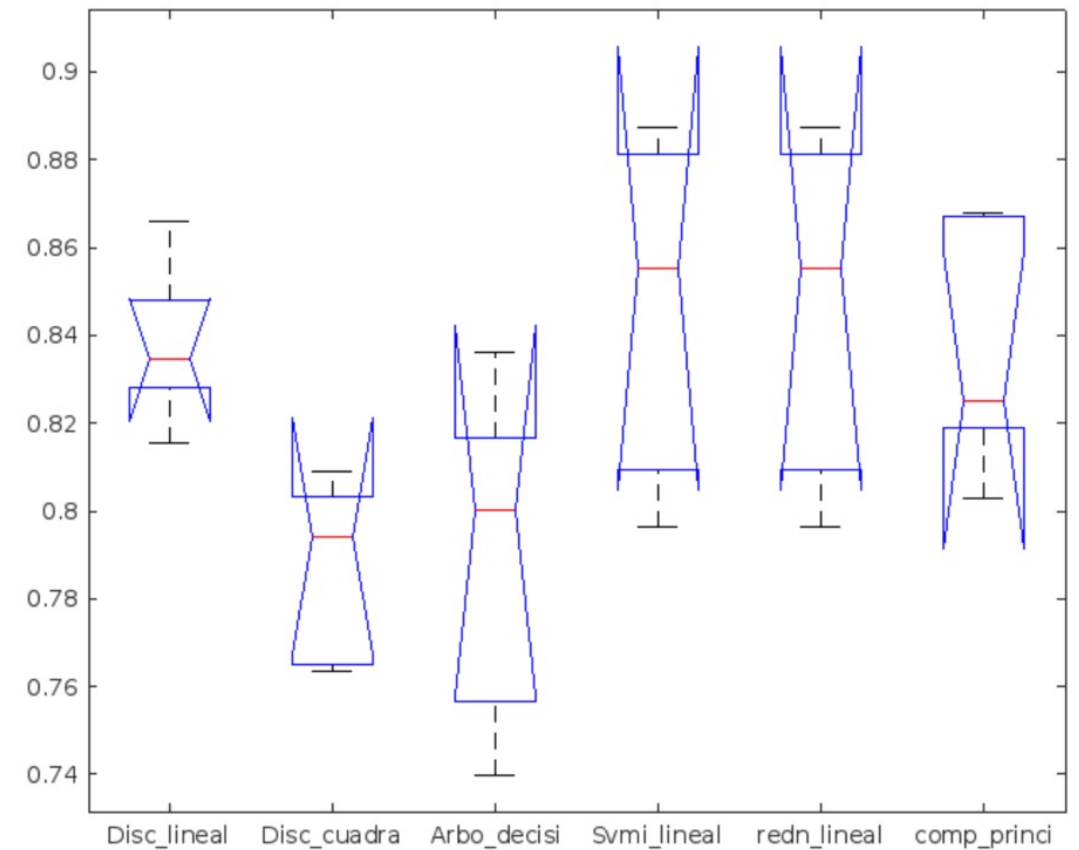
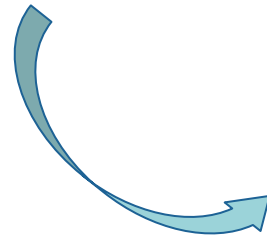
Modelo	F1-Score (media \pm desviación típica)
Discriminante lineal	0,8374 \pm 0,0770
Discriminante cuadrático	0,7768 \pm 0,0567
Árbol de decisión	0,7896 \pm 0,0385
SVM lineal	0,8578 \pm 0,0550
Red neuronal	0,8818 \pm 0,0546
Análisis de componentes principales	0,8116 \pm 0,1371

COMPARACIÓN DE MODELOS.

ANÁLISIS DE COMPONENTES PRINCIPALES Y TÉCNICAS ANTERIORES:

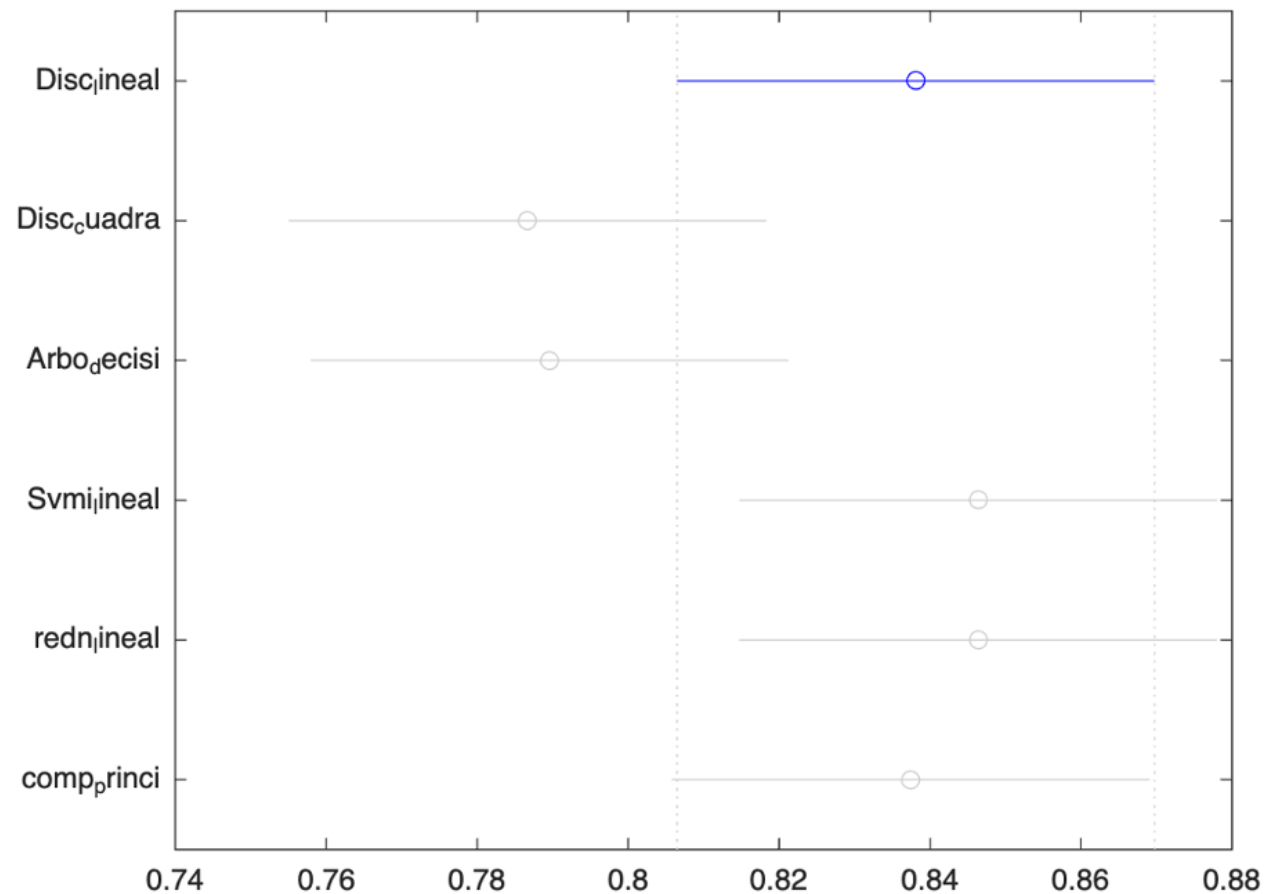
■ FI-Score en test:

Source	SS	df	MS	F	Prob>F
Columns	0.01976	5	0.00395	3.77	0.0116
Error	0.02513	24	0.00105		
Total	0.04488	29			



COMPARACIÓN DE MODELOS. Test de comparación múltiple.

ANÁLISIS DE COMPONENTES PRINCIPALES Y TÉCNICAS ANTERIORES:



COMPARACIÓN DE MODELOS.

RELIEF Y TÉCNICAS ANTERIORES:

- **F1-Score** en test:

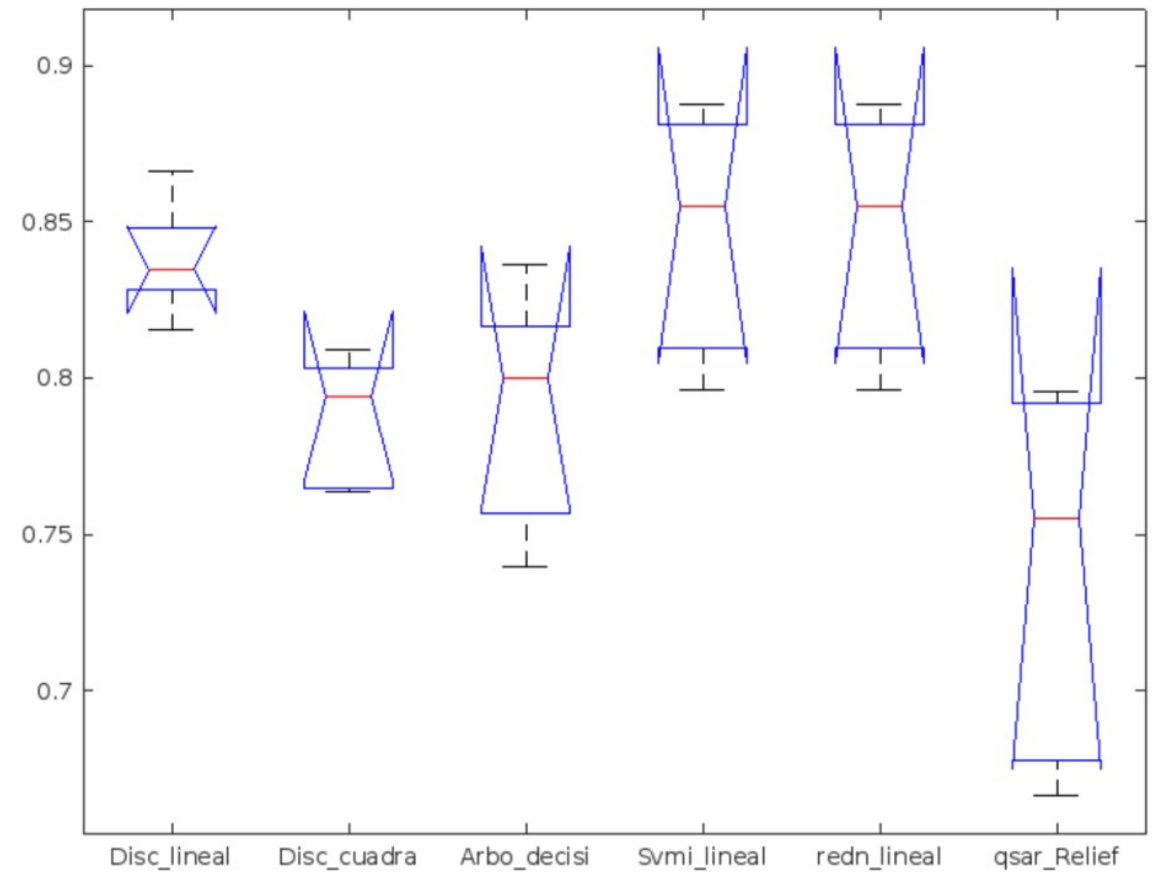
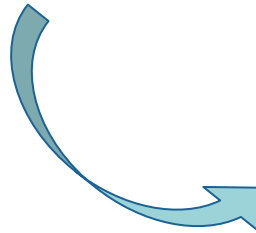
Modelo	F1-Score (media \pm desviación típica)
Discriminante lineal	0,8374 \pm 0,0770
Discriminante cuadrático	0,7768 \pm 0,0567
Árbol de decisión	0,7896 \pm 0,0385
SVM lineal	0,8578 \pm 0,0550
Red neuronal	0,8818 \pm 0,0546
Relief	0,7486 \pm 0,1138

COMPARACIÓN DE MODELOS.

RELIEF Y TÉCNICAS ANTERIORES:

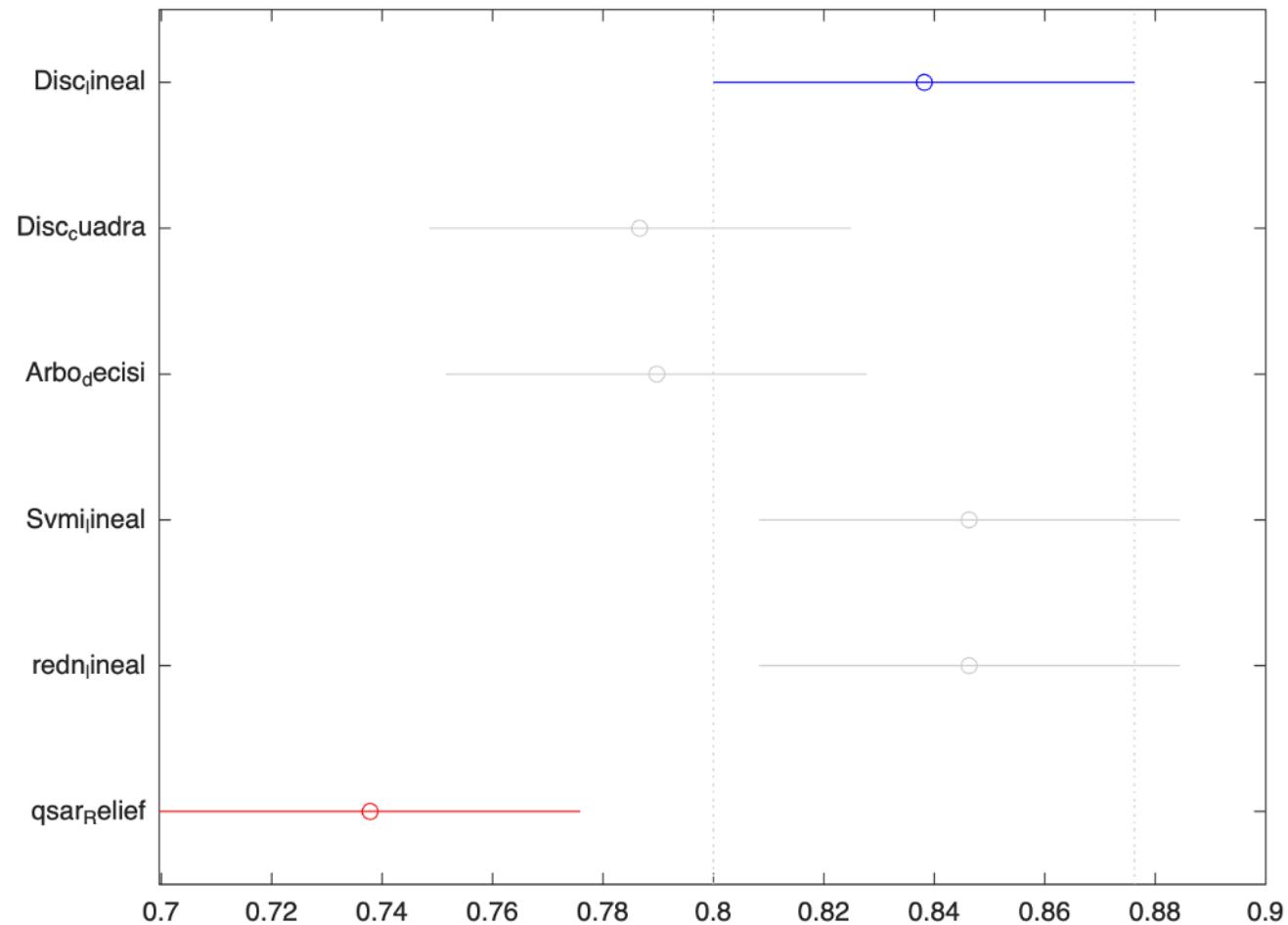
■ FI-Score en test:

Source	SS	df	MS	F	Prob>F
Columns	0.04786	5	0.00957	6.3	0.007
Error	0.03649	24	0.00152		
Total	0.08436	29			



COMPARACIÓN DE MODELOS. Test de comparación múltiple.

RELIEF Y TÉCNICAS ANTERIORES:





CONCLUSIONES.



CONCLUSIONES

ANÁLISIS DE COMPONENTES PRINCIPALES:

- ❑ Se alcanzan 22 componentes mediante la técnica de análisis de componentes principales.
- ❑ La medida de rendimiento empeora con la versión de red neuronal entrenada con estas entradas.
- ❑ Posible asociación con los resultados por la varianza mínima establecida.
- ❑ En general: mejora el rendimiento en comparación con el modelo de discriminante lineal y árbol de decisión.

CONCLUSIONES

RELIEF:

- ❑ La selección de características genera una **medida de rendimiento similar al análisis de componentes principales**.
- ❑ Se alcanzan las siguientes 10 características y pesos proporcionados por MatLab:

Característica	25	39	13	30	38	1	3	36	37	24
Peso	0,0230	0,0211	0,0171	0,0124	0,0087	0,0083	0,0063	0,0062	0,0034	0,0014

- ❑ Determinación de las **2 primeras características**, empleadas en la mitad (o más) de las ejecuciones efectuadas.
- ❑ *Futuros trabajos*: realización de más pruebas para determinar parámetros adecuados y mejorar resultados.

GRACIAS POR VUESTRA ATENCIÓN.

Pedro Sánchez García.
Inteligencia Computacional para Bioinformática.
MUBICS | Universidade da Coruña.



UNIVERSIDADE DA CORUÑA