



Master Universitario en Bioinformática para Ciencias de la Salud  
Facultad de Informática

UNIVERSIDADE DA CORUÑA



# Informe de las Prácticas 3 y 4.

---

INTELIGENCIA COMPUTACIONAL PARA BIOINFORMÁTICA.

Pedro Sánchez García | Andrea Santisteban Veiga

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA PARA CIENCIAS DE LA SALUD.  
PROFESORA: DRA. NOELIA SÁNCHEZ MAROÑO.

---

# PRÁCTICA 3. APRENDIZAJE BASADO EN *KERNELS* Y DISEÑO EXPERIMENTAL

---

---

## PARTE 1

---

## OBJETIVO.

El objetivo principal de esta práctica es alcanzar aquella combinación de máquinas de vectores soporte (SVM) binarias que permitan discriminar entre las clases tanto de Iris como de Qsar Biodegradation. Esto debe realizarse minimizando el error empírico en el entrenamiento y logrando el margen entre clases más amplio, de modo que se eviten problemas de generalización inadecuada. Hay que destacar que se utilizarán los dos conjuntos de datos normalizados obtenidos en las prácticas anteriores.

De esta forma, se entrenan 3 versiones de SVM ajustando varios parámetros asociados y se alcanza una medida de rendimiento, que, junto con el correspondiente test estadístico, permitirá decidir la versión más adecuada.

## APRENDIZAJE.

### CONFIGURACIÓN DEL MÉTODO DE APRENDIZAJE.

#### SVMs MULTICLASE.

La extensión de las SVMs a problemas multiclase implica una combinación de SVMs binarias, con dos posibles aproximaciones: *Uno contra el resto* y *uno contra uno*. La primera consiste en la comparación de una determinada clase con el resto, de modo que se procede a la asignación de etiquetas correspondientes en el proceso de clasificación mediante el uso de un esquema de votación entre los  $k$  clasificadores binarios posibles.

Con respecto al método *una contra uno*, en el que se basa la presente práctica, se construyen  $k(k-1)/2$  hiperplanos que separan una clase de otra con la que se compara. Posteriormente, para cada par  $(i,j)$  de etiquetas de la clase, se establece un clasificador binario, clasificando los puntos de las clases  $i$  y  $j$ . Debemos tener en cuenta un aspecto interesante, que consiste en que este método, a diferencia del *uno contra el resto*, mantiene el balance en torno al análisis comparativo. Por tanto, lo habitual es que se elija el *uno contra uno* para problemas con SVMs multiclase, como es el caso de Iris.

Por otro lado, dado que se trabaja con varias versiones de SVMs, es preciso tener en cuenta un conjunto de conceptos y parámetros, que se exponen brevemente en las siguientes secciones:

#### SVMs LINEALES.

A partir de un conjunto de  $N$  datos de entrenamiento que se encuentra dividido en dos clases, se persiguen aquellos hiperplanos que permitan la separación de las clases minimizando el error empírico del entrenamiento y lograr un margen amplio que evite una generalización inapropiada.

Por tanto, se determinarán dos rectas que actúan como extremos del margen en la separación de clases. Dichas rectas constan de las entradas proporcionadas ( $x_i$ ), unos pesos asignados ( $w^T$ ) y unas guías ( $b$ ) para su formación:

$$H1 = w^T x_i + b \geq 1 \text{ para } y_i = 1$$

$$H2 = w^T x_i + b \leq -1 \text{ para } y_i = -1$$

De acuerdo con la función de coste o error, se trata de maximizar el margen y minimizar los pesos sujetos al hiperplano. No obstante, un aspecto de gran relevancia es que la mayoría de los problemas, como los tratados en esta práctica, no son linealmente separables, de modo que establecer el margen resulta complejo. Por tanto, la ecuación sobre la que se basa es la siguiente:

$$E(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

Se centra la atención en el parámetro C, que se encuentra relacionado con el margen flexible del clasificador, por lo que junto con el parámetro  $\xi$ , se penalizan errores de clasificación. La tendencia es que, con valores elevados de C, se incrementa la restricción en torno a errores de clasificación, es decir, que prácticamente no cometa errores. En cambio, con un valor de C próximo a 0, se permite un mayor margen, el cual implica a su vez un mayor error de clasificación en el entrenamiento. En este caso, teniendo en cuenta lo anterior y tomando como referencia el valor de 10 en la  $k$ -fold, se plantea la SVM lineal bajo un parámetro C de 10.



#### SVMs NO LINEALES.

Tal y como se ha mencionado anteriormente, la mayoría de los problemas reales no son linealmente separables, de modo que los clasificadores lineales son muy limitados. Con el fin de mantener su uso, por medio de una transformación no lineal, se plantea un espacio de características en lugar del espacio original de los datos.

En este cambio en el espacio, entra en juego la función *kernel* ( $K$ ), que permite producir regiones de decisión interesantes en dicho espacio de forma sencilla incluso para grandes dimensiones. Para esta práctica, se emplean la función *kernel* de tipo gaussiana (RBF) con  $\sigma = 1$  y de tipo polinomial de grado 2. En ambos casos, el parámetro C tiene un valor de 10.



#### MÉTODO DE ESTIMACIÓN DEL ERROR: VALIDACIÓN CRUZADA.

Al igual que en las anteriores prácticas, para llevar a cabo la estimación del error del modelo se opta por la validación cruzada tanto para el conjunto de Iris como para el de Qsar. Así, se calculan los errores sobre el conjunto de test, con un proceso que se itera  $k$  veces y consta de 3 repeticiones, tal y como se ha establecido. En el caso del conjunto Iris  $k = 10$  y para Qsar  $k = 5$ .

#### MEDIDAS DE RENDIMIENTO.

Para las versiones de SVM lineal, SVM no lineal con *kernel* polinómico y gaussiano planteadas para ambos conjuntos de datos, se calculan las medidas de rendimiento analizadas en las prácticas anteriores, tanto de entrenamiento como de test: sensibilidad (recall), especificidad, precisión, valor predictivo negativo, exactitud (accuracy) y F1-Score.

#### SELECCIÓN DE MEDIDA DE RENDIMIENTO PARA EL TEST ESTADÍSTICO.

Para la realización del test estadístico para comparar las SVMs planteadas para cada conjunto de datos, se utilizará como medida de rendimiento F1-Score.

# CONJUNTO DE DATOS: IRIS

## RESULTADOS.

### MEDIDAS DE RENDIMIENTO.

#### SVM LINEAL.

Tabla 1. Medidas de rendimiento global para entrenamiento y test en la SVM lineal del conjunto de datos Iris.

	Entrenamiento	Test
Sensibilidad (Recall)	0,9758 ± 0,0280	0,9600 ± 0,0909
Especificidad	0,9879 ± 0,0014	0,9800 ± 0,0455
Precisión	0,9766 ± 0,0264	0,9648 ± 0,0789
Valor predictivo negativo	0,9881 ± 0,0136	0,9814 ± 0,0419
Exactitud (accuracy)	0,9839 ± 0,0125	0,9733 ± 0,0477
F1-Score	0,9758 ± 0,0188	0,9595 ± 0,0729

Tal y como se refleja en la Tabla 1, se alcanzan unos resultados similares en las medidas de rendimiento para entrenamiento y test de forma global. Dado que, en el caso de las medidas para entrenamiento, los valores son próximos a 1, nos muestra un funcionamiento adecuado de la SVM lineal planteada. Debemos tener en cuenta que ciertas medidas como la exactitud, que nos proporciona la tasa de acierto global del sistema, no sería adecuada en aquellos conjuntos que no se encuentren balanceados. De esta forma, se atiende a la sensibilidad y especificidad, que nos informa de la probabilidad de que un caso positivo se clasifique como positivo o que un caso negativo sea clasificado como negativo, respectivamente. En este caso, para ambas medidas, los valores son próximos a 1, aunque más elevada en el caso de la especificidad. En términos generales, la SVM lineal que se ha generado ofrece un comportamiento adecuado ante el problema de clasificación en Iris.

## SVM NO LINEAL CON *KERNEL* POLINÓMICO.

Tabla 2. Medidas de rendimiento global para entrenamiento y test en la SVM con *kernel* polinómico del conjunto de datos Iris.

	Entrenamiento	Test
Sensibilidad (Recall)	0,9916 ± 0,0108	0,9511 ± 0,1052
Especificidad	0,9958 ± 0,0054	0,9756 ± 0,0526
Precisión	0,9917 ± 0,0011	0,9597 ± 0,0819
Valor predictivo negativo	0,9958 ± 0,0053	0,9780 ± 0,0456
Exactitud (accuracy)	0,9944 ± 0,0054	0,9674 ± 0,0482
F1-Score	0,9916 ± 0,0079	0,9499 ± 0,0772

Con respecto a la SVM no lineal con *kernel* polinómico, se aprecia una tendencia similar a la de la SVM lineal (Tabla 2). No obstante, si comparamos en el caso de entrenamiento, los valores de las medidas incrementan notablemente, lo que pone de manifiesto un funcionamiento adecuado de esta SVM. De nuevo, si nos fijamos en las medidas de sensibilidad y especificidad en test, a pesar de que no se produzcan cambios muy notables, estas indican que el patrón de esta SVM planteada también resulta adecuado para la clasificación de Iris ante nuevos datos proporcionados.

## SVM NO LINEAL CON *KERNEL* GAUSSIANO.

Tabla 3. Medidas de rendimiento global para entrenamiento y test en la SVM no lineal, *kernel* gaussiano del conjunto de datos Iris.

	Entrenamiento	Test
Sensibilidad (Recall)	0,9938 ± 0,0100	0,9400 ± 0,0106
Especificidad	0,9969 ± 0,0050	0,9700 ± 0,0570
Precisión	0,9940 ± 0,0098	0,9512 ± 0,0896
Valor predictivo negativo	0,9969 ± 0,0049	0,9728 ± 0,0471
Exactitud (accuracy)	0,9959 ± 0,0037	0,9600 ± 0,0477
F1-Score	0,9938 ± 0,0055	0,9394 ± 0,0729

En lo que respecta a la SVM no lineal con *kernel* gaussiano, se observan diferencias de interés en los resultados de las medidas de rendimiento (Tabla 3). En comparación con las otras SVMs, los valores obtenidos se reducen. Atendiendo a las medidas de sensibilidad y especificidad, los cambios informan de que el comportamiento para esta SVM resulta peor para la clasificación de Iris ante nuevos datos proporcionados.

## COMPARACIÓN DE MODELOS.

### RESULTADOS PARA LOS MODELOS IMPLICADOS. SVMs

A partir de las medias del F1-Score en test para cada SVM, se procede al test estadístico para su comparación. Se muestran a continuación las medias y correspondientes desviaciones típicas en cada caso:

Tabla 4. Media y desviación típica de la medida de rendimiento F1-Score para las SVMs planteadas del conjunto de datos Iris.

Modelo	F1-Score (media $\pm$ desviación típica)
SVM lineal	0,9595 $\pm$ 0,0729
SVM no lineal ( <i>kernel polinómico</i> )	0,9499 $\pm$ 0,0772
SVM no lineal ( <i>kernel gaussiano</i> )	0,9394 $\pm$ 0,0729

Bajo un valor crítico de 0.10, se determina si las muestras correspondientes a la medida de rendimiento en las SVMs resultan o no estadísticamente similares. A continuación, se muestra la tabla ANOVA obtenida del contraste no paramétrico de Kruskal-Wallis:

Tabla 5. Tabla ANOVA de F1-Score para las SVMs en el conjunto de datos Iris.

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	36.45	2	18.225	0.58	0.7483
Error	1786.05	27	66.15		
Total	1822.5	29			

Dado que el p-valor obtenido es superior al valor crítico de 0.10 establecido, se mantiene la hipótesis nula de ausencia de diferencias estadísticamente significativas entre la versión de SVM lineal y no lineales con *kernel* polinómico y gaussiano respectivamente.

### DIAGRAMA DE CAJAS Y BIGOTES.

Mediante el test estadístico, también se proporciona el resultado correspondiente al diagrama de cajas y bigotes para el F1-Score sobre el conjunto de test como medida elegida para la comparación de SVMs:

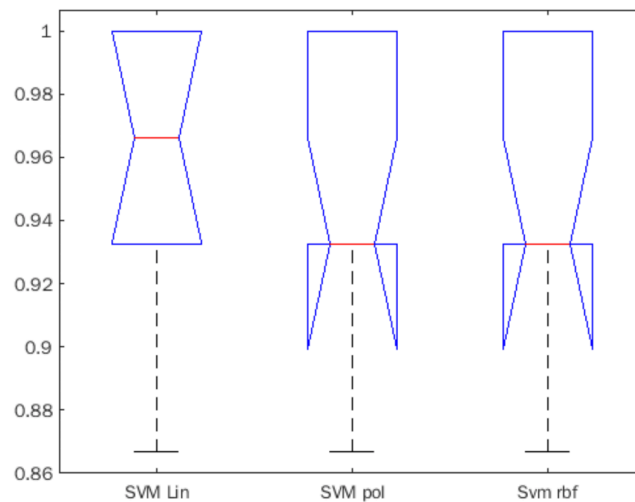


Figura 1. Diagrama de cajas y bigotes para F1-Score en las SVMs del conjunto de datos Iris.

Del diagrama obtenido, es preciso tener en cuenta que la SVM lineal presenta un rango intercuartílico más reducido en comparación con las SVMs no lineales planteadas. Además, se aprecia la tendencia en la mediana con respecto a la medida de rendimiento analizada. En general, este resultado muestra la ausencia de diferencias estadísticamente significativas reflejadas con el contraste de Kruskal-Wallis. En consecuencia, se determina que la SVM lineal es la mejor aproximación entre las SVMs, de tal forma que se elige esta como la versión que se compara con las técnicas vistas en las anteriores prácticas.

## DISCUSIÓN Y CONCLUSIONES.

Se han planteado versiones de SVM lineal, así como, no lineal con *kernel gaussiano* y *kernel* polinómico respectivamente para el conjunto de Iris del UCI Machine Learning Repository. El planteamiento de las aproximaciones y la posterior estimación del error real de estas mediante validación cruzada conducen a unas medidas de rendimiento que no varían de forma estadísticamente significativa entre sí, tal y como se verifica con el test estadístico a través del contraste de Kruskal-Wallis y el diagrama de cajas y bigotes correspondiente. De esta forma, teniendo en cuenta las medidas del F1-Score analizadas, se elige la versión de SVM lineal como la más adecuada para el tratamiento de este problema de clasificación de Iris.



# CONJUNTO DE DATOS: QSAR BIODEGRADATION

## RESULTADOS.

### MEDIDAS DE RENDIMIENTO.

#### SVM LINEAL.

Tabla 6. Medidas de rendimiento global para entrenamiento y test de la SVM lineal del conjunto de datos Qsar.

	Entrenamiento	Test
Sensibilidad (Recall)	0,8750 $\pm$ 0,0559	0,8554 $\pm$ 0,0652
Especificidad	0,8750 $\pm$ 0,0559	0,8554 $\pm$ 0,0652
Precisión	0,8828 $\pm$ 0,0294	0,8609 $\pm$ 0,0510
Valor predictivo negativo	0,8828 $\pm$ 0,0294	0,8609 $\pm$ 0,0510
Exactitud (accuracy)	0,8925 $\pm$ 0,0060	0,8736 $\pm$ 0,0231
F1-Score	0,8787 $\pm$ 0,0424	0,8578 $\pm$ 0,0550

Los resultados de las medidas de rendimiento son similares para entrenamiento y test de forma global (Tabla 6). No obstante, se puede observar que las medidas en el caso del entrenamiento no presentan valores próximos a 1, lo que parece indicar que no se produce una generalización adecuada y, en consecuencia, un entrenamiento adecuado del modelo. Esto conduce a unos valores de las medidas para el caso de test que también se distancian de 1.

Al igual que se hizo en el caso del conjunto de Iris, centramos nuestra atención en la sensibilidad y especificidad. En base a estas, no podemos afirmar que la SVM lineal proporcione un buen comportamiento para la clasificación de los compuestos químicos.

## SVM NO LINEAL CON KERNEL POLINÓMICO.

Tabla 7. Medidas de rendimiento global para entrenamiento y test de la SVM no lineal (kernel polinómico) del conjunto Qsar.

	Entrenamiento	Test
Sensibilidad (Recall)	0,9954 $\pm$ 0,0041	0,8118 $\pm$ 0,0510
Especificidad	0,9954 $\pm$ 0,0041	0,8118 $\pm$ 0,0510
Precisión	0,9970 $\pm$ 0,0021	0,8017 $\pm$ 0,0889
Valor predictivo negativo	0,9970 $\pm$ 0,0021	0,8017 $\pm$ 0,0889
Exactitud (accuracy)	0,9967 $\pm$ 0,0011	0,8224 $\pm$ 0,0246
F1-Score	0,9966 $\pm$ 0,0018	0,8052 $\pm$ 0,0646

Para la SVM no lineal con *kernel* polinómico, se aprecian valores más elevados en las medidas de rendimiento para entrenamiento en comparación con la SVM lineal (Tabla 7). No obstante, en comparación con la SVM lineal, se reducen notablemente los valores en el caso de test. Atendiendo a las medidas de sensibilidad y especificidad en test, estas nos informan de un peor comportamiento de esta SVM para la clasificación de compuestos ante nuevos datos de entrada.

## SVM NO LINEAL CON KERNEL GAUSSIANO.

Tabla 8. Medidas de rendimiento global para entrenamiento y test de la SVM no lineal (kernel gaussiano) del conjunto de datos Qsar.

	Entrenamiento	Test
Sensibilidad (Recall)	0,9951 $\pm$ 0,0049	0,6762 $\pm$ 0,2975
Especificidad	0,9951 $\pm$ 0,0049	0,6762 $\pm$ 0,2975
Precisión	0,9969 $\pm$ 0,0023	0,8048 $\pm$ 0,0563
Valor predictivo negativo	0,9969 $\pm$ 0,0023	0,8048 $\pm$ 0,0563
Exactitud (accuracy)	0,9964 $\pm$ 0,0012	0,7703 $\pm$ 0,0176
F1-Score	0,9960 $\pm$ 0,0020	0,6886 $\pm$ 0,1672

En lo que respecta a la SVM no lineal con *kernel* gaussiano, se aprecia una tendencia similar a la SVM con *kernel* polinómico en torno a las medidas de rendimiento para entrenamiento (Tabla 8). Por otra parte, en test, hay una drástica reducción de las medidas en comparación con el resto de SVMs. Si nos centramos en las medidas de sensibilidad y especificidad en test, estas nos informan

claramente de un peor comportamiento de esta SVM para la clasificación de compuestos ante nuevos datos de entrada en comparación con las otras versiones.

## COMPARACIÓN DE MODELOS.

### RESULTADOS PARA LOS MODELOS IMPLICADOS. SVMs

Se muestran las medias y correspondientes desviaciones típicas para las medidas del F1-Score empleadas en el test estadístico para la comparación de las SVMs entre sí:

Tabla 9. Media y desviación típica de la medida de rendimiento F1-Score en test para comparar la SVMs en el conjunto de datos Qsar.

Modelo	F1-Score (media $\pm$ desviación típica)
SVM lineal	0,8578 $\pm$ 0,0550
SVM no lineal ( <i>kernel polinómico</i> )	0,8052 $\pm$ 0,0646
SVM no lineal ( <i>kernel gaussiano</i> )	0,6886 $\pm$ 0,1672

Para el test estadístico se estableció un valor crítico de 0.10, que determina si las muestras correspondientes a las medidas de rendimiento en los modelos resultan o no similares estadísticamente. A continuación, se muestra la correspondiente tabla resultante del ANOVA de 1 vía:

Tabla 10. Tabla ANOVA de F1-Score en test para la comparación de SVMs en el conjunto de datos Qsar.

Source	SS	df	MS	F	Prob>F
Columns	0.07969	2	0.03985	42.11	3.76423e-06
Error	0.01136	12	0.00095		
Total	0.09105	14			

El p-valor resultante es inferior al valor crítico de 0.10 establecido, de tal manera que se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las versiones de SVMs.

## DIAGRAMA DE CAJAS Y BIGOTES.

La existencia de diferencias estadísticamente significativas se verifica posteriormente con la representación del diagrama de cajas y bigotes para el F1-Score sobre el conjunto de test como medida para la comparación de modelos:

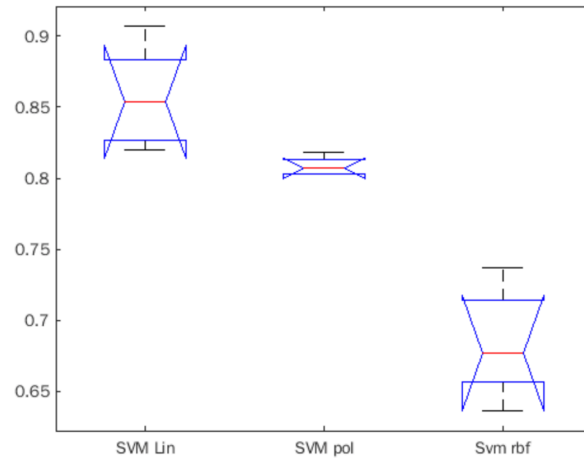


Figura 2. Diagrama de cajas y bigotes para F1-Score en test de las SVMs del conjunto de datos Qsar.

Del diagrama, se distingue principalmente que la SVM lineal presenta la mayor mediana y rango intercuartílico para el F1-Score empleado en la comparación. En el caso de la SVM con *kernel* polinómico, la mediana se reduce, así como el rango intercuartílico de forma drástica en comparación con la SVM lineal. Por otro lado, con la SVM con *kernel* gaussiano, se logra el resultado más bajo y el rango intercuartílico es similar al observado en la SVM lineal.

### TEST DE COMPARACIÓN MÚLTIPLE.

La existencia de diferencias estadísticamente significativas entre las SVMs da lugar a una representación de un test de comparación múltiple, como un complemento al diagrama de cajas y bigotes:

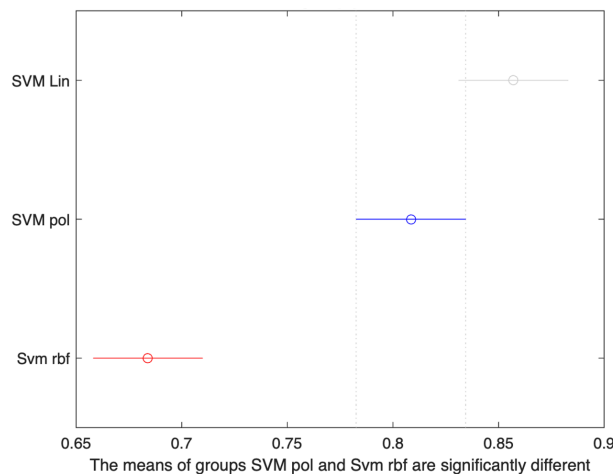


Figura 3. Representación del test de comparación múltiple para las SVMs en el conjunto de datos Qsar.

Este nos muestra la tendencia observada anteriormente, distinguiendo claramente las SVMs entre las que se producen las mayores diferencias que, en este caso, son la SVM lineal y la SVM con *kernel* gaussiano.

## DISCUSIÓN Y CONCLUSIONES.

Con el conjunto de datos QSAR se alcanzan resultados de interés en la presente práctica. Las medidas de rendimiento en las SVMs entrenadas varían de forma estadísticamente significativa al compararlas entre sí, de forma que se elige la SVM lineal como la mejor aproximación entre las SVM.

---

# PRÁCTICA 3. APRENDIZAJE BASADO EN *KERNELS* Y DISEÑO EXPERIMENTAL

---

---

## PARTE 2

---

# SERIE DE HENON

## PREPROCESADO DE DATOS.

Para esta parte 2 de la práctica 3, se toma como punto de partida un conjunto de datos que corresponde con la serie de Henon. En este caso, nos encontramos ante la aproximación de las SVMs para la resolución de problemas de regresión, donde perseguimos que todos los datos de un conjunto permanezcan delimitados en el interior de un margen determinado:

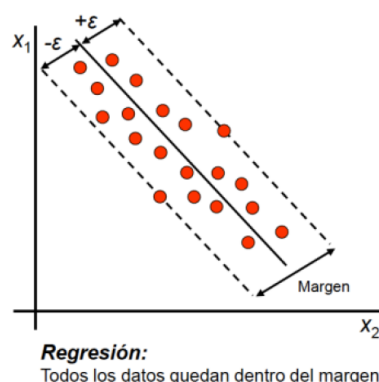


Figura 4. Representación del problema de regresión de las SVMs.

De esta forma, tratamos aquel hiperplano que mejor describa los datos por medio de las SVMs y los márgenes evaluados.

La serie de Henon, que consta de 300.000 muestras, se puede describir en términos generales como un conjunto de datos que conforma una señal a predecir o modelar, de tal forma que se necesita como primera fase, un análisis general de los datos. En este, se aprecia el notable tamaño muestral, que conduce a la selección de 1.500 muestras, con el fin de evitar problemas asociados al tiempo empleado para el correspondiente entrenamiento.

Posteriormente, se generan 2 conjuntos de datos, de modo que la señal tratada (dimensión  $1 \times N$ ) se transforma en muestras de varias dimensiones empleando la técnica de “ventanas deslizantes”. Es aquí cuando entra en juego el parámetro  $d$ , que corresponde con el tamaño de ventana. A través de este, se emplea ese número de variables de entrada, para generar las correspondientes salidas en un proceso iterativo por medio del desplazamiento de la ventana. En consecuencia, estamos tratando una matriz de datos con dimensión  $d \times (N - d)$ , lo que sirve como una referencia en la decisión del parámetro  $d$ .

Finalmente, en esta fase de preparación de los datos, cabe destacar que se aprecia una diferencia en magnitud de los valores en estos, lo que implica la normalización, para reducir la dispersión y la aparición del sesgo en el modelo de aprendizaje.

## APRENDIZAJE.

### CONFIGURACIÓN DEL MÉTODO DE APRENDIZAJE.

Como se ha mencionado anteriormente, han de determinarse dos valores para el parámetro  $d$ , con el que se lleva a cabo la configuración de las SVMs y los conjuntos de entrenamiento y test.

Para esta etapa, se opta en primer lugar, por la configuración de las versiones de SVMs que se van a emplear, llevando a cabo una serie de pruebas modificando los valores del parámetro  $d$  y eligiendo aquellos que lleven a unos resultados coherentes en la medida de rendimiento. Con esto, se trata de que no haya valores elevados que podríamos considerar como resultados anómalos.

Por tanto, se plantean las siguientes versiones de SVMs, sobre las que se llevan a cabo las evaluaciones correspondientes:

Tabla 11. Versiones de SVMs planteadas para el conjunto de Henon bajo los parámetros  $d$  establecidos.

Versión de SVM	Descripción general
SVM lineal	Parámetro $C = 10$
SVM no lineal ( <i>kernel polinómico</i> )	Parámetro $C = 10$   grado de polinomio = 2
SVM no lineal ( <i>kernel gaussiano</i> )	Parámetro $C = 10$   $\sigma = 1$



En base a las pruebas efectuadas, se opta finalmente por unos valores de 15 y 35 en el parámetro  $d$ , pues nos permiten alcanzar unos resultados en error cuadrático medio que poseen una magnitud y tendencia similar entre entrenamiento y test. De este modo, tenemos constancia de que no se produce un sobreajuste en el proceso.

### MÉTODO DE ESTIMACIÓN DEL ERROR: VALIDACIÓN CRUZADA.

Con el fin de realizar la estimación del error de cada una de las aproximaciones realizadas, se determina la validación cruzada con  $k=10$  subconjuntos disjuntos y 3 iteraciones.

### MEDIDAS DE RENDIMIENTO.

En esta práctica, la medida de rendimiento analizada es el error cuadrático medio, cuyo fundamento nos permite conocer la diferencia exacta que existe entre la salida real ( $y_i$ ) y esperada ( $d_i$ ). A continuación, se muestra la fórmula, donde  $N$  es el número de ejemplos para entrenamiento/test:

$$ECM = \frac{1}{N} \sum_{i=1}^N (y_i - d_i)$$

Cabe destacar que se persigue un error cuadrático medio con un valor próximo a 0, lo que nos muestra una tendencia adecuada para el problema tratado.



## COMPARACIÓN DE MODELOS.

### RESULTADOS PARA LOS MODELOS IMPLICADOS.

Se muestran las medias y correspondientes desviaciones típicas alcanzadas para las medidas del error cuadrático medio en entrenamiento y test. Estas medidas se emplean en el test estadístico para la comparación de las SVMs entre sí:

Tabla 12. Media y desviación típica del error cuadrático medio para entrenamiento y test bajo los parámetros d establecidos.

Modelo	Error cuadrático medio (media $\pm$ desviación típica)	
	Conjunto: d = 15	Conjunto: d = 35
SVM lineal	Entrenamiento: 0,8061 $\pm$ 0,0087	Entrenamiento: 0,8019 $\pm$ 0,0110
	Test: 0,8254 $\pm$ 0,1286	Test: 0,8388 $\pm$ 0,1133
SVM no lineal (kernel polinómico)	Entrenamiento: 0,0069 $\pm$ 0,0002	Entrenamiento: 0,0086 $\pm$ 0,0002
	Test: 0,0070 $\pm$ 0,0004	Test: 0,0188 $\pm$ 0,0025
SVM no lineal (kernel gaussiano)	Entrenamiento: 0,0262 $\pm$ 0,0007	Entrenamiento: 0,1232 $\pm$ 0,0293
	Test: 0,23375 $\pm$ 0,0381	Test: 1,0062 $\pm$ 0,1075

En la realización del test estadístico, se establece un valor crítico de 0.10, para determinar si las medidas de error cuadrático medio en test resultan o no similares estadísticamente para las versiones de SVMs generadas en los conjuntos.

A continuación, se muestra la correspondiente tabla resultante del ANOVA de 1 vía para el conjunto de datos bajo un parámetro d = 15:

Tabla 13. Tabla ANOVA empleando error cuadrático medio en test para el conjunto de datos con d = 15.

Source	SS	df	MS	F	Prob>F
Columns	3.55679	2	1.77839	334.74	8.79369e-20
Error	0.14345	27	0.00531		
Total	3.70023	29			

Tal y como se puede observar, el p-valor resultante es inferior al valor crítico de 0.10 establecido, de tal forma que se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las versiones de SVMs analizadas para este conjunto.

En lo que respecta al conjunto de datos mediante un parámetro d = 35, se logra la siguiente tabla resultante del ANOVA de 1 vía:

Tabla 14. Tabla ANOVA empleando error cuadrático medio en test para el conjunto de datos con  $d = 35$ .

Source	SS	df	MS	F	Prob>F
Columns	5.59282	2	2.79641	245.13	4.87796e-18
Error	0.30801	27	0.01141		
Total	5.90083	29			

Al igual que en el caso anterior, para este conjunto, el p-valor obtenido es notablemente inferior al valor crítico de 0.10 establecido, de tal forma que existen diferencias estadísticamente significativas entre las versiones de SVMs evaluadas.

### DIAGRAMA DE CAJAS Y BIGOTES.

Para ambos conjuntos planteados, se verifica posteriormente la existencia de diferencias estadísticamente significativas mediante la representación del diagrama de cajas y bigotes para el error cuadrático medio sobre el conjunto de test como medida:

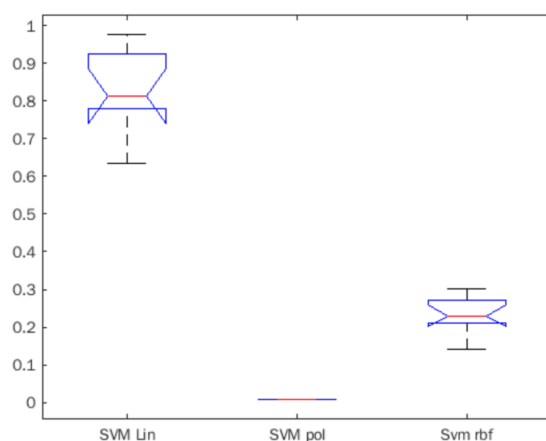


Figura 5. Diagrama de cajas y bigotes para error cuadrático medio del conjunto de datos con  $d = 15$ .

En el caso del conjunto de datos con  $d = 15$  (Figura 5), existe un mayor rango intercuartílico en el caso de las SVMs lineal y no lineal con *kernel* gaussiano si se compara con la SVM no lineal con *kernel* polinómico. Se distingue la mayor mediana del error cuadrático medio en la SVM lineal, con una drástica reducción en la SVM no lineal con *kernel* polinómico.

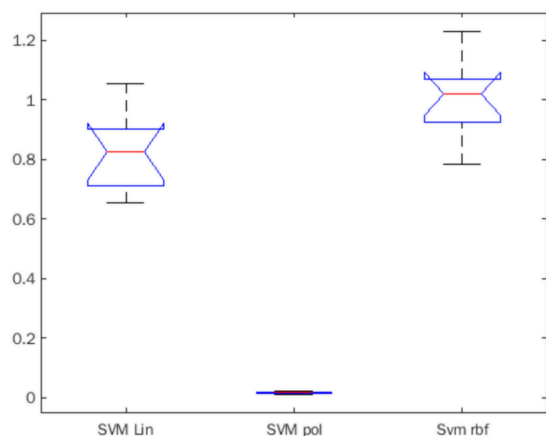


Figura 6. Diagrama de cajas y bigotes para error cuadrático medio del conjunto de datos con  $d = 35$ .

Por otra parte, en el conjunto de datos con  $d = 35$  se produce un ligero cambio de tendencia. En la Figura 6 se puede observar el valor notablemente más elevado para el error cuadrático medio para la SVM no lineal con *kernel* gaussiano, así como un patrón similar en el valor de la SVM lineal. De nuevo, al igual que en el caso anterior, con la SVM con *kernel* polinómico se alcanza el menor rango intercuartílico y el valor más bajo en la medida.

### TEST DE COMPARACIÓN MÚLTIPLE.

Para los conjuntos planteados, se visualiza el test de comparación múltiple, donde se distinguen aquellas aproximaciones entre las que se producen diferencias estadísticamente significativas. Para el conjunto con  $d = 15$ , estas se producen al comparar todas las versiones de SVM entre sí:

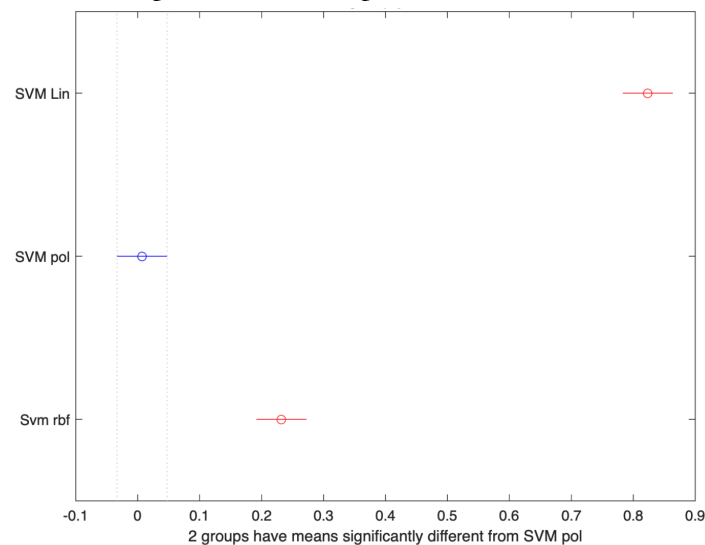


Figura 7. Representación del test de comparación múltiple para el conjunto de datos con  $d = 15$ .

Del mismo modo, se analiza para el conjunto con  $d = 35$ , donde las diferencias estadísticamente significativas tienen lugar al comparar todas las versiones de SVM entre sí, observándose una notable proximidad para que se produzca solapamiento entre las SVMs lineal y no lineal con *kernel* gaussiano:

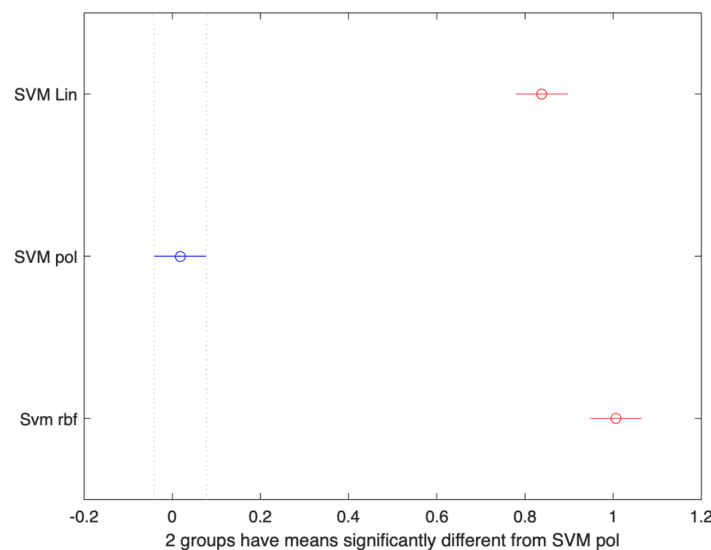


Figura 8. Representación del test de comparación múltiple para el conjunto de datos con  $d = 35$ .

## DISCUSIÓN Y CONCLUSIONES.

En esta parte 2 correspondiente a la práctica 3, se han planteado diferentes versiones de SVMs para 2 conjuntos de datos basados en la serie de Henon. Bajo un parámetro  $d$  con valor de 15, el valor más reducido de error cuadrático medio en test se logra con la versión de SVM no lineal con *kernel* polinómico, mientras que con la SVM lineal se consigue el resultado más alto, lo que pone de manifiesto que esta última conforma una aproximación inadecuada en el tratamiento del problema. Del mismo modo, se produce un patrón similar en el caso del conjunto con un valor de  $d$  igual a 35, aunque hay un incremento en el error cuadrático medio para el caso de la SVM no lineal con *kernel* gaussiano.

Además, en ambos conjuntos de datos existen diferencias estadísticamente significativas entre las versiones de SVMs al compararlas todas entre sí.

De esta forma, dado que las medidas de error cuadrático medio en test se reducen notablemente en la versión de SVM no lineal con *kernel* polinómico, se valoraría el uso de esta aproximación para el tratamiento de los problemas planteados con los diferentes tamaños de ventana por el parámetro  $d$  elegido.

---

# PRÁCTICA 4: APRENDIZAJE BASADO EN REDES NEURONALES Y DISEÑO EXPERIMENTAL

---

## OBJETIVO

El objetivo de esta práctica es llevar a cabo el problema de clasificación mediante Redes Neuronales Artificiales (RNA) del conjunto de datos de Iris, Qsar y Henon normalizados utilizados en las prácticas anteriores para determinar la pertenencia o no de una entrada a una determinada clase.

De esta forma, se entrenan 6 versiones de redes neuronales ajustando varios parámetros asociados para cada conjunto y se alcanza una medida de rendimiento, que, junto con el correspondiente test estadístico, permitirá decidir la versión más adecuada.

## PREPROCESADO DE DATOS.

Para poder realizar el entrenamiento de las RNAs necesitamos que las entradas se traspongan en los tres conjuntos de datos. Sin embargo, dependiendo del conjunto de datos debemos procesar las salidas de diferente forma: para el conjunto de Iris y Qsar necesitamos que las salidas formen una matriz binaria en la cual, se establecen 3 columnas de forma que cada una representa una clase y si corresponde con dicha clase se pone un 1 y sino un 0; mientras que para el problema de regresión de Henon, tanto para el tamaño de ventana igual a 15 como a 35 estudiado en la práctica anterior, necesitamos que las salidas solo se traspongan como las entradas

## APRENDIZAJE.

### CONFIGURACIÓN DEL MÉTODO DE APRENDIZAJE.

Una Red de Neuronas Artificiales (RNA) es un modelo computacional, paralelo y distribuido, basado en la conexión de varios elementos de proceso (neuronas) capaz de aprender a partir de datos. Entre estas RNAs nosotros nos centraremos en las multicapa alimentadas hacia adelante en las cuales existe una capa de neuronas de entrada y otra de salida y múltiples ocultas. En esta práctica, solo se tomará una única capa con diferente número de neuronas de acuerdo con la composición de los datos. En general, tomaremos la mitad, el doble y un valor intermedio de neuronas en comparación con la capa de entrada.

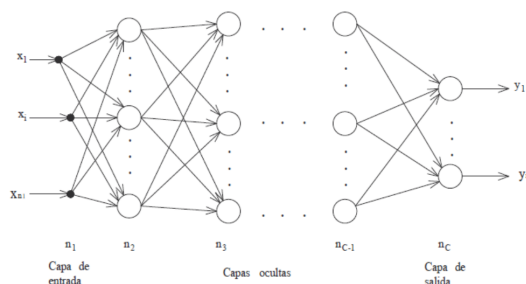


Figura 9. Representación de una RNAs multicapa.

Las RNAs permiten resolver problemas básicos tanto de clasificación (Iris y Qsar) como de regresión (Henon) mediante su aprendizaje con diferentes algoritmos, los cuales permiten modificar los pesos de las conexiones de forma que la red aprende a partir de los datos introducidos. Dicho aprendizaje se consigue mediante la minimización de una función de error, que en el caso

de los problemas de clasificación se utiliza la entropía y en los de regresión el MSE. En cuanto al algoritmo de aprendizaje, nosotros nos centraremos en dos métodos:

El **método de gradiente conjugado escalado** es un algoritmo de aprendizaje de redes neuronales que utiliza la información de la segunda derivada de la función de error para la actualización de los pesos, de forma que se basa en realizar una búsqueda a través de las direcciones conjugadas. Sigue la siguiente ecuación para determinar cuánto hay que moverse en la dirección escogida y para determinar la nueva dirección en la red:

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) + \alpha(n)\mathbf{p}(n) \text{ donde } \alpha(n) \text{ es el paso de aprendizaje} \\ \mathbf{p}(n) &= -\mathbf{g}(n) + \beta(n)\mathbf{p}(n-1) \text{ donde } \beta(n) \text{ es una constante} \end{aligned}$$

El **método Levenberg-Marquardt** es un método alternativo al anterior para una optimización rápida, de forma que, considerando el error cuadrático como función de coste y la matriz hessiana, la regla de optimización es la siguiente:

$$\mathbf{w}(n+1) = \mathbf{w}(n)(\mathbf{J}^T(n)\mathbf{J}(n) + \mu(n)\mathbf{I})^{-1}\mathbf{g}(n)$$

Donde  $\mathbf{J}^T(n)\mathbf{J}(n)$  corresponde con una aproximación de la matriz hessiana, la cual produce cálculos computacionales elevados, por lo que no es recomendable para redes grandes ni para conjuntos de datos grandes.

Por otro lado, para que las RNAs funcionen adecuadamente hay que activar las neuronas y aplicarles la función de activación o transferencia. En nuestro caso, para Iris y Qsar utilizaremos la función de la tangente hiperbólica para la capa oculta y la sigmoidea para la de salida, mientras que para el conjunto Henon, se empleará para la capa de salida la función lineal.

Finalmente, hay que destacar que en la metodología experimental utilizaremos 3 conjuntos de datos (entrenamiento, test y validación), donde el conjunto de test es intocable. Por su parte, el conjunto de validación sirve para apreciar el parámetro que mejor funciona en el contexto. De esta forma, se efectúa el entrenamiento de la RNA mientras el error de validación siga descendiendo.

## MÉTODO DE ESTIMACIÓN DEL ERROR: VALIDACIÓN CRUZADA.

Al igual que en las anteriores prácticas, se opta por la validación cruzada de forma que se calculan los errores sobre el conjunto de test, con un proceso que se itera  $k$  veces y consta de 3 repeticiones. Tal y como se estableció en las prácticas anteriores, para el conjunto Iris y Henon  $k = 10$  y para Qsar  $k = 5$ .

## MEDIDAS DE RENDIMIENTO.

Para las seis versiones de RNAs con diferente tamaño de la capa oculta y diferente algoritmo de aprendizaje (Gradiente conjugado escalado y Levenberg Marquardt) en cada conjunto de datos se calculan las medidas de rendimiento. En el caso de Iris y Qsar, se han calculado las medidas de rendimiento para entrenamiento y test analizadas en las prácticas anteriores: sensibilidad (recall), especificidad, precisión, valor predictivo negativo, exactitud (accuracy) y F1-Score. Sin embargo, para el conjunto de datos Henon se calcula el error cuadrático medio (MSE) para el entrenamiento y test.

# CONJUNTO DE DATOS: IRIS

## RESULTADOS.

### MEDIDAS DE RENDIMIENTO.

Se calculado las siguientes medidas de rendimiento mencionadas anteriormente para las seis RNAs con diferentes tamaños de la capa oculta (2, 5 y 8) y diferente algoritmo de aprendizaje (Gradiente conjugado escalado y Levenberg Marquardt).

Tabla 15. Medidas de rendimiento global para entrenamiento y test de la RNAs con gradiente conjugado escalado del conjunto de datos Iris.

		Gradiente conjugado escalado		
		2	5	8
Sensibilidad (Recall)	Train	9.8025x10 <sup>-1</sup> ±2.4353x10 <sup>-2</sup>	9.8420x10 <sup>-1</sup> ±1.5712x10 <sup>-2</sup>	9.8346x10 <sup>-1</sup> ±1.6842x10 <sup>-2</sup>
	Test	9.7778x10 <sup>-1</sup> ±6.3206x10 <sup>-2</sup>	9.8000x10 <sup>-1</sup> ±6.7375x10 <sup>-2</sup>	9.8222x10 <sup>-1</sup> ±6.4613x10 <sup>-2</sup>
Especificidad	Train	9.9012x10 <sup>-1</sup> ±1.2177x10 <sup>-2</sup>	9.9210x10 <sup>-1</sup> ±7.8558x10 <sup>-3</sup>	9.9173x10 <sup>-1</sup> ±8.4211x10 <sup>-3</sup>
	Test	9.8889x10 <sup>-1</sup> ±3.1603x10 <sup>-2</sup>	9.9000x10 <sup>-1</sup> ±3.3687x10 <sup>-2</sup>	9.9111x10 <sup>-1</sup> ±3.2306x10 <sup>-2</sup>
Precisión	Train	9.8087x10 <sup>-1</sup> ±2.3262x10 <sup>-2</sup>	9.8429x10 <sup>-1</sup> ±1.5494x10 <sup>-2</sup>	9.8356x10 <sup>-1</sup> ±1.6657x10 <sup>-2</sup>
	Test	9.8148x10 <sup>-1</sup> ±5.2672x10 <sup>-2</sup>	9.8386x10 <sup>-1</sup> ±5.3279x10 <sup>-2</sup>	9.8571x10 <sup>-1</sup> ±5.0828x10 <sup>-2</sup>
Valor predictivo negativo	Train	9.9028x10 <sup>-1</sup> ±1.1895x10 <sup>-2</sup>	9.9212x10 <sup>-1</sup> ±7.8000x10 <sup>-3</sup>	9.9175x10 <sup>-1</sup> ±8.3734x10 <sup>-3</sup>
	Test	9.8990x10 <sup>-1</sup> ±2.8730x10 <sup>-2</sup>	9.9108x10 <sup>-1</sup> ±2.9692x10 <sup>-2</sup>	9.9209x10 <sup>-1</sup> ±2.8389x10 <sup>-2</sup>
Exactitud (accuracy)	Train	9.8683x10 <sup>-1</sup> ±1.0753x10 <sup>-2</sup>	9.8947x10 <sup>-1</sup> ±9.3551x10 <sup>-3</sup>	9.8897x10 <sup>-1</sup> ±1.0131x10 <sup>-2</sup>
	Test	9.8519x10 <sup>-1</sup> ±2.7871x10 <sup>-2</sup>	9.8667x10 <sup>-1</sup> ±3.0313x10 <sup>-2</sup>	9.8815x10 <sup>-1</sup> ±2.9270x10 <sup>-2</sup>
F1-Score	Train	9.8024x10 <sup>-1</sup> ±1.6149x10 <sup>-2</sup>	9.8420x10 <sup>-1</sup> ±1.4035x10 <sup>-2</sup>	9.8346x10 <sup>-1</sup> ±1.5199x10 <sup>-2</sup>
	Test	9.7755x10 <sup>-1</sup> ±4.2500x10 <sup>-2</sup>	9.7966x10 <sup>-1</sup> ±4.7084x10 <sup>-2</sup>	9.8190x10 <sup>-1</sup> ±4.5526x10 <sup>-2</sup>



Tabla 16. Medidas de rendimiento global para entrenamiento y test de la RNAs con Levenberg Marquardt del conjunto de datos Iris.

		Levenberg Marquardt		
		2	5	8
Sensibilidad (Recall)	Train	$9.9259 \times 10^{-1} \pm 1.2011 \times 10^{-2}$	$9.9679 \times 10^{-1} \pm 7.8558 \times 10^{-3}$	$9.8914 \times 10^{-1} \pm 1.2123 \times 10^{-2}$
	Test	$9.9333 \times 10^{-1} \pm 3.6102 \times 10^{-2}$	$9.9111 \times 10^{-1} \pm 5.1154 \times 10^{-2}$	$9.8889 \times 10^{-1} \pm 4.6069 \times 10^{-2}$
Especificidad	Train	$9.9630 \times 10^{-1} \pm 6.0055 \times 10^{-3}$	$9.9840 \times 10^{-1} \pm 3.9279 \times 10^{-3}$	$9.9457 \times 10^{-1} \pm 6.0617 \times 10^{-3}$
	Test	$9.9667 \times 10^{-1} \pm 1.8051 \times 10^{-2}$	$9.9556 \times 10^{-1} \pm 2.5577 \times 10^{-2}$	$9.9444 \times 10^{-1} \pm 2.3034 \times 10^{-2}$
Precisión	Train	$9.9275 \times 10^{-1} \pm 1.1696 \times 10^{-2}$	$9.9684 \times 10^{-1} \pm 7.7380 \times 10^{-3}$	$9.8922 \times 10^{-1} \pm 1.2004 \times 10^{-2}$
	Test	$9.9444 \times 10^{-1} \pm 3.0085 \times 10^{-2}$	$9.9167 \times 10^{-1} \pm 4.6825 \times 10^{-2}$	$9.9074 \times 10^{-1} \pm 3.8391 \times 10^{-2}$
Valor predictivo negativo	Train	$9.9634 \times 10^{-1} \pm 5.9251 \times 10^{-3}$	$9.9841 \times 10^{-1} \pm 3.8981 \times 10^{-3}$	$9.9459 \times 10^{-1} \pm 6.0314 \times 10^{-3}$
	Test	$9.9697 \times 10^{-1} \pm 1.6410 \times 10^{-2}$	$9.9574 \times 10^{-1} \pm 2.4179 \times 10^{-2}$	$9.9495 \times 10^{-1} \pm 2.0940 \times 10^{-2}$
Exactitud (accuracy)	Train	$9.9506 \times 10^{-1} \pm 5.2083 \times 10^{-3}$	$9.9786 \times 10^{-1} \pm 4.0411 \times 10^{-3}$	$9.9276 \times 10^{-1} \pm 6.7524 \times 10^{-3}$
	Test	$9.9556 \times 10^{-1} \pm 1.6723 \times 10^{-2}$	$9.9407 \times 10^{-1} \pm 3.1036 \times 10^{-2}$	$9.9259 \times 10^{-1} \pm 2.1069 \times 10^{-2}$
F1-Score	Train	$9.9259 \times 10^{-1} \pm 7.8172 \times 10^{-3}$	$9.9679 \times 10^{-1} \pm 6.0623 \times 10^{-3}$	$9.8914 \times 10^{-1} \pm 1.0130 \times 10^{-2}$
	Test	$9.9327 \times 10^{-1} \pm 2.5473 \times 10^{-2}$	$9.9102 \times 10^{-1} \pm 4.7268 \times 10^{-2}$	$9.8878 \times 10^{-1} \pm 3.2101 \times 10^{-2}$

Como se observa tanto en la Tabla 15 como en la Tabla 16, se obtienen unos resultados similares en las medidas de rendimiento para entrenamiento y test de forma global. Además, las medidas para entrenamiento de las redes, en general, son muy próximas a 1, por lo que tanto para el algoritmo de aprendizaje de gradiente conjugado escalado como para el algoritmo de Levenberg Marquardt las redes funcionan adecuadamente para realizar la clasificación del conjunto Iris.

## COMPARACIÓN DE MODELOS.

### RESULTADOS PARA LOS MODELOS IMPLICADOS DE RNAs.

Al igual que en el resto de las prácticas realizadas para el conjunto de Iris, utilizaremos F1-Score para realizar la comparación de las RNAs. Se muestran a continuación las medias y correspondientes desviaciones típicas en cada caso para test:

Tabla 17. Media y desviación típica de la medida de rendimiento F1-Score para las RNAs planteadas del conjunto de datos Iris.

Modelo		F1-Score (media $\pm$ desviación típica)
Gradiente conjugado escalado	2	$9.7755 \times 10^{-1} \pm 4.2500 \times 10^{-2}$
	5	$9.7966 \times 10^{-1} \pm 4.7084 \times 10^{-2}$
	8	$9.8190 \times 10^{-1} \pm 4.5526 \times 10^{-2}$
Levenberg Marquardt	2	$9.9327 \times 10^{-1} \pm 2.5473 \times 10^{-2}$
	5	$9.9102 \times 10^{-1} \pm 4.7268 \times 10^{-2}$
	8	$9.8878 \times 10^{-1} \pm 3.2101 \times 10^{-2}$

A continuación, se muestra la tabla ANOVA de 1 vía del test estadístico realizado con un valor crítico de 0.10 para las 6 RNAs:

Tabla 18. Tabla ANOVA de F1-Score en test para comparar las RNAs del conjunto de datos Iris.

Source	SS	df	MS	F	Prob>F
Columns	315.7	5	63.14	2.15	0.8284
Error	8357.3	54	154.765		
Total	8673	59			

El p-valor resultante es mayor que el valor crítico de 0.10 establecido, de tal manera que se mantiene la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las RNAs utilizadas con diferente algoritmo de aprendizaje y tamaño de la capa oculta.

## DIAGRAMA DE CAJAS Y BIGOTES.

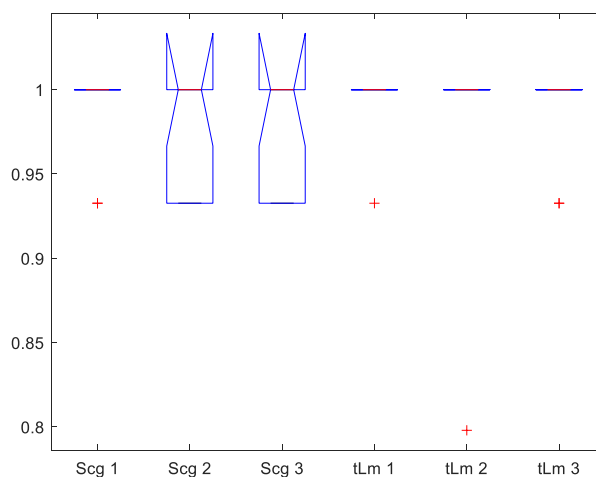


Figura 10. Diagrama de cajas y bigotes para F1-Score en comparación de RNAs del conjunto de datos Iris.

Para el conjunto de Iris, se observa en el diagrama de cajas y bigotes que las RNAs entrenadas con ambos algoritmos de aprendizaje dan lugar a medianas similares. Sin embargo, el empleo del algoritmo de gradiente conjugado escalado con tamaño de capa oculta de 5 y 8 da lugar a un rango intercuartílico mayor al obtenido con el resto de RNAs. Además, hay que destacar que, en el resto

de las redes, la mayoría de los datos se concentran en torno a un valor de 1, a excepción de algún dato atípico que se encuentra más alejado.

## DISCUSIÓN Y CONCLUSIONES.

El planteamiento de las RNAs con diferente tamaño de capa oculta (2, 5 y 8) y algoritmo de aprendizaje (gradiente conjugado escalado y Levenberg Marquardt) para el conjunto de Iris y la posterior estimación del error real de estas mediante validación cruzada conducen a unas medidas de rendimiento que no varían de forma estadísticamente significativa entre sí, tal y como se verifica con el test estadístico a través del contraste de Kruskal-Wallis y el diagrama de cajas y bigotes correspondiente. De esta forma, teniendo en cuenta las medidas del F1-Score analizadas, se elige la versión de RNA con Levenberg y tamaño de capa oculta de 2 como la más adecuada para el tratamiento de este problema de clasificación de Iris.

# CONJUNTO DE DATOS: QSAR BIODEGRADATION

## RESULTADOS.

### MEDIDAS DE RENDIMIENTO.

Al igual que en el conjunto de Iris, se han calculado las siguientes medidas de rendimiento para las seis RNAs con diferentes tamaños de la capa oculta (21, 56 y 84) y diferentes algoritmos de aprendizaje mencionados.

Tabla 19. Medidas de rendimiento global para entrenamiento y test de la RNA con gradiente conjugado escalado del conjunto de datos Qsar.

		Gradiente conjugado escalado		
		21	56	84
Sensibilidad (Recall)	Train	$8.5466 \times 10^{-1} \pm 5.4532 \times 10^{-2}$	$8.6763 \times 10^{-1} \pm 5.7876 \times 10^{-2}$	$8.5292 \times 10^{-1} \pm 6.5670 \times 10^{-2}$
	Test	$8.4863 \times 10^{-1} \pm 6.8922 \times 10^{-2}$	$8.6677 \times 10^{-1} \pm 6.4667 \times 10^{-2}$	$8.5080 \times 10^{-1} \pm 6.8969 \times 10^{-2}$
Especificidad	Train	$8.5466 \times 10^{-1} \pm 5.4532 \times 10^{-2}$	$8.6763 \times 10^{-1} \pm 5.7876 \times 10^{-2}$	$8.5292 \times 10^{-1} \pm 6.5670 \times 10^{-2}$
	Test	$8.4863 \times 10^{-1} \pm 6.8922 \times 10^{-2}$	$8.6677 \times 10^{-1} \pm 6.4667 \times 10^{-2}$	$8.5080 \times 10^{-1} \pm 6.8969 \times 10^{-2}$
Precisión	Train	$8.5734 \times 10^{-1} \pm 4.4588 \times 10^{-2}$	$8.7497 \times 10^{-1} \pm 3.3271 \times 10^{-2}$	$8.6053 \times 10^{-1} \pm 3.9822 \times 10^{-2}$
	Test	$8.5342 \times 10^{-1} \pm 5.5406 \times 10^{-2}$	$8.7264 \times 10^{-1} \pm 4.6637 \times 10^{-2}$	$8.5765 \times 10^{-1} \pm 4.9511 \times 10^{-2}$
Valor predictivo negativo	Train	$8.5734 \times 10^{-1} \pm 4.4588 \times 10^{-2}$	$8.7497 \times 10^{-1} \pm 3.3271 \times 10^{-2}$	$8.6053 \times 10^{-1} \pm 3.9822 \times 10^{-2}$
	Test	$8.5342 \times 10^{-1} \pm 5.5406 \times 10^{-2}$	$8.7264 \times 10^{-1} \pm 4.6637 \times 10^{-2}$	$8.5765 \times 10^{-1} \pm 4.9511 \times 10^{-2}$
Exactitud (accuracy)	Train	$8.7164 \times 10^{-1} \pm 7.2127 \times 10^{-3}$	$8.8578 \times 10^{-1} \pm 7.1741 \times 10^{-3}$	$8.7299 \times 10^{-1} \pm 1.4185 \times 10^{-2}$
	Test	$8.6698 \times 10^{-1} \pm 2.4608 \times 10^{-2}$	$8.8404 \times 10^{-1} \pm 2.4253 \times 10^{-2}$	$8.7046 \times 10^{-1} \pm 2.4048 \times 10^{-2}$
F1-Score	Train	$8.5592 \times 10^{-1} \pm 4.9217 \times 10^{-2}$	$8.7105 \times 10^{-1} \pm 4.5144 \times 10^{-2}$	$8.5646 \times 10^{-1} \pm 5.2499 \times 10^{-2}$
	Test	$8.5038 \times 10^{-1} \pm 5.8376 \times 10^{-2}$	$8.6929 \times 10^{-1} \pm 5.3426 \times 10^{-2}$	$8.5385 \times 10^{-1} \pm 5.7584 \times 10^{-2}$

Tabla 20. Medidas de rendimiento global para entrenamiento y test de la RNA con Levenberg Marquardt del conjunto de datos Qsar.

		Levenberg Marquardt		
		21	56	84
Sensibilidad (Recall)	Train	$8.8908 \times 10^{-1} \pm 3.9530 \times 10^{-2}$	$8.8483 \times 10^{-1} \pm 5.4777 \times 10^{-2}$	$9.2431 \times 10^{-1} \pm 1.9345 \times 10^{-2}$
	Test	$8.8906 \times 10^{-1} \pm 5.1494 \times 10^{-2}$	$8.8029 \times 10^{-1} \pm 6.4020 \times 10^{-2}$	$9.2102 \times 10^{-1} \pm 3.5495 \times 10^{-2}$
Especificidad	Train	$8.8908 \times 10^{-1} \pm 3.9530 \times 10^{-2}$	$8.8483 \times 10^{-1} \pm 5.4777 \times 10^{-2}$	$9.2431 \times 10^{-1} \pm 1.9345 \times 10^{-2}$
	Test	$8.8906 \times 10^{-1} \pm 5.1494 \times 10^{-2}$	$8.8029 \times 10^{-1} \pm 6.4020 \times 10^{-2}$	$9.2102 \times 10^{-1} \pm 3.5495 \times 10^{-2}$
Precisión	Train	$8.9023 \times 10^{-1} \pm 3.5816 \times 10^{-2}$	$8.9328 \times 10^{-1} \pm 2.5586 \times 10^{-2}$	$9.1940 \times 10^{-1} \pm 3.5095 \times 10^{-2}$
	Test	$8.9129 \times 10^{-1} \pm 4.7995 \times 10^{-2}$	$8.8886 \times 10^{-1} \pm 3.9278 \times 10^{-2}$	$9.1707 \times 10^{-1} \pm 4.3921 \times 10^{-2}$
Valor predictivo negativo	Train	$8.9023 \times 10^{-1} \pm 3.5816 \times 10^{-2}$	$8.9328 \times 10^{-1} \pm 2.5586 \times 10^{-2}$	$9.1940 \times 10^{-1} \pm 3.5095 \times 10^{-2}$
	Test	$8.9129 \times 10^{-1} \pm 4.7995 \times 10^{-2}$	$8.8886 \times 10^{-1} \pm 3.9278 \times 10^{-2}$	$9.1707 \times 10^{-1} \pm 4.3921 \times 10^{-2}$
Exactitud (accuracy)	Train	$9.0142 \times 10^{-1} \pm 5.7576 \times 10^{-3}$	$9.0158 \times 10^{-1} \pm 7.5254 \times 10^{-3}$	$9.2954 \times 10^{-1} \pm 3.6517 \times 10^{-3}$
	Test	$9.0142 \times 10^{-1} \pm 2.3030 \times 10^{-2}$	$8.9763 \times 10^{-1} \pm 2.5373 \times 10^{-2}$	$9.2670 \times 10^{-1} \pm 1.7324 \times 10^{-2}$
F1-Score	Train	$8.8962 \times 10^{-1} \pm 3.7298 \times 10^{-2}$	$8.8873 \times 10^{-1} \pm 3.9726 \times 10^{-2}$	$9.2169 \times 10^{-1} \pm 2.5578 \times 10^{-2}$
	Test	$8.8964 \times 10^{-1} \pm 4.5169 \times 10^{-2}$	$8.8422 \times 10^{-1} \pm 5.0545 \times 10^{-2}$	$9.1849 \times 10^{-1} \pm 3.3048 \times 10^{-2}$

Como se observa en las tablas anteriores, se obtienen unos resultados similares en las medidas de rendimiento para entrenamiento y test para todas las redes configuradas. Sin embargo, se aprecia que hay una mejora de dichas medidas, en general, cuando se utiliza el algoritmo de aprendizaje Levenberg Marquardt con un tamaño de capa oculta de 84, ya que en todas las medidas de rendimiento se alcanza casi un valor próximo a 1.

## COMPARACIÓN DE MODELOS.

### RESULTADOS PARA LOS MODELOS IMPLICADOS DE RNAs.

Al igual que en el resto de las prácticas realizadas para el conjunto de Qsar, utilizaremos F1-Score para realizar la comparación de las RNAs. Se muestran a continuación las medias y correspondientes desviaciones típicas en cada caso para test:

Tabla 21. Media y desviación típica de la medida de rendimiento F1-Score para las RNAs planteadas del conjunto de datos Qsar.

Modelo		F1-Score (media $\pm$ desviación típica)
Gradiente conjugado escalado	21	$8.5038 \times 10^{-1} \pm 5.8376 \times 10^{-2}$
	56	$8.6929 \times 10^{-1} \pm 5.3426 \times 10^{-2}$
	84	$8.5385 \times 10^{-1} \pm 5.7584 \times 10^{-2}$
Levenberg Marquardt	21	$8.8964 \times 10^{-1} \pm 4.5169 \times 10^{-2}$
	56	$8.8422 \times 10^{-1} \pm 5.0545 \times 10^{-2}$
	84	$9.1849 \times 10^{-1} \pm 3.3048 \times 10^{-2}$

A continuación, se muestra la tabla ANOVA de 1 vía del test estadístico realizado con un valor crítico de 0.10 para las 6 RNAs:

Tabla 22. Tabla ANOVA de F1-Score en test para comparar las RNAs del conjunto de datos Qsar.

Source	SS	df	MS	F	Prob>F
Columns	0.02119	5	0.00424	4.14	0.0075
Error	0.0246	24	0.00102		
Total	0.04579	29			

El p-valor resultante es inferior al valor crítico de 0.10 establecido, de manera que se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las versiones de RNAs analizadas.

### DIAGRAMA DE CAJAS Y BIGOTES.

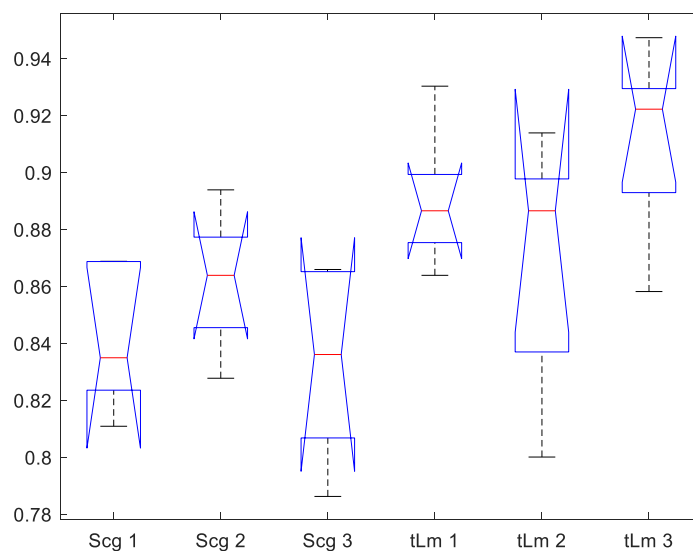


Figura 11. Diagrama de cajas y bigotes para F1-Score en test de las RNAs del conjunto de datos Qsar.

En el diagrama de cajas y bigotes se puede observar que las RNAs con algoritmo de Levenberg Marquardt presentan una mayor mediana para F1-Score que las que utilizan el algoritmo de

gradiente conjugado escalado. Además, la red con el tamaño de capa oculta de 21 y algoritmo de Levenberg Marquardt presenta el menor rango intercuartílico de entre todas las RNAs analizadas.

### TEST DE COMPARACIÓN MÚLTIPLE.

La existencia de diferencias estadísticamente significativas entre las RNAs da lugar a una representación de un test de comparación múltiple, como un complemento al diagrama de cajas y bigotes:

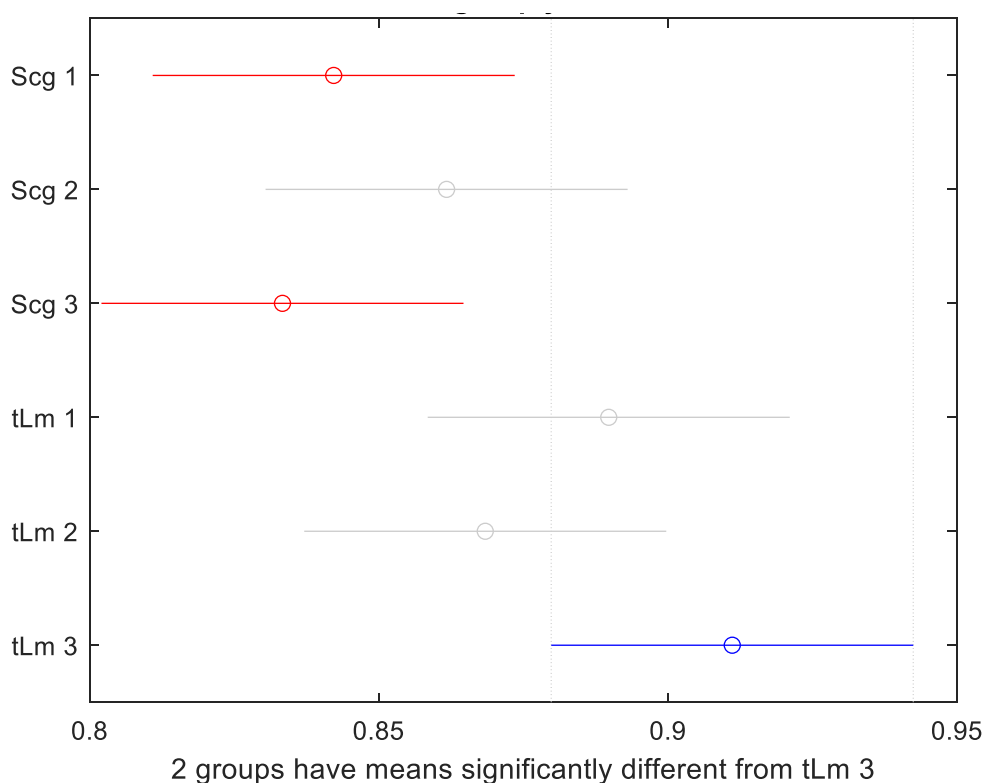


Figura 12. Representación del test de comparación múltiple para las RNAs en el conjunto de datos Qsar.

Se observa en la Figura 12 que existen diferencias significativas entre el uso del algoritmo de aprendizaje de gradiente conjugado escalado con tamaño de la capa oculta de 21 u 84 y la RNA con el algoritmo de aprendizaje Levenberg Marquardt con tamaño de la capa oculta de 84.

### DISCUSIÓN Y CONCLUSIONES.

De esta forma, teniendo en cuenta las medidas de rendimiento de entrenamiento y test obtenidas y el test estadístico realizado para F1-Score, la RNA con tamaño de capa oculta de 84 y algoritmo de aprendizaje Levenberg Marquardt es la mejor aproximación para resolver el problema de clasificación para el conjunto de datos Qsar.

# SERIE DE HENON

## RESULTADOS.

### MEDIDAS DE RENDIMIENTO.

Al igual que en el modelo con SVMs para la resolución de problemas de regresión, la medida de rendimiento analizada es el error cuadrático medio. En este caso, se calculará para 6 RNAs con diferentes tamaños de la capa oculta (15, 45, 70) y diferentes algoritmos de aprendizaje mencionados para cada conjunto de Henon, uno con tamaño de ventana igual a 15 y otro igual a 35.

Tabla 23. Media y desviación típica del error cuadrático medio para entrenamiento y test bajo los parámetros d establecidos.

Modelo			Error cuadrático medio (media $\pm$ desviación típica)	
			Conjunto: d = 15	Conjunto: d = 35
Gradiente conjugado escalado	15	Entrenamiento	$4.4236 \times 10^{-4} \pm 5.7155 \times 10^{-6}$	$3.1659 \times 10^{-3} \pm 4.7359 \times 10^{-3}$
		Test	$4.4236 \times 10^{-4} \pm 5.0759 \times 10^{-5}$	$3.1900 \times 10^{-3} \pm 4.8472 \times 10^{-3}$
	45	Entrenamiento	$2.5891 \times 10^{-4} \pm 4.0282 \times 10^{-6}$	$1.0185 \times 10^{-3} \pm 2.6002 \times 10^{-5}$
		Test	$2.5922 \times 10^{-4} \pm 3.6378 \times 10^{-5}$	$1.0196 \times 10^{-3} \pm 2.3906 \times 10^{-4}$
	70	Entrenamiento	$1.8336 \times 10^{-4} \pm 3.8845 \times 10^{-6}$	$6.8566 \times 10^{-3} \pm 2.1559 \times 10^{-4}$
		Test	$1.8347 \times 10^{-4} \pm 3.5303 \times 10^{-5}$	$6.8661 \times 10^{-3} \pm 1.9624 \times 10^{-3}$
Levenberg Marquardt	15	Entrenamiento	$1.4842 \times 10^{-8} \pm 2.2619 \times 10^{-10}$	$8.7122 \times 10^{-10} \pm 5.3660 \times 10^{-11}$
		Test	$1.4845 \times 10^{-8} \pm 2.0373 \times 10^{-9}$	$8.8328 \times 10^{-10} \pm 2.0299 \times 10^{-10}$
	45	Entrenamiento	$1.2809 \times 10^{-10} \pm 2.1517 \times 10^{-12}$	$1.1315 \times 10^{-7} \pm 1.3205 \times 10^{-7}$
		Test	$1.2810 \times 10^{-10} \pm 1.9389 \times 10^{-11}$	$1.8214 \times 10^{-7} \pm 4.8878 \times 10^{-7}$
	70	Entrenamiento	$3.8945 \times 10^{-10} \pm 4.5179 \times 10^{-10}$	$1.4323 \times 10^{-3} \pm 7.0465 \times 10^{-3}$
		Test	$7.1225 \times 10^{-10} \pm 2.2247 \times 10^{-9}$	$4.9985 \times 10^{-3} \pm 2.4901 \times 10^{-2}$



En la Tabla 23, se refleja que se han obtenido unos buenos resultados para todas las redes entrenadas en los conjuntos de datos, ya que presentan valores cercanos a 0. Además, el MSE disminuye todavía más cuando se utiliza el algoritmo de aprendizaje Levenberg Marquardt tanto para el conjunto de datos Henon  $d = 15$  como para  $d = 35$ , llegando a magnitudes de  $\times 10^{-10}$ .

## COMPARACIÓN DE MODELOS.

### RESULTADOS PARA LOS MODELOS IMPLICADOS DE RNAs.

Al igual que en las prácticas anteriores, se realiza un test estadístico en el que se establece un valor crítico de 0.10, para determinar si las medidas de MSE en test resultan o no similares estadísticamente para las versiones de RNAs generadas en los conjuntos.

A continuación, se muestra la correspondiente tabla resultante del ANOVA de 1 vía para el conjunto de datos bajo un parámetro  $d = 15$  y bajo  $d = 35$ :

**Tabla 24. Tabla ANOVA de MSE en test para comparar las RNAs del conjunto de datos Henon  $d = 15$ .**

Source	SS	df	MS	F	Prob>F
Columns	17269	5	3453.8	56.62	6.05711e-11
Error	726	54	13.44		
Total	17995	59			

**Tabla 25. Tabla ANOVA de MSE en test para comparar las RNAs del conjunto de datos Henon  $d = 35$ .**

Source	SS	df	MS	F	Prob>F
Columns	13406.4	5	2681.28	43.96	2.365e-08
Error	4588.6	54	84.97		
Total	17995	59			

Tal y como se puede observar en la Tabla 24 y Tabla 25, el p-valor resultante es inferior al valor crítico de 0.10 establecido, de tal forma que se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las versiones de RNAs analizadas para el conjunto Henon  $d = 15$  y Henon  $d = 35$ .

## DIAGRAMA DE CAJAS Y BIGOTES.

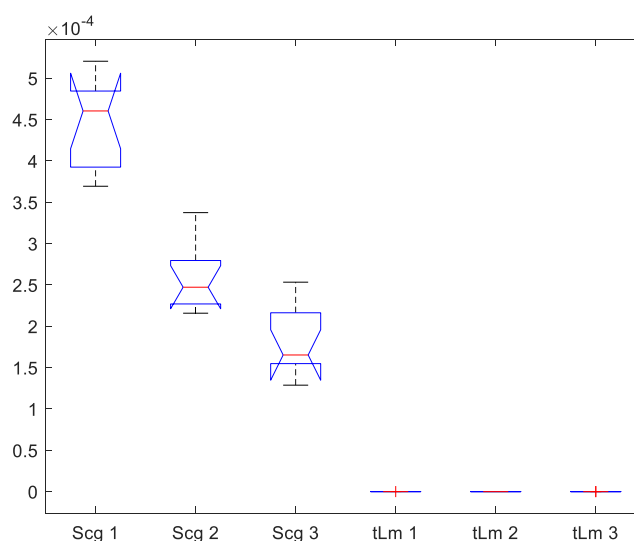


Figura 13. Diagrama de cajas y bigotes para error cuadrático medio del conjunto de datos con  $d = 15$ .

Como se puede apreciar en el caso del conjunto de datos con  $d = 15$  (Figura 13), existe gran diferencia entre las redes que utilizan el algoritmo de aprendizaje de gradiente conjugado escalado y las que usan Levenberg Marquardt. En concreto, estas últimas presentan una mediana más cercana a 0 y, además, el rango intercuartílico que presentan es mínimo en comparación con las RNAs con gradiente conjugado escalado. También, se puede observar que el uso de una capa oculta de menor tamaño para este tipo de algoritmo empeora el entrenamiento de la red.

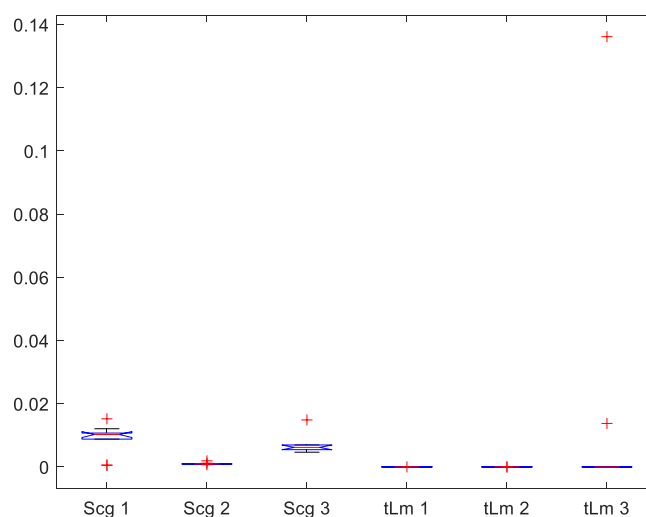


Figura 14. Diagrama de cajas y bigotes para error cuadrático medio del conjunto de datos con  $d = 35$ .

Por otro lado, para el conjunto de Henon  $d = 35$ , se observa que los dos algoritmos de aprendizaje siguen una tendencia similar para la mediana, que es cercana a cero. Sin embargo, para el algoritmo de gradiente conjugado escalado se observa un mayor rango intercuartílico, mientras que para Levenberg Marquardt y capa oculta de tamaño 70, aparecen más datos atípicos.

## TEST DE COMPARACIÓN MÚLTIPLE.

Para los conjuntos planteados, se visualiza el test de comparación múltiple, para distinguir las diferencias estadísticamente significativas entre RNAs.

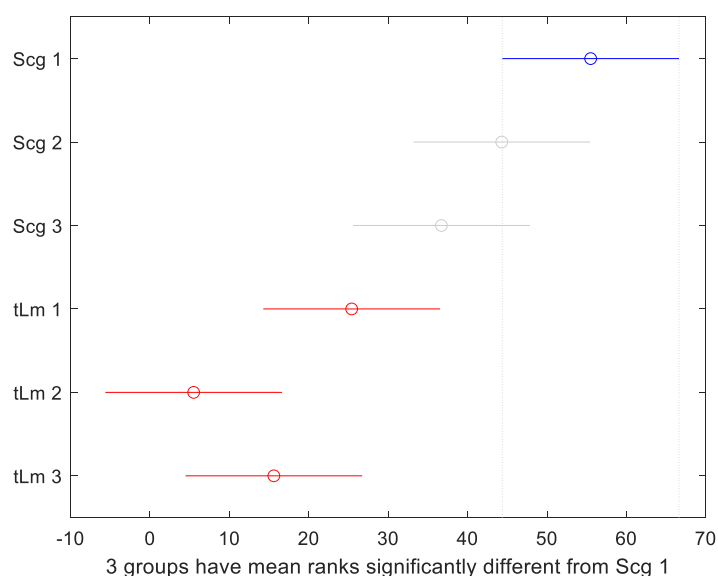


Figura 15. Representación del test de comparación múltiple para el conjunto de datos con  $d = 15$ .

Para el conjunto Henon  $d = 15$ , en general, existen diferencias significativas entre las RNAs que usan el algoritmo de aprendizaje de gradiente conjugado escalado y las que usan el algoritmo de Levenberg Marquardt. En concreto, el resultado más destacable es la existencia de diferencias entre todas las redes con algoritmo de Levenberg Marquardt y el de gradiente conjugado escalado con el menor tamaño de capa oculta (scg 1).

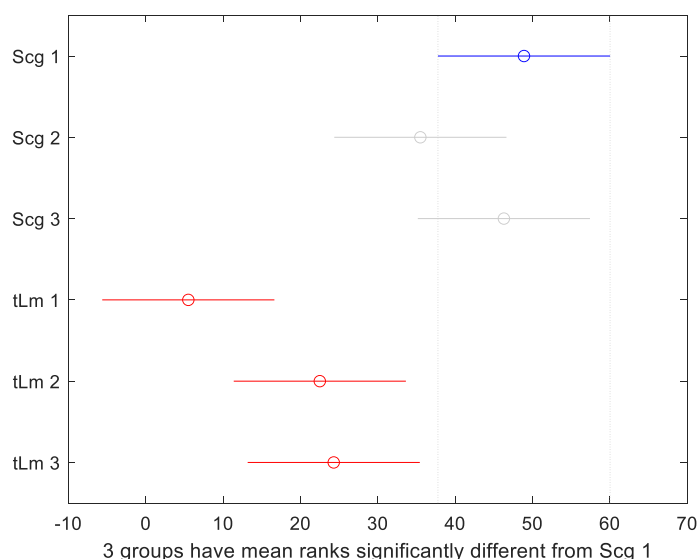


Figura 16. Representación del test de comparación múltiple para el conjunto de datos con  $d = 35$ .

Por otro lado, y al igual que para Henon  $d = 15$ , se producen diferencias significativas entre las RNAs que usan el algoritmo de aprendizaje de gradiente conjugado escalado y las que usan el

algoritmo de Levenberg Marquardt. Además, también hay que destacar que existen dichas diferencias entre todas las RNAs con algoritmo de Levenberg Marquardt y el de gradiente conjugado escalado con menor tamaño de capa oculta (scg 1).

## DISCUSIÓN Y CONCLUSIONES.

De acuerdo con las diferentes versiones de RNAs planteadas y debido a los resultados obtenidos en los test estadísticos realizados para MSE en el conjunto de datos de Henon, optaremos como mejor RNA para el problema de regresión del conjunto Henon con  $d = 15$ , aquella que utiliza el algoritmo de aprendizaje de Levenberg Marquardt y un tamaño de capa oculta de 45.

Por otro lado, para el conjunto de datos Henon  $d = 35$  elegiremos como mejor RNA aquella que utiliza el algoritmo de aprendizaje de Levenberg Marquardt y un tamaño de capa oculta de 15.

# COMPARACIÓN DE TODOS LOS MODELOS ESTUDIADOS

# CONJUNTO DE DATOS: IRIS

## RESULTADOS PARA TODOS LOS MODELOS ESTUDIADOS

Se lleva a cabo la comparación del modelo de discriminante lineal, discriminante cuadrático, árbol de decisión, SVM lineal y RNA de las anteriores prácticas para el conjunto de Iris. A continuación, se muestran las medias y desviaciones típicas para las medidas del F1-Score empleadas en el test estadístico:

Tabla 26. Media y desviación típica de la medida de rendimiento F1-Score para todas las técnicas vistas en el conjunto de datos Iris.

Modelo	F1-Score (media ± desviación típica)
Discriminante lineal	0,9798 ± 0,0592
Discriminante cuadrático	0,9728 ± 0,0592
Árbol de decisión	0,9464 ± 0,0527
SVM lineal	0,9595 ± 0,0729
RNA Levenberg Marquardt 2	0,9933 ± 0,0255

De nuevo, para el test estadístico se estableció un valor crítico de 0.10. A continuación, se muestra la correspondiente tabla resultante del ANOVA de 1 vía:

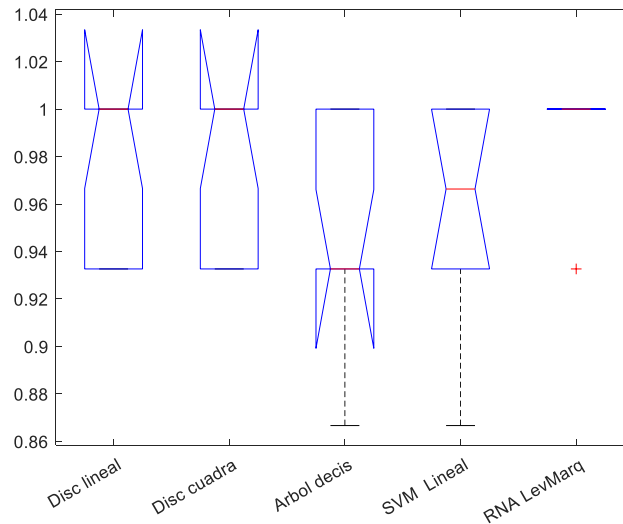
Tabla 27. Tabla ANOVA de F1-Score en test para comparar las SVMs y técnicas vistas del conjunto de datos Iris.

Source	SS	df	MS	F	Prob>F
Columns	1096.25	4	274.062	7.08	0.1319
Error	6494.25	45	144.317		
Total	7590.5	49			

El p-valor resultante es superior al valor crítico de 0.10 establecido, de tal manera que se mantiene la hipótesis nula de ausencia de diferencias estadísticamente significativas entre todas las técnicas evaluadas.

## DIAGRAMA DE CAJAS Y BIGOTES.

Gráficamente, con la representación del diagrama de cajas y bigotes se verifica la ausencia de diferencias estadísticamente significativas para el F1-Score sobre el conjunto de test como medida para la comparación de las técnicas:



**Figura 17. Diagrama de cajas y bigotes para F1-Score para comparar todas las técnicas vistas del conjunto de datos Iris.**

Tal y como se puede apreciar en el diagrama obtenido, el modelo de discriminante lineal y cuadrático, junto con la RNA presentan la mediana más elevada para el F1-Score empleado en la comparación. Sin embargo, los dos casos de discriminante presentan un mayor rango intercuartílico en comparación con la red. Por otro lado, el árbol de decisión presenta la menor mediana junto con la SVM.

## DISCUSIÓN Y CONCLUSIONES.

En base a los resultados obtenidos al comparar todos los modelos estudiados en prácticas, es preciso destacar que las aproximaciones idóneas para el problema son los modelos de discriminante lineal, discriminante cuadrático y la RNA ya que presentan los valores más altos para F1-Score. Sin embargo, hay que destacar que sería una buena elección seleccionar el discriminante lineal como modelo para solucionar el problema de clasificación debido a que presenta unos valores similares a otros modelos más complejos, como es la RNA.

# CONJUNTO DE DATOS: QSAR BIODEGRADATION

## RESULTADOS PARA LOS MODELOS ESTUDIADOS

A continuación, se muestran las medias y correspondientes desviaciones típicas para las medidas del F1-Score empleadas en el test estadístico para la correspondiente comparación:

Tabla 28. Media y desviación típica de la medida de rendimiento F1-Score para comparación de todas las técnicas anteriores en el conjunto de datos Qsar.

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
RNA Levenberg Marquardt 84	0,9185 $\pm$ 0,0331

De nuevo, para el test estadístico se estableció un valor crítico de 0.10. A continuación, se muestra la correspondiente tabla resultante del ANOVA de 1 vía:

Tabla 29. Tabla ANOVA de F1-Score en test para la comparación de todas las técnicas vistas en el conjunto de datos Qsar.

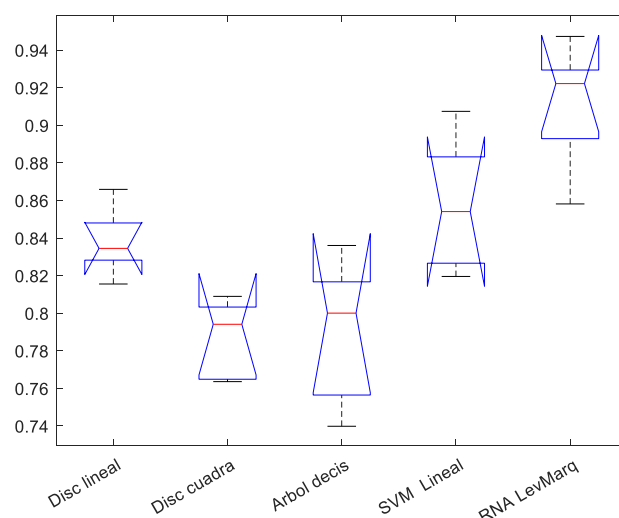
Source	SS	df	MS	F	Prob>F
Columns	0.05335	4	0.01334	14.4	1.09075e-05
Error	0.01853	20	0.00093		
Total	0.07187	24			

El p-valor resultante es inferior al valor crítico de 0.10 establecido, de tal manera que se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las técnicas evaluadas.

## DIAGRAMA DE CAJAS Y BIGOTES.

Se verifica posteriormente la existencia de diferencias estadísticamente significativas con la representación del diagrama de cajas y bigotes para el F1-Score sobre el conjunto de test como medida para la comparación de modelos:



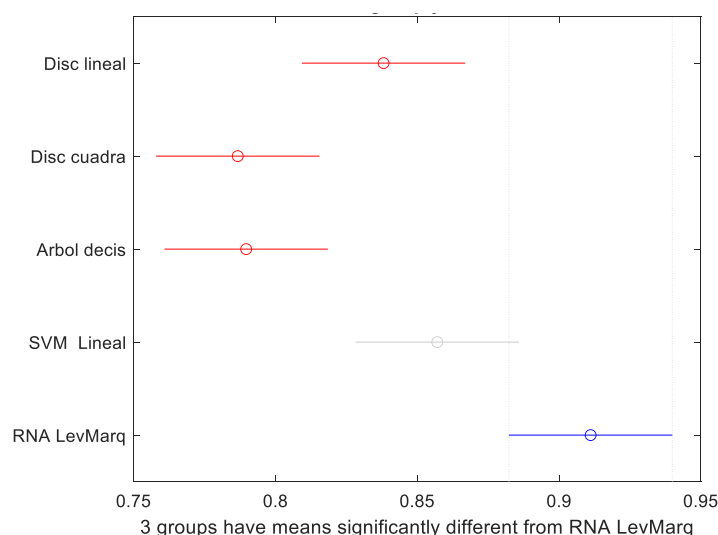


**Figura 18.** Diagrama de cajas y bigotes para F1-Score en la comparación de todas las técnicas vistas en el conjunto de datos Qsar.

Como se puede observar en la Figura 18, la RNA presenta la mediana más elevada para el F1-Score empleado, siendo este casi similar a 1. Por el contrario, el discriminante cuadrático y el árbol de decisión presentan los valores más bajos. En lo que se refiere al rango intercuartílico, cabe destacar que este es mayor en el árbol de decisión, mientras que el discriminante lineal presenta el menor rango en comparación con las otras aproximaciones.

### TEST DE COMPARACIÓN MÚLTIPLE.

Mediante la representación complementaria del test de comparación múltiple, se distinguen aquellas aproximaciones entre las que se producen diferencias estadísticamente significativas.



**Figura 19.** Representación del test de comparación múltiple para todas las técnicas vistas en el conjunto de datos Qsar.

Tal y como se refleja en la Figura 19, existen diferencias significativas entre la RNA y el discriminante lineal, cuadrático y el árbol de decisión. Además, estos últimos modelos también son diferentes estadísticamente en comparación con la SVM.

## DISCUSIÓN Y CONCLUSIONES.

En base a los resultados alcanzados en el test de comparación y en las medidas de F1-Score para cada modelo analizado, podemos concluir que el mejor modelo para realizar la clasificación del conjunto de datos Qsar es la RNA con algoritmo de aprendizaje Levenberg Marquardt y tamaño de capa oculta igual a 84.

Además, es preciso destacar que, para este conjunto de datos, la reducción de la dimensionalidad se debería incluir como una fase en el preprocesado de los datos. El motivo es que, entre las 41 variables, es posible la existencia de redundancia entre los datos, lo que determinaría la extracción de variables para eliminar el posible ruido en el sistema de aprendizaje del modelo por medio de las aproximaciones aplicadas.

# SERIE DE HENON

## RESULTADOS PARA LOS MODELOS ESTUDIADOS

A continuación, se muestran las medias y correspondientes desviaciones típicas para el MSE empleado en el test estadístico para la correspondiente comparación:

Tabla 30. Media y desviación típica de MSE para la comparación de SVM polinomial y RNA Levenberg Marquardt para el conjunto de datos Henon  $d = 15$  y  $d = 35$ .

Modelo	Error cuadrático medio (media $\pm$ desviación típica)	
	Conjunto: $d = 15$	Conjunto: $d = 35$
SVM polinomial	0,0070 $\pm$ 0,0004	0,0188 $\pm$ 0,0025
RNA Levenberg Marquardt (45 y 15)	1.2810 $\times 10^{-10}$ $\pm$ 1.9389 $\times 10^{-11}$	8.8328 $\times 10^{-10}$ $\pm$ 2.0299 $\times 10^{-10}$

De nuevo, para el test estadístico se estableció un valor crítico de 0.10. A continuación, se muestra la correspondiente tabla resultante del ANOVA de 1 vía para el conjunto Henon con  $d = 15$ :

Tabla 31. Tabla ANOVA de MSE en test para comparar la SVM y RNA del conjunto de datos Henon  $d = 15$ .

Source	SS	df	MS	F	Prob>F
Columns	0.00025	1	0.00025	3228.94	9.19483e-22
Error	0	18	0		
Total	0.00025	19			

Tabla 32. Tabla ANOVA de MSE en test para comparar la SVM y RNA del conjunto de datos Henon  $d = 35$ .

Source	SS	df	MS	F	Prob>F
Columns	500	1	500	14.29	0.0002
Error	165	18	9.1667		
Total	665	19			

Para ambos conjuntos de datos, el p-valor resultante es inferior al valor crítico de 0.10 establecido, de tal manera que se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las técnicas evaluadas.

## DIAGRAMA DE CAJAS Y BIGOTES.

Se verifica posteriormente la existencia de diferencias estadísticamente significativas con la representación del diagrama de cajas y bigotes para el MSE sobre el conjunto de test como medida para la comparación de modelos:

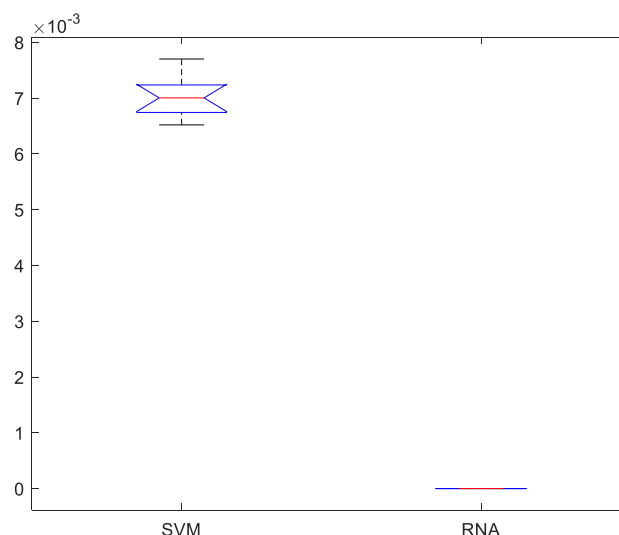


Figura 20. Diagrama de cajas y bigotes para error cuadrático medio del conjunto de datos con  $d = 15$ .

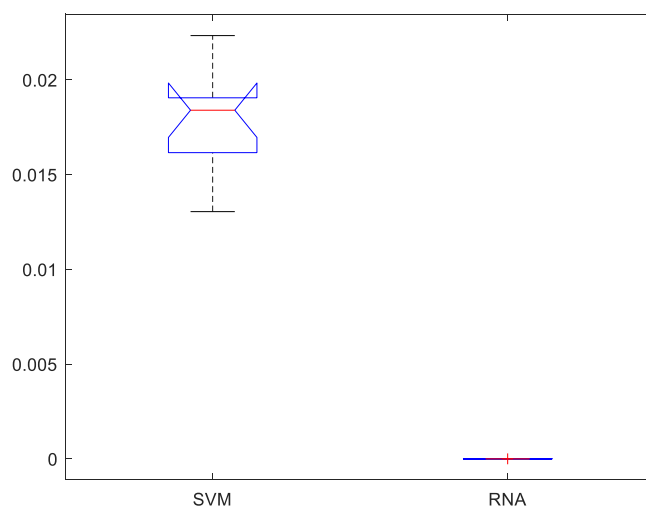


Figura 21. Diagrama de cajas y bigotes para error cuadrático medio del conjunto de datos con  $d = 35$ .

Como se puede observar en la Figura 20 y 21, tanto para el conjunto de datos Henon  $d = 15$  como para  $d = 35$ , la RNA presenta el valor de la mediana más próximo a 0, además de presentar un rango intercuartílico mínimo en comparación con el de la SVM.

## TEST DE COMPARACIÓN MÚLTIPLE.

Mediante la representación complementaria del test de comparación múltiple, se distinguen aquellas aproximaciones entre las que se producen diferencias estadísticamente significativas.

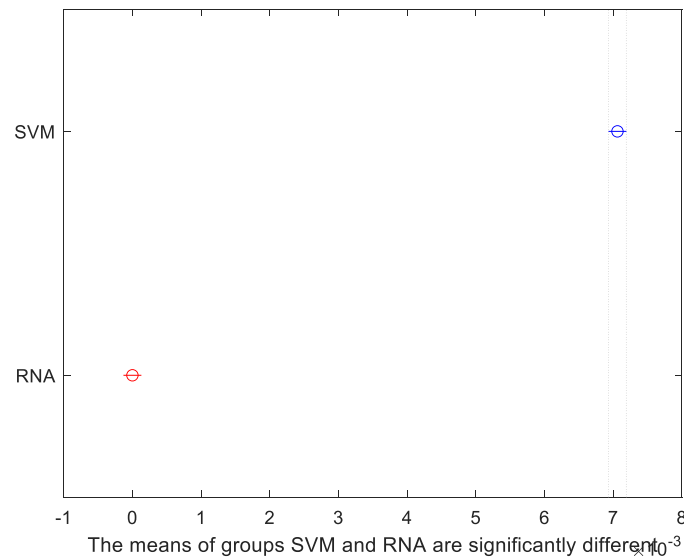


Figura 22. Representación del test de comparación múltiple para el conjunto de datos con  $d = 15$ .

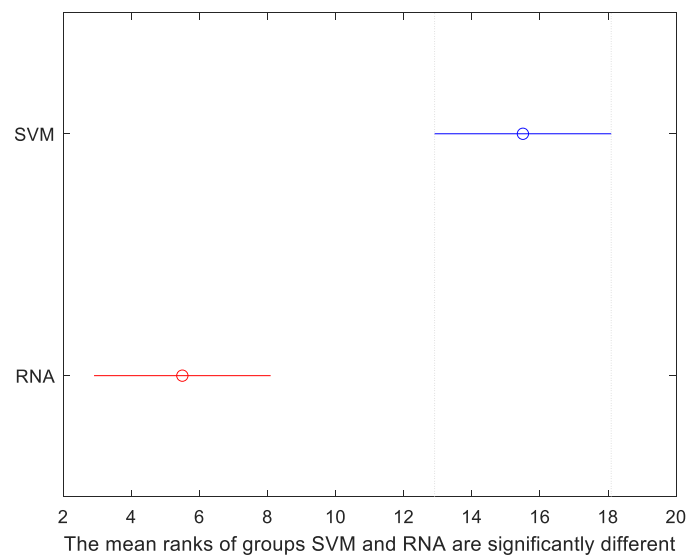


Figura 23. Representación del test de comparación múltiple para el conjunto de datos con  $d = 35$ .

Como era de esperar teniendo en cuenta el diagrama de cajas y bigotes y el test estadístico obtenido anteriormente, existen diferencias significativas entre la SVM polinómica y la RNA con algoritmo de aprendizaje Levenberg Marquardt.

## DISCUSIÓN Y CONCLUSIONES.

De acuerdo con los resultados alcanzados en el test de comparación, podemos concluir que existen diferencias estadísticamente significativas entre la SVM polinomial y la RNA, de forma que el mejor modelo para resolver el problema de regresión para el conjunto de datos Henon  $d = 15$  y  $d = 35$  es la RNA con algoritmo de aprendizaje Levenberg Marquardt con tamaño de capa oculta 45 y 15, respectivamente.