



TALLER DE MODELADO MATEMÁTICO II
PARTE II / 25P

TRABAJO FINAL

REDES NEURONALES ESTADÍSTICAMENTE INFORMADAS

ALAN BADILLO SALAS

Dr. Joaquín Delgado Fernández

CONTENIDO

INTRODUCCIÓN	3
AVANCES RELEVANTES Y ÁREAS DE EXPLORACIÓN	4
CASO DE ESTUDIO - CLASIFICACIÓN IRIS	11
FASE 1 - EXPLORACIÓN Y SEPARABILIDAD DE LAS CLASES	13
FASE 2 - RED NEURONAL PROFUNDA	16
FASE 3 - DNN ESTADÍSTICAMENTE INFORMADA (MANUAL)	20
FASE 4 - DNN ESTADÍSTICAMENTE INFORMADA (OVR)	22
CONCLUSIONES	24

¿Qué son las Redes Neuronales Estadísticamente Informadas?

Son **modelos de deep learning** que incorporan principios, estructuras o **transformaciones estadísticas** explícitamente en la arquitectura, el preprocesamiento, o la interpretación de los datos.

Se busca **evitar supuestos** ingenuos como la normalidad multivariada, o la linealidad simple, integrando herramientas estadísticas más robustas o **estructuras informadas** por la teoría del dominio.

(1) Aumento de Dimensionalidad con Variables Derivadas Estadísticamente

¿Qué es?

Agregar variables cuadráticas, interacciones, polinomios, estadísticas resumen, componentes principales, o funciones kernel antes de entrenar redes.

Método	Descripción
Polynomial Networks	Redes que reciben no solo las variables originales sino también sus combinaciones polinomiales.
Deep Feature Synthesis (AutoML tools / Featuretools)	Generan automáticamente interacciones y estadísticas cruzadas.
Higher-Order Neural Units (HONUs)	Redes con nodos que modelan interacciones cuadráticas o cúbicas directamente.
Functional ANOVA Embedding	Usa el análisis de varianza funcional para definir interacciones significativas.

(2) Espacios Latentes con Significado Estadístico

¿Qué es?

Transformar las variables de entrada a espacios donde los datos están mejor distribuidos sin imponer normalidad (como PCA lo hace bajo varianzas).

Método	Descripción
Normalizing Flows	Aprenden una transformación invertible que convierte la distribución original a una distribución simple (como la normal), pero sin perder información.
Variational Autoencoders (VAEs)	Encoders con prior no normal o basados en copulas.
Copula-based Neural Networks	Modelan explícitamente las dependencias entre variables sin asumir su distribución marginal.

(3) Normalización Informada (Statistical Normalization Layers)

¿Qué es?

Son transformaciones que se pueden aplicar en lugar de **StandardScaler** o **BatchNorm**.

Método	Descripción
Quantile Normalization Layers	Genera normalizaciones por agrupaciones cuantílicas
Rank Transformations	Genera transformaciones basadas en ponderaciones
Group-wise normalization	Separa por grupos estadísticos (por ejemplo, clases, regiones) antes de normalizar.

(4) Modelos con Componentes Estadísticos Internos

¿Qué es?

Son modelos que pueden ajustar componentes internos mediante redes neuronales, por ejemplo:

$$y \approx f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2)$$

donde f_1, f_2, f_{12} son pequeñas redes neuronales.

Modelo	Descripción
GLM-Nets	Redes que incorporan partes de modelos lineales generalizados.
Neural Additive Models (NAMs)	Redes donde cada variable tiene su propia subred, manteniendo interpretabilidad tipo regresión aditiva.
Spline-based models	Incorporan funciones splines dentro de la arquitectura (usado en medicina, bioestadística).

(5) Regularización basada en estadística

¿Qué es?

Usar la teoría estadística para regularizar modelos de **deep learning**.

Método	Descripción
Informative Priors	Utiliza los priors informativos en redes bayesianas.
Dropout no uniforme	Regularización informada por la varianza o la importancia estadística de las variables.
Sparse regression layers	Funciona como LASSO o Ridge dentro de una red.

(6) Modelos Cuadráticos / Interacciones Explícitas

¿Qué es?

Son estructuras que permiten expresar:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \dots$$

Modelo	Descripción
Factorization Machines (FMs)	Genera factorizaciones complejas
Field-aware FMs (FFMs)	Modelos de factorización más avanzados
PolyNet, Deep Cross Networks	Funciona como en los modelos de recomendación
Tensor Neural Networks	Aprenden interacciones cruzadas de alto orden.

(7) Modelos Deep + Estadísticos en Series de Tiempo

¿Qué es?

Son modelos que permiten usar información temporal para las predicciones (**forecasting**).

Modelo	Descripción
DeepAR	Redes recurrentes con modelos probabilísticos por serie.
Deep State Space Models	Combinación de redes con modelos espacio-estado estadísticos.

¿Qué es? Problema de clasificación:

4 características continuas / **3 respuestas** binarias
150 muestras (50 de cada clase)

iris setosa



petal sepal

iris versicolor

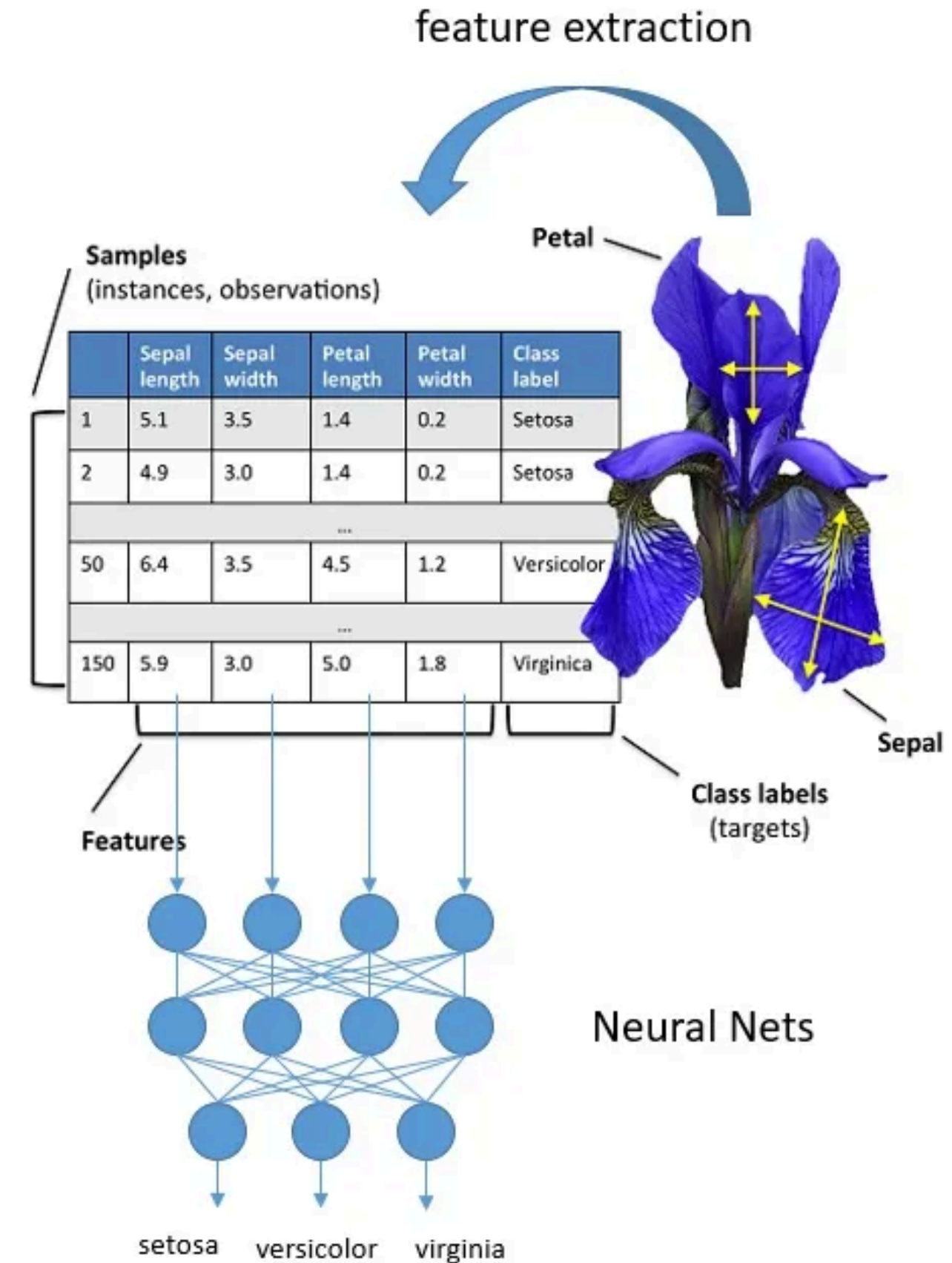


petal sepal

iris virginica



petal sepal



Solución: Red neuronal profunda

1 Capa de entrada (4 características)

3 Capas ocultas (ReLU con 16, 16 y 8 nodos)

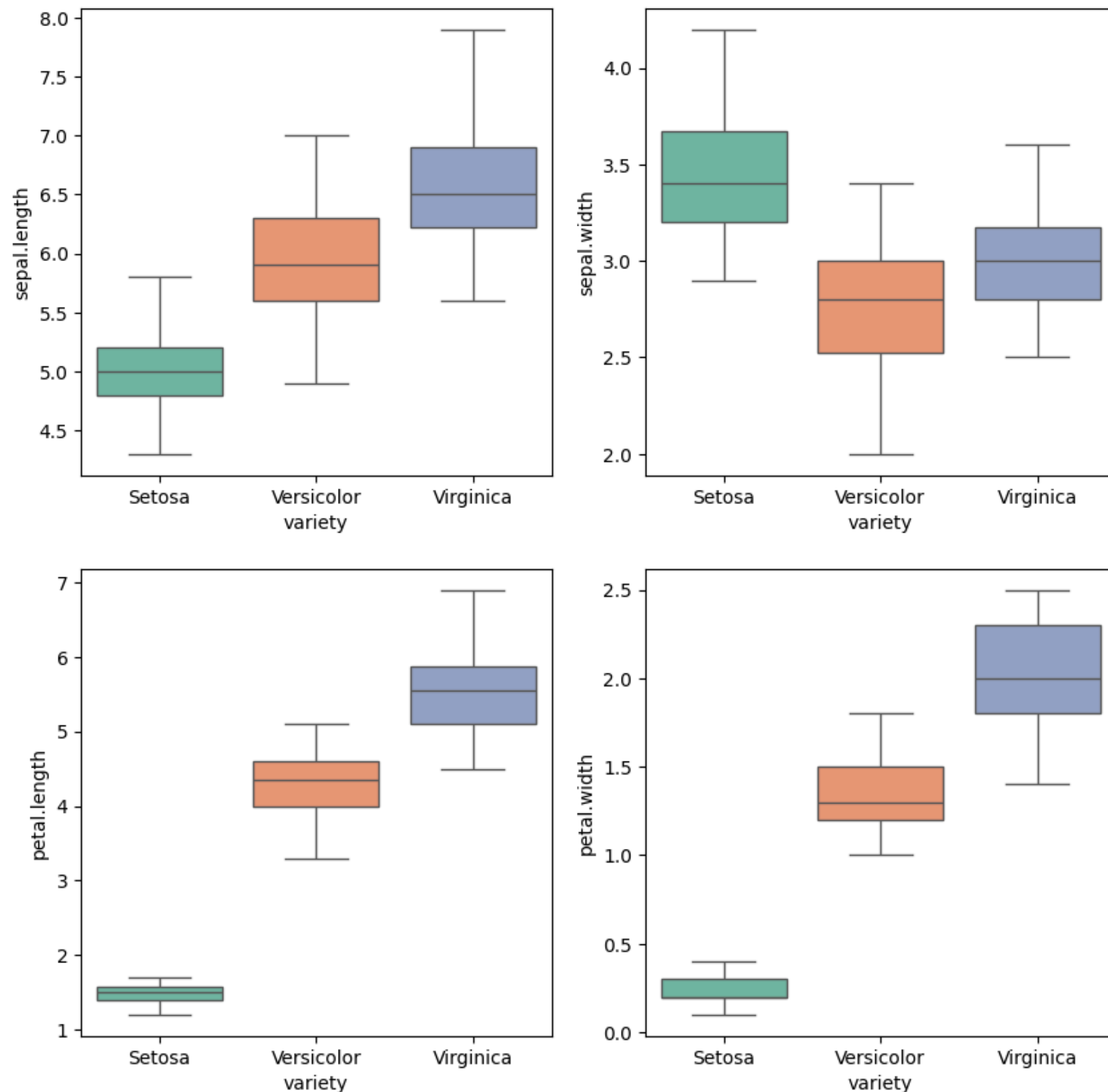
1 Capa de salida (Softmax con 3 nodos)

Capa	Número de nodos	Activación	Descripción
Entrada	4	-	4 características (ancho y largo de pétalo y sépalo)
Oculto 1	16	ReLU	16 nodos en la primera capa oculta
Oculto 2	16	ReLU	16 nodos en la segunda capa oculta
Oculto 3	8	ReLU	8 nodos en la tercera capa oculta
Salida	3	Softmax	3 nodos en la capa de salida (Setosa, Versicolor, Virgínica)

FASE 1 - EXPLORACIÓN Y SEPARABILIDAD DE LAS CLASES

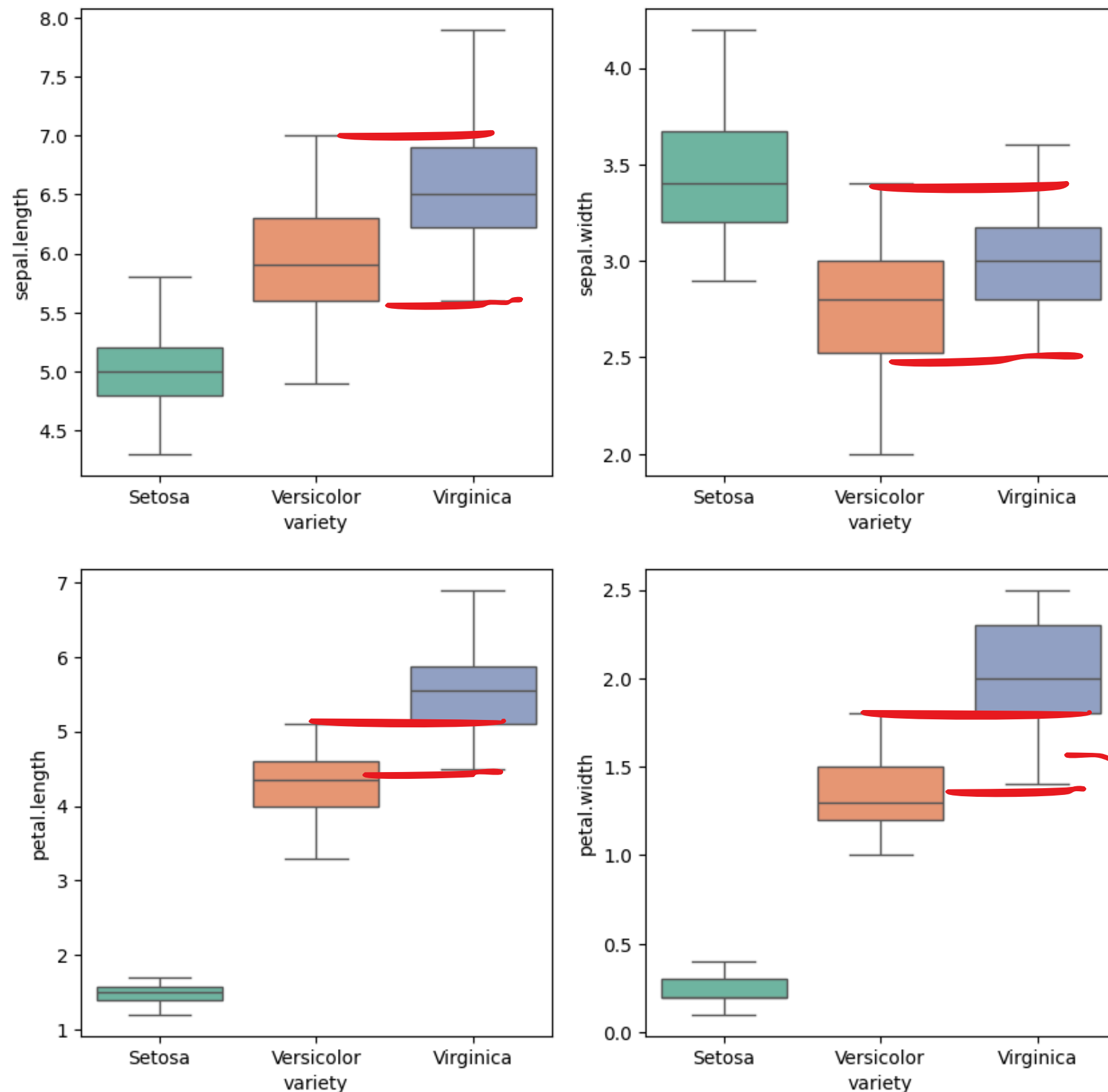
	sepal.length	sepal.width	petal.length	petal.width	variety
29	4.7	3.2	1.6	0.2	Setosa
101	5.8	2.7	5.1	1.9	Virginica
2	4.7	3.2	1.3	0.2	Setosa
17	5.1	3.5	1.4	0.3	Setosa
117	7.7	3.8	6.7	2.2	Virginica
4	5.0	3.6	1.4	0.2	Setosa
34	4.9	3.1	1.5	0.2	Setosa
75	6.6	3.0	4.4	1.4	Versicolor
76	6.8	2.8	4.8	1.4	Versicolor
146	6.3	2.5	5.0	1.9	Virginica
62	6.0	2.2	4.0	1.0	Versicolor
61	5.9	3.0	4.2	1.5	Versicolor
23	5.1	3.3	1.7	0.5	Setosa
111	6.4	2.7	5.3	1.9	Virginica
104	6.5	3.0	5.8	2.2	Virginica
82	5.8	2.7	3.9	1.2	Versicolor
70	5.9	3.2	4.8	1.8	Versicolor
72	6.3	2.5	4.9	1.5	Versicolor
7	5.0	3.4	1.5	0.2	Setosa
120	6.9	3.2	5.7	2.3	Virginica

Total	
variety	
Setosa	50
Versicolor	50
Virginica	50



Dispersión de las características respecto a las clases

Se observa una alta separación de la clase **Setosa** (verde), sin embargo, las clases **Versicolor** (naranja) y **Virgínica** (morado) se solapan en todos los ejes, teniendo una mayor separación en el **Petal Length** (ancho de pétalo).



Dispersión de las características respecto a las clases

Se observa una alta separación de la clase **Setosa** (verde), sin embargo, las clases **Versicolor** (naranja) y **Virgínica** (morado) se solapan en todos los ejes, teniendo una mayor separación en el **Petal Length** (ancho de pétalo).

D zona de confusión

X

	x1	x2	x3	x4
0	1.4	0.2	5.1	3.5
1	1.4	0.2	4.9	3.0
2	1.3	0.2	4.7	3.2
3	1.5	0.2	4.6	3.1
4	1.4	0.2	5.0	3.6

	Muestras	Dimensión
X (train)	120	4
X (test)	30	4
y (train)	120	3
y (test)	30	3

y

	y1	y2	y3
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0















Layer (type)	Output Shape	Param #
dense (Dense)	(None, 16)	80
dense_1 (Dense)	(None, 16)	272
dense_2 (Dense)	(None, 8)	136
dense_3 (Dense)	(None, 3)	27

1/1  0s 87ms/step - accuracy: 0.8667 - loss: 0.3480

[0.3480192720890045, 0.8666666746139526]

87.1

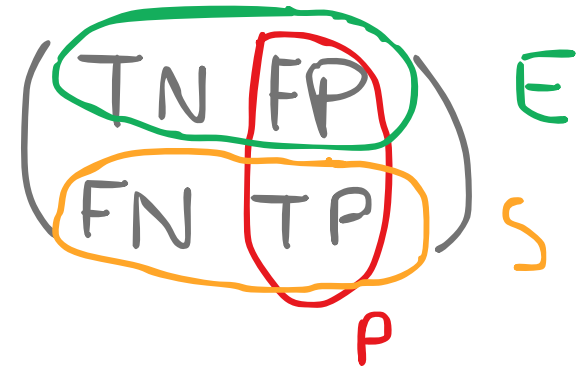
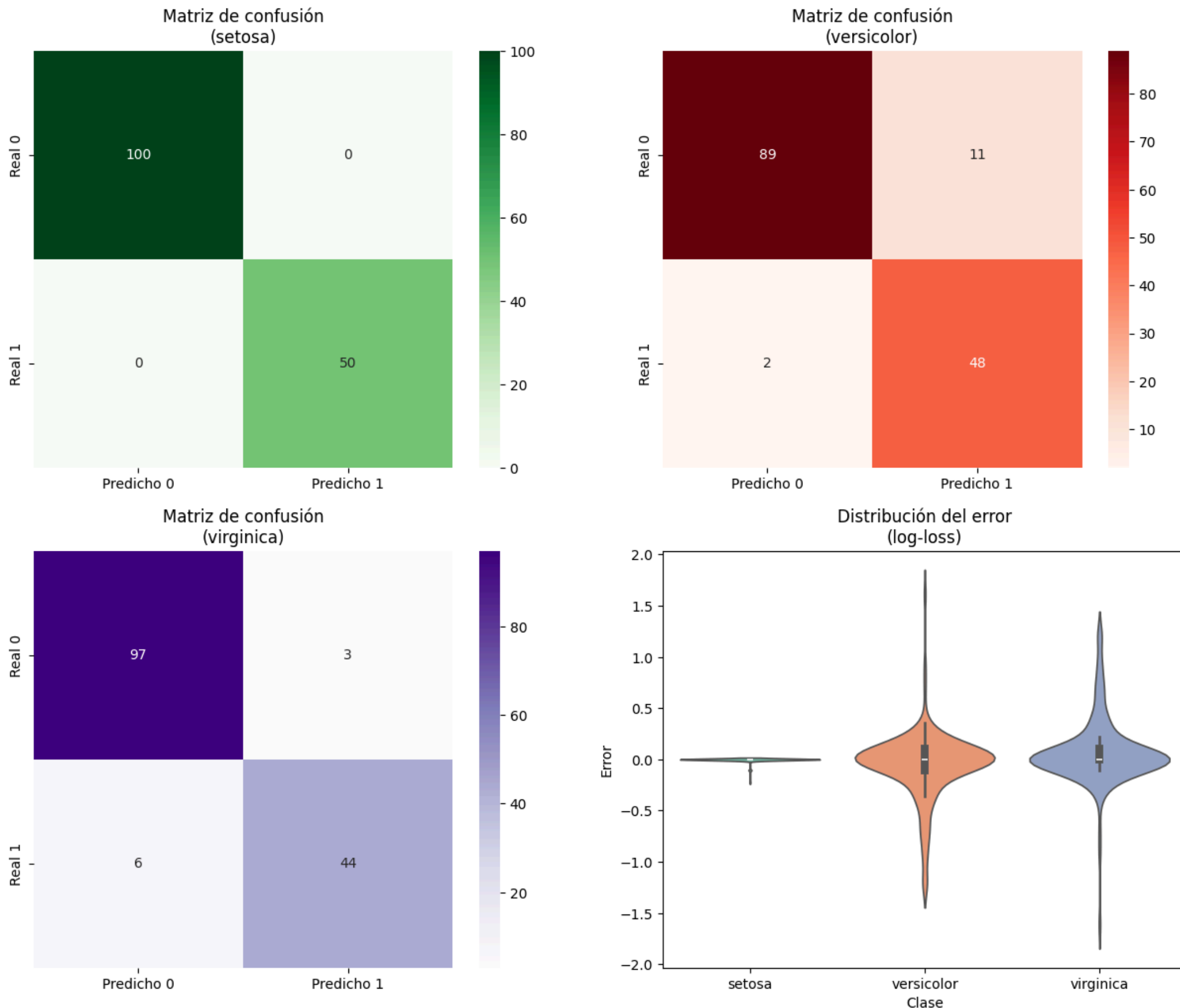
Se observa un aprendizaje real del 86.6%

Epoch 1/100
6/6  0s 3ms/step - accuracy: 0.2787 - loss: 5.2651
Epoch 2/100
6/6  0s 2ms/step - accuracy: 0.5644 - loss: 1.0264
Epoch 3/100
6/6  0s 2ms/step - accuracy: 0.6476 - loss: 0.9639
Epoch 4/100
6/6  0s 2ms/step - accuracy: 0.7081 - loss: 0.9058
Epoch 5/100
6/6  0s 2ms/step - accuracy: 0.6630 - loss: 0.8934
Epoch 6/100
6/6  0s 2ms/step - accuracy: 0.6831 - loss: 0.8501
Epoch 7/100
6/6  0s 2ms/step - accuracy: 0.7046 - loss: 0.8278
Epoch 8/100
6/6  0s 2ms/step - accuracy: 0.6539 - loss: 0.8025
Epoch 9/100
6/6  0s 2ms/step - accuracy: 0.6576 - loss: 0.7922
Epoch 10/100
6/6  0s 2ms/step - accuracy: 0.6592 - loss: 0.7720
Epoch 11/100
6/6  0s 2ms/step - accuracy: 0.5779 - loss: 0.7644
Epoch 12/100
6/6  0s 2ms/step - accuracy: 0.5079 - loss: 0.7283
Epoch 13/100
...
Epoch 99/100
6/6  0s 2ms/step - accuracy: 0.9499 - loss: 0.3600
Epoch 100/100
6/6  0s 2ms/step - accuracy: 0.9156 - loss: 0.3520

FASE 2 - RED NEURONAL PROFUNDA

	setosa	setosa (p)	versicolor	versicolor (p)	virginica	virginica (p)
102	0	0.0	0	0.1	1	0.9
82	0	0.0	1	1.0	0	0.0
74	0	0.0	1	1.0	0	0.0
101	0	0.0	0	0.1	1	0.9
24	1	1.0	0	0.0	0	0.0
49	1	1.0	0	0.0	0	0.0
92	0	0.0	1	0.9	0	0.1
104	0	0.0	0	0.1	1	0.9
46	1	1.0	0	0.0	0	0.0
75	0	0.0	1	1.0	0	0.0
143	0	0.0	0	0.1	1	0.9
45	1	1.0	0	0.0	0	0.0
55	0	0.0	1	0.8	0	0.2
40	1	1.0	0	0.0	0	0.0
23	1	1.0	0	0.0	0	0.0
114	0	0.0	0	0.1	1	0.9
3	1	1.0	0	0.0	0	0.0
96	0	0.0	1	0.9	0	0.1
110	0	0.0	0	0.7	1	0.3
39	1	1.0	0	0.0	0	0.0

	setosa	setosa (p)	versicolor	versicolor (p)	virginica	virginica (p)
102	0	0	0	0	1	1
82	0	0	1	1	0	0
74	0	0	1	1	0	0
101	0	0	0	0	1	1
24	1	1	0	0	0	0
49	1	1	0	0	0	0
92	0	0	1	1	0	0
104	0	0	0	0	1	1
46	1	1	0	0	0	0
75	0	0	1	1	0	0
143	0	0	0	0	1	1
45	1	1	0	0	0	0
55	0	0	1	1	0	0
40	1	1	0	0	0	0
23	1	1	0	0	0	0
114	0	0	0	0	1	1
3	1	1	0	0	0	0
96	0	0	1	1	0	0
110	0	0	0	1	1	0
39	1	1	0	0	0	0



Matrices de confusión

Setosa

- **Precisión alta** (TP=50 | FP=0)
- **Sensibilidad alta** (TP=50 | FN=0)
- **Especificidad alta** (TN=100 | FP=0)

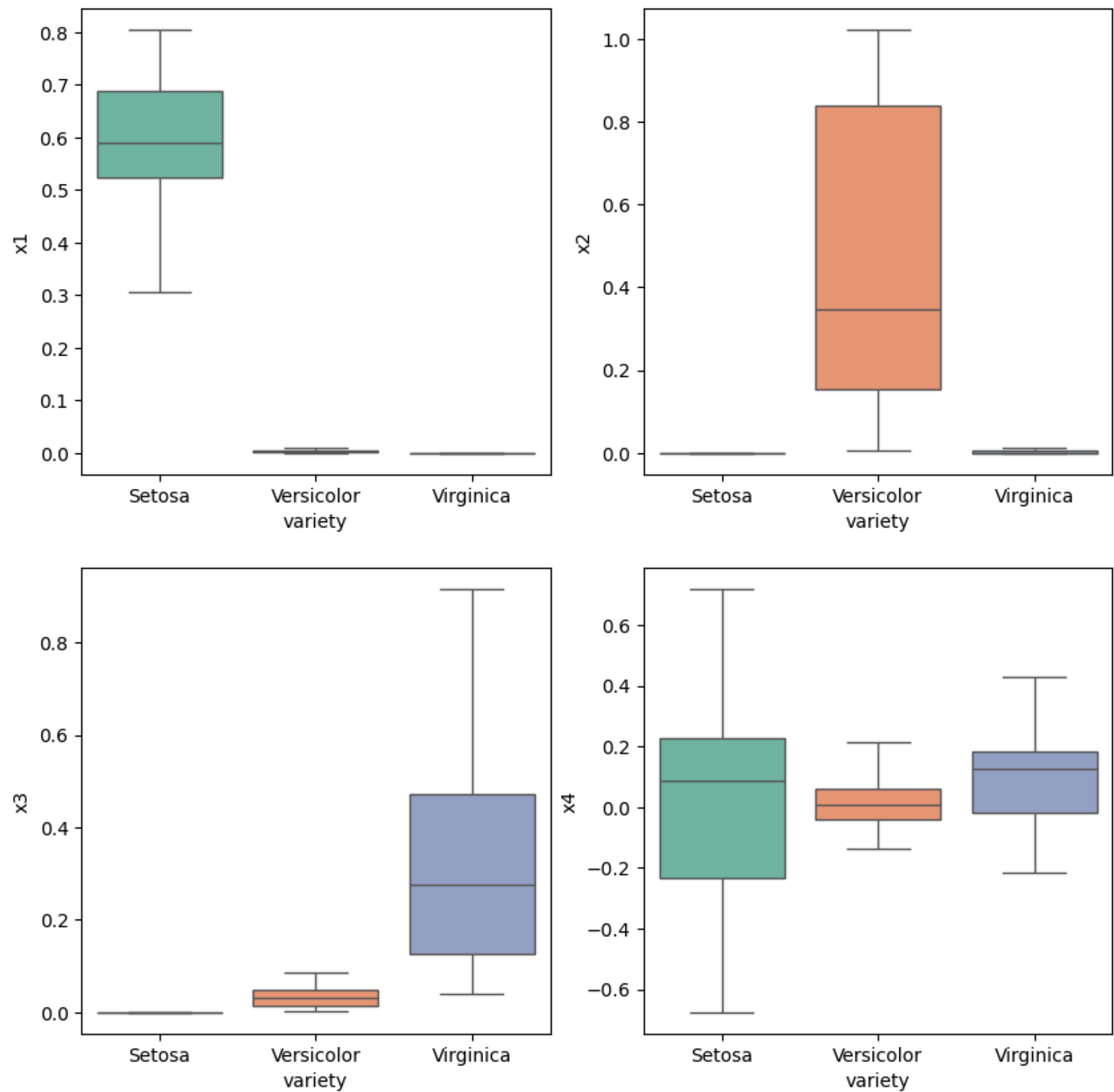
Versicolor

- Precisión baja (TP=48 | FP=11)
- **Sensibilidad alta** (TP=48 | FN=2)
- Especificidad baja (TN=89 | FP=11)

Virgínica

- **Precisión alta** (TP=44 | FP=3)
- Sensibilidad media (TP=44 | FN=6)
- **Especificidad alta** (TN=97 | FP=3)

FASE 3 - DNN ESTADÍSTICAMENTE INFORMADA (MANUAL)



	x1	x2	x3	x4	x5	x6	x7	x8
0	1.4	0.2	5.1	3.5	0.689593	3.956157e-10	0.000007	-1.012109
1	1.4	0.2	4.9	3.0	0.689593	3.956157e-10	0.000007	0.188400
2	1.3	0.2	4.7	3.2	0.805827	1.514370e-11	0.000005	0.236297
3	1.5	0.2	4.6	3.1	0.588917	6.245790e-09	0.000011	0.165577
4	1.4	0.2	5.0	3.6	0.689593	3.956157e-10	0.000007	-0.675357

6/6 0s 3ms/step - accuracy: 0.6130 - loss: 0.9003
Epoch 12/100

6/6 0s 3ms/step - accuracy: 0.6318 - loss: 0.8693
Epoch 13/100

...

Epoch 99/100

6/6 0s 3ms/step - accuracy: 0.9962 - loss: 0.0987
Epoch 100/100

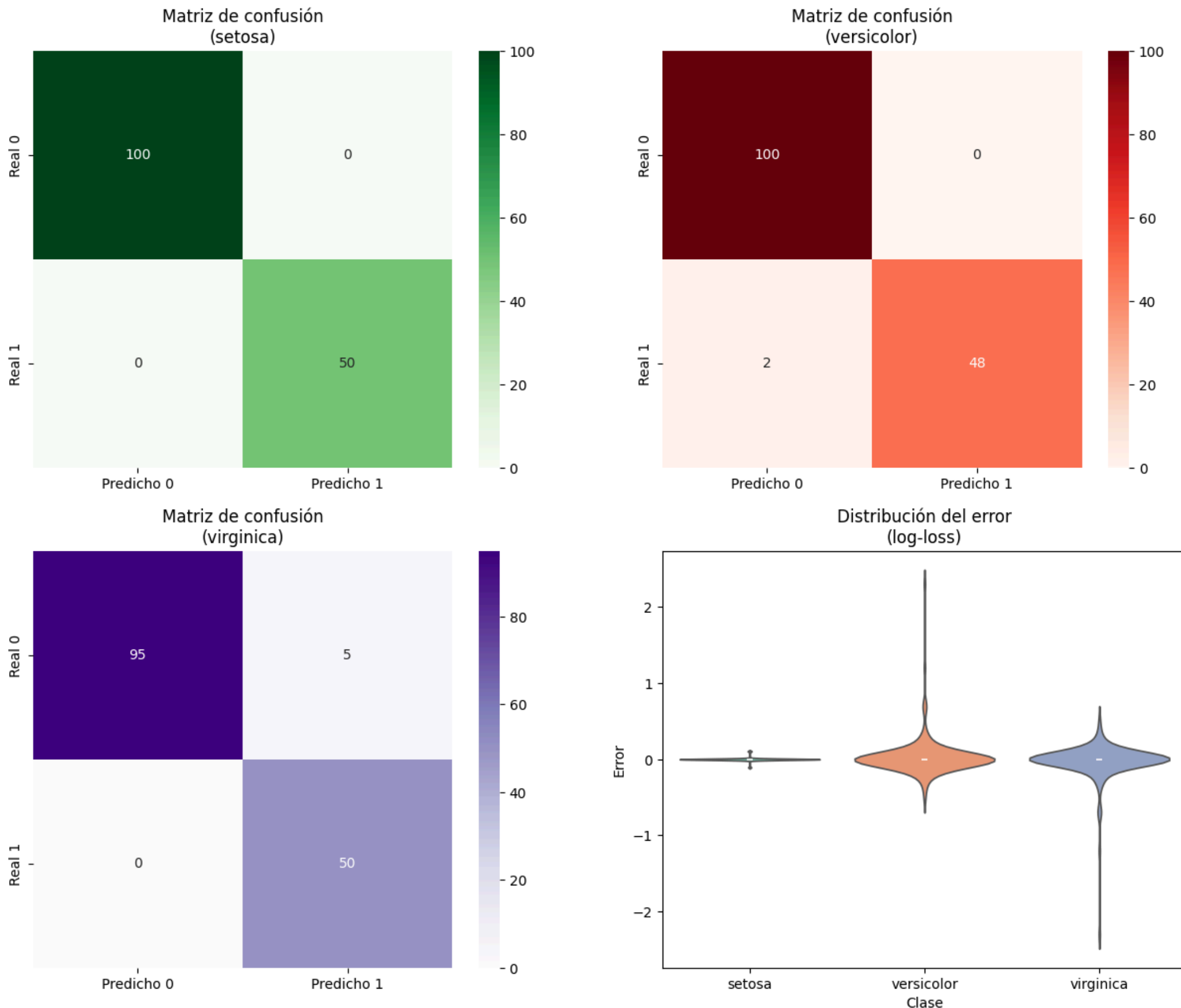
6/6 0s 2ms/step - accuracy: 0.9757 - loss: 0.1122

1/1 0s 88ms/step - accuracy: 0.9667 - loss: 0.1392

[0.13919782638549805, 0.9666666388511658]

97%

Observamos un aprendizaje real del 96.6%



Matrices de confusión

Setosa

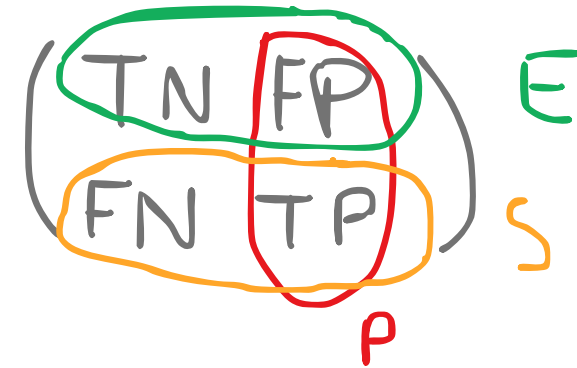
- **Precisión alta** (TP=50 | FP=0)
- **Sensibilidad alta** (TP=50 | FN=0)
- **Especificidad alta** (TN=100 | FP=0)

Versicolor

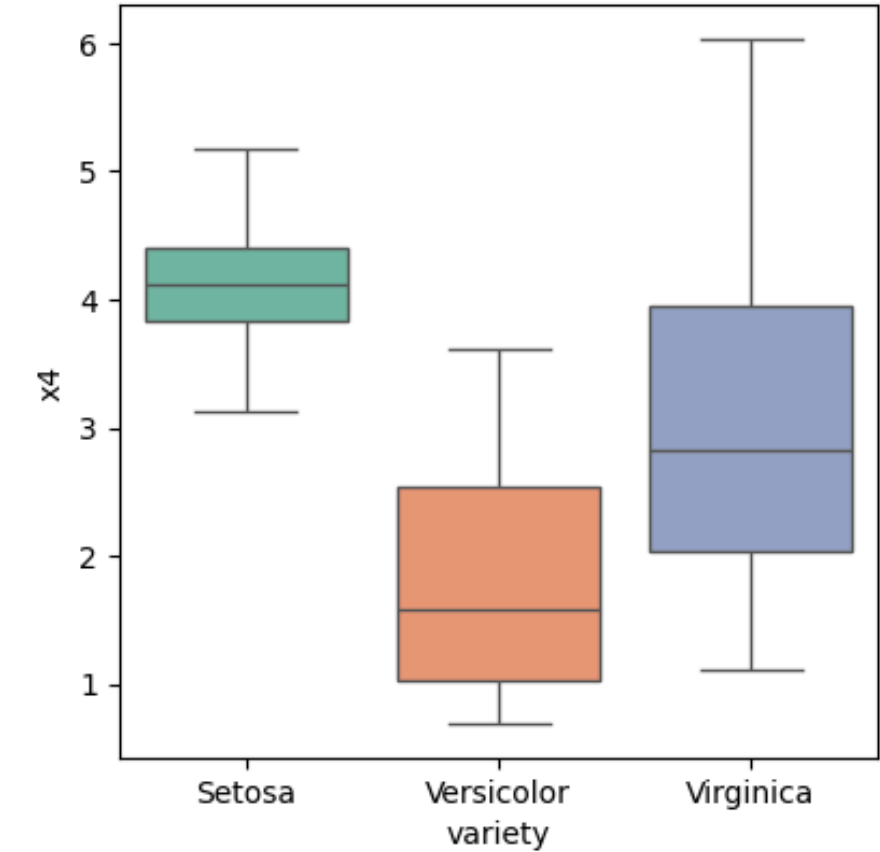
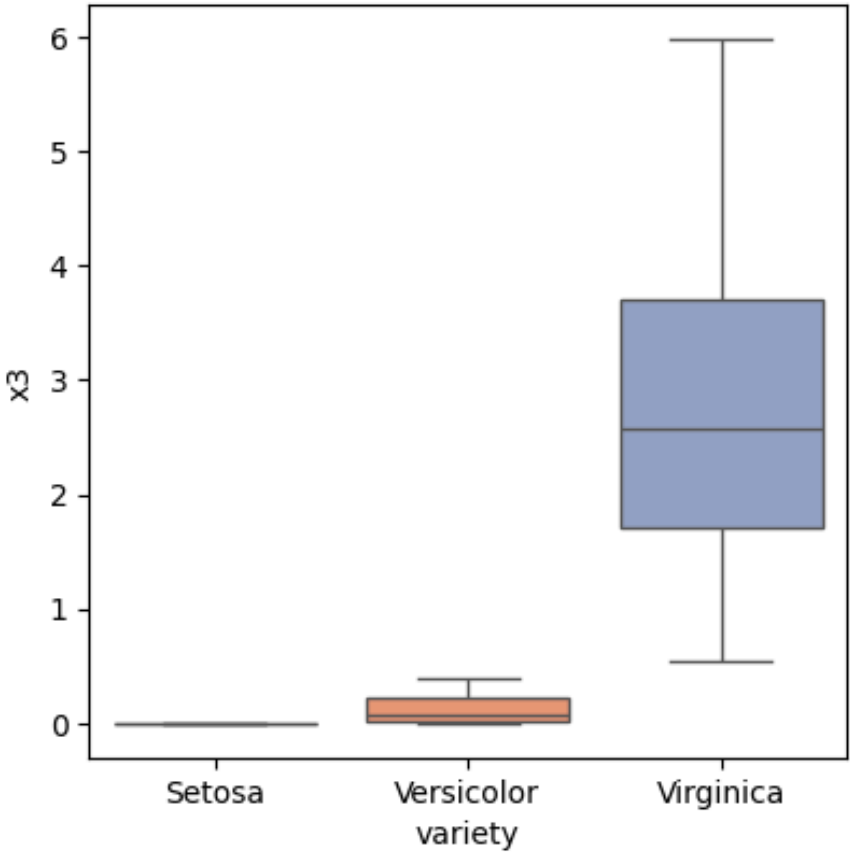
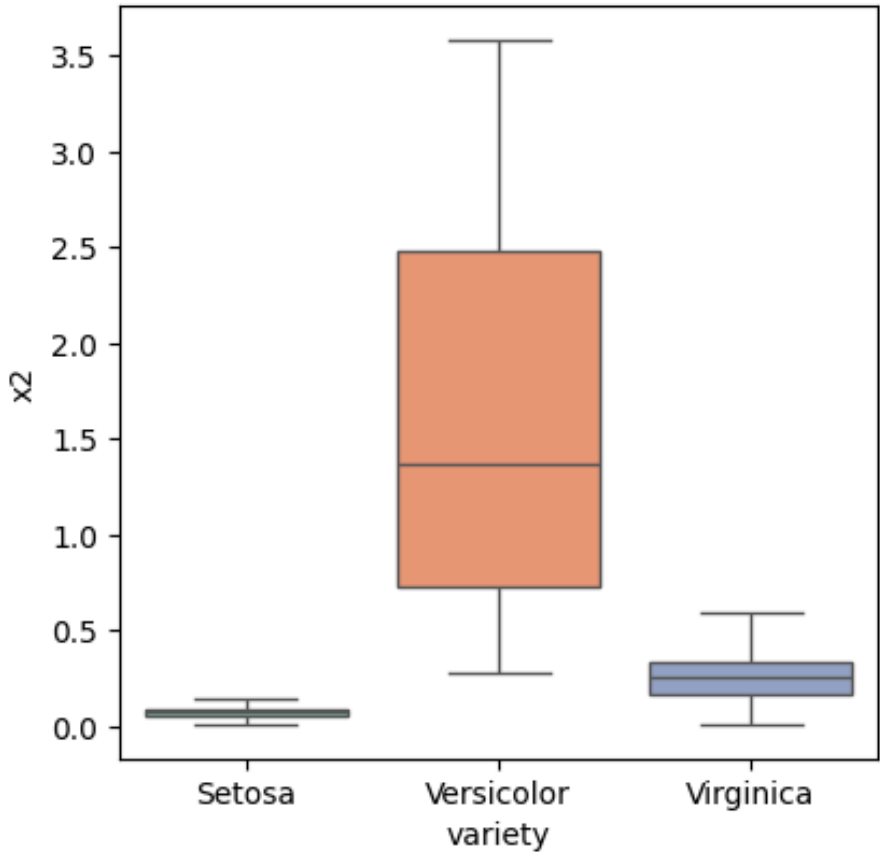
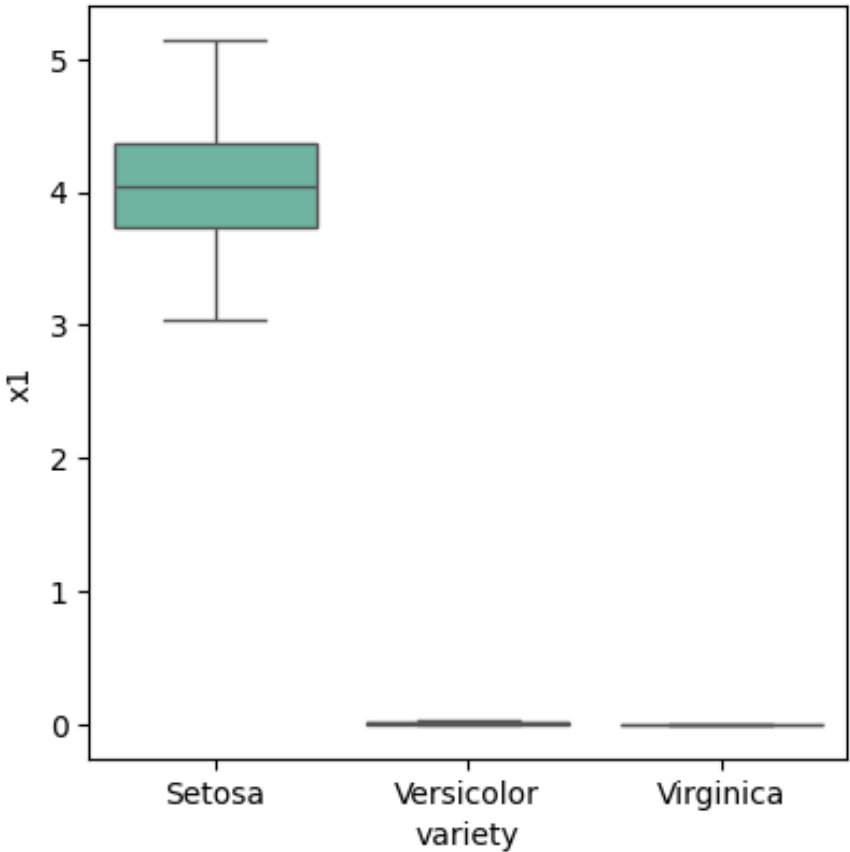
- **Precisión alta** (TP=48 | FP=0)
- **Sensibilidad alta** (TP=48 | FN=2)
- **Especificidad alta** (TN=100 | FP=0)

Virgínica

- **Precisión alta** (TP=50 | FP=5)
- **Sensibilidad alta** (TP=50 | FN=0)
- **Especificidad alta** (TN=95 | FP=5)



FASE 4 - DNN ESTADÍSTICAMENTE INFORMADA (OVR)



	x1	x2	x3	x4	x5	x6	x7	x8
0	1.4	0.2	5.1	3.5	4.200271	0.063124	0.000003	4.263398
1	1.4	0.2	4.9	3.0	3.941659	0.089701	0.000005	4.031365
2	1.3	0.2	4.7	3.2	4.466517	0.051071	0.000004	4.517591
3	1.5	0.2	4.6	3.1	3.879415	0.090080	0.000007	3.969502
4	1.4	0.2	5.0	3.6	4.339543	0.054067	0.000003	4.393614

6/6 0s 2ms/step - accuracy: 0.7357 - loss: 0.7991
Epoch 12/100

6/6 0s 2ms/step - accuracy: 0.8299 - loss: 0.7396
Epoch 13/100

...
Epoch 99/100

6/6 0s 2ms/step - accuracy: 0.9805 - loss: 0.0953
Epoch 100/100

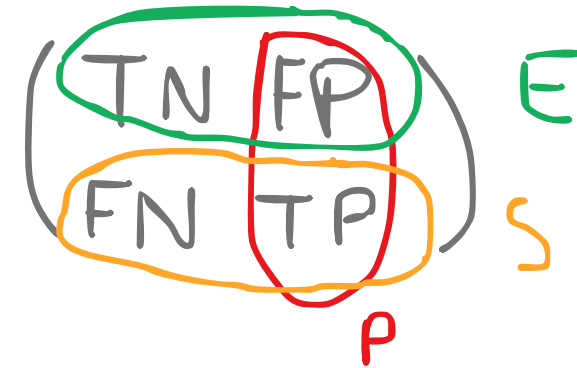
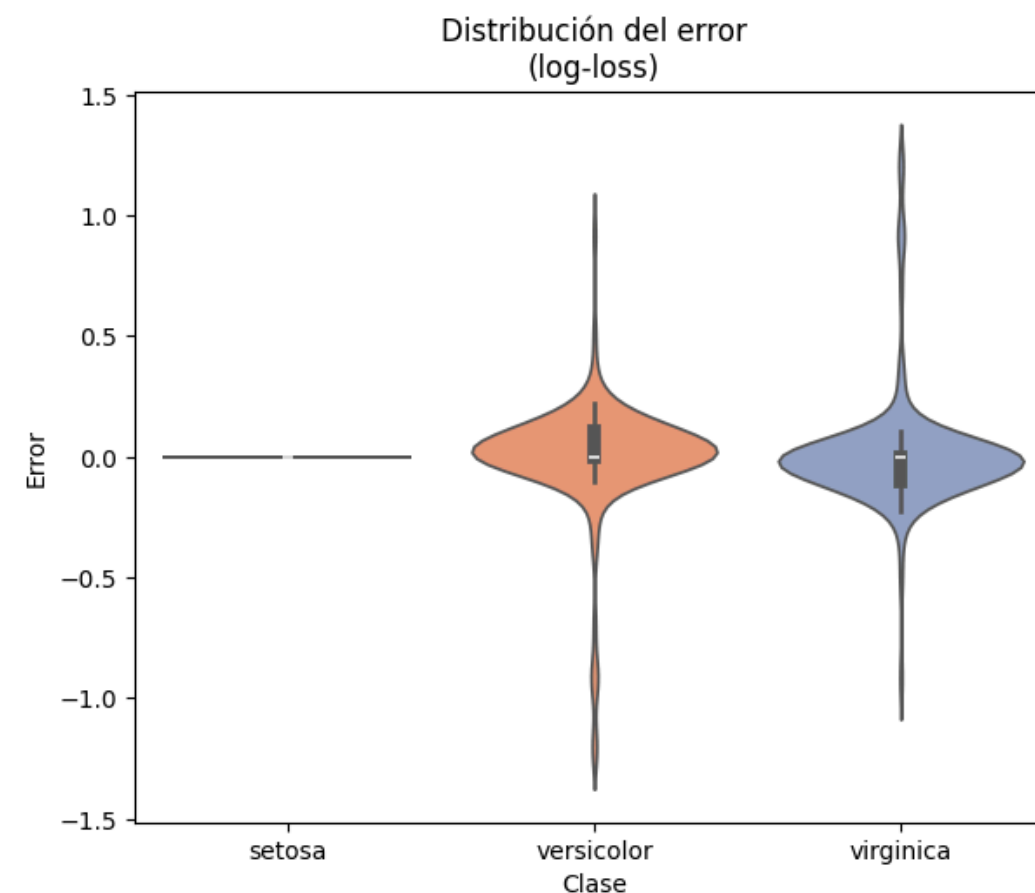
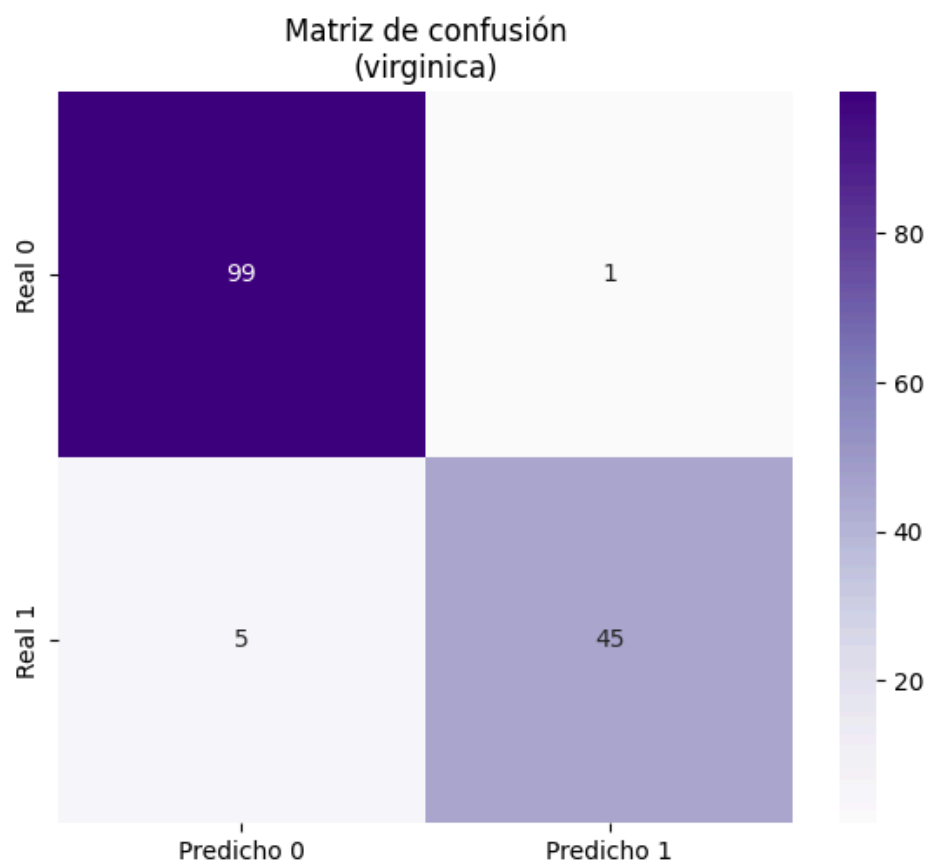
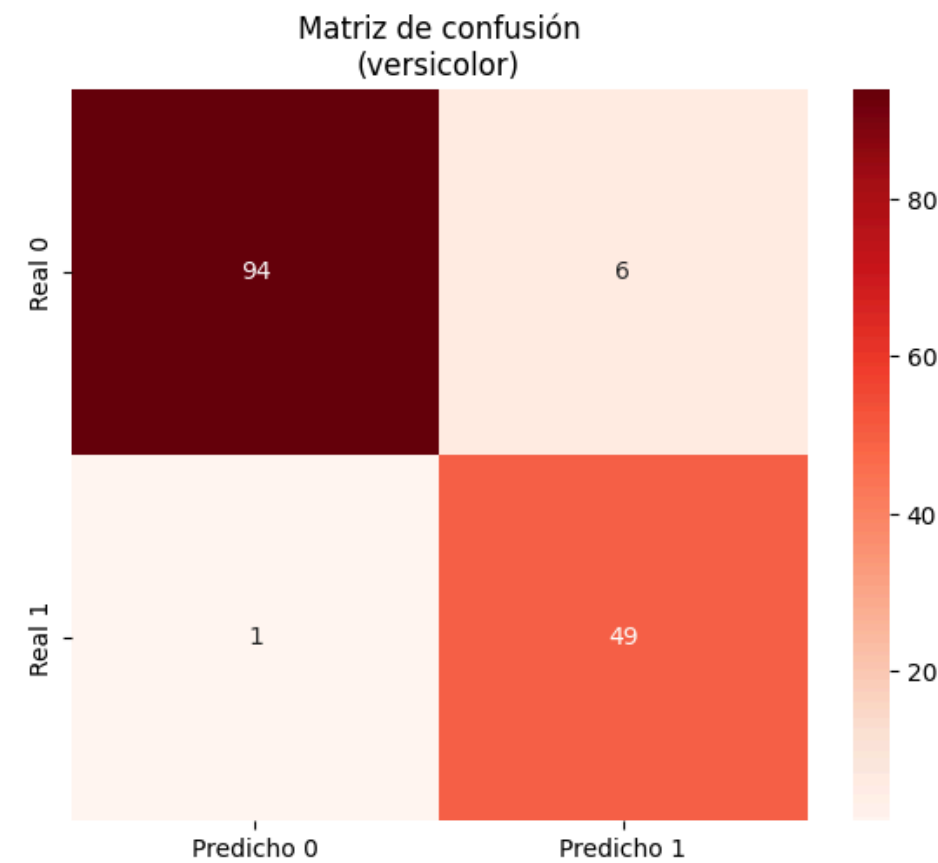
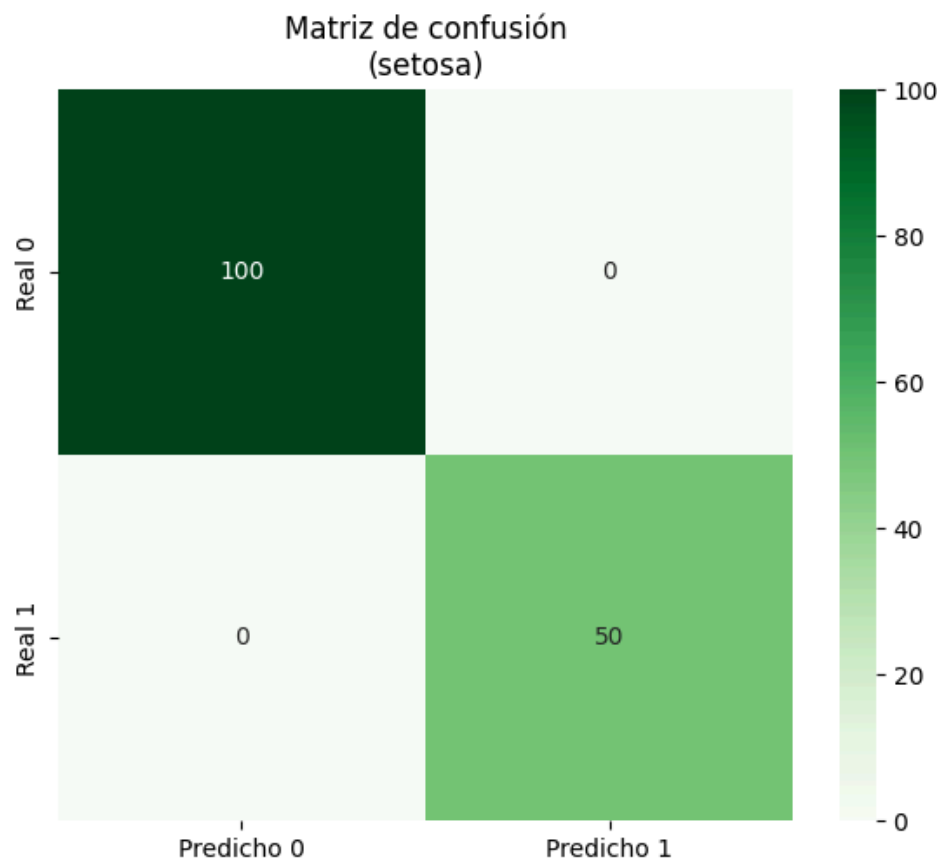
6/6 0s 2ms/step - accuracy: 0.9598 - loss: 0.0951

1/1 0s 84ms/step - accuracy: 0.9000 - loss: 0.1444

[0.14439982175827026, 0.8999999761581421]

90%

Observamos un aprendizaje real del 90%



Matrices de confusión

Setosa

- **Precisión alta** (TP=50 | FP=0)
- **Sensibilidad alta** (TP=50 | FN=0)
- **Especificidad alta** (TN=100 | FP=0)

Versicolor

- **Precisión alta** (TP=49 | FP=6)
- **Sensibilidad alta** (TP=49 | FN=1)
- **Especificidad alta** (TN=94 | FP=6)

Virgínica

- **Precisión alta** (TP=45 | FP=1)
- **Sensibilidad alta** (TP=45 | FN=5)
- **Especificidad alta** (TN=99 | FP=1)

Conclusiones

1. Al **aumentar la información estadística**, las redes neuronales profundas con arquitectura similar lograron mejorar el aprendizaje
2. El **método manual arrojó mejores resultados**, pero no es un método automático, ya que requiere el ajuste manual de las variables aumentadas
3. El método OvR logró mejorar los resultados, y este **si es un método automático** que determina las variables aumentadas
4. Otros métodos listados deberían ser estudiados para mejorar el aprendizaje en las redes neuronales