

# UNIVERSIDAD DE GUADALAJARA



## CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍAS

### Sistemas Inteligentes IV

#### Actividad 5. Clasificadores y Métricas de Clasificación

Pacheco Quintero Marco Antonio      213535019

23 de abril de 2021

Semestre 2021A

Sección D01

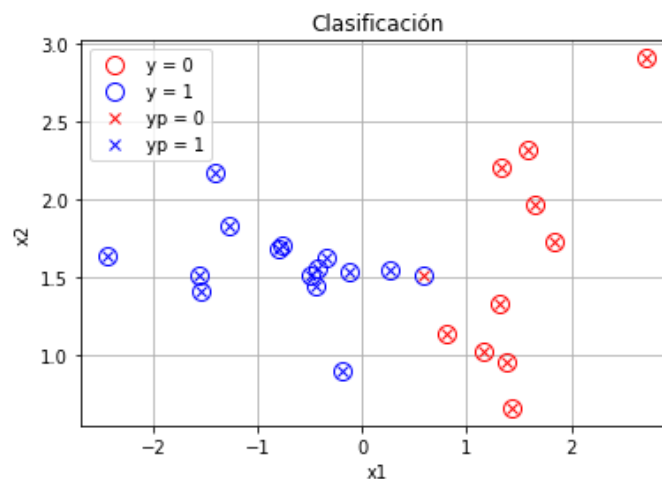
## Objetivo

- Para la parte 1, realizar un programa en Python para aplicar una clasificación binaria a los datos de los archivos adjuntos a la actividad. Utilizar las herramientas de sklearn para el modelo de clasificación.
- Para la parte 2, plantear algún problema de clasificación binaria y realiza un programa en Python para resolverlo. Utilizar las herramientas de sklearn para el modelo de clasificación.
- Para cada caso utilizar la métrica de *F1 score macro average* para demostrar que el entrenamiento tiene una buena generalización. Además, utilizar los siguientes modelos de clasificación:
  - Árboles de decisión
  - K vecinos más cercanos
  - Máquinas de soporte vectorial
  - Perceptrón Multicapa

## Resultados

### Parte 1

- Para el archivo adjunto “[df\\_clasificacion\\_1.csv](#)” se muestran a continuación los mejores resultados obtenidos para cada modelo de clasificación.
  - Árboles de decisión



Los resultados de las métricas globales son:

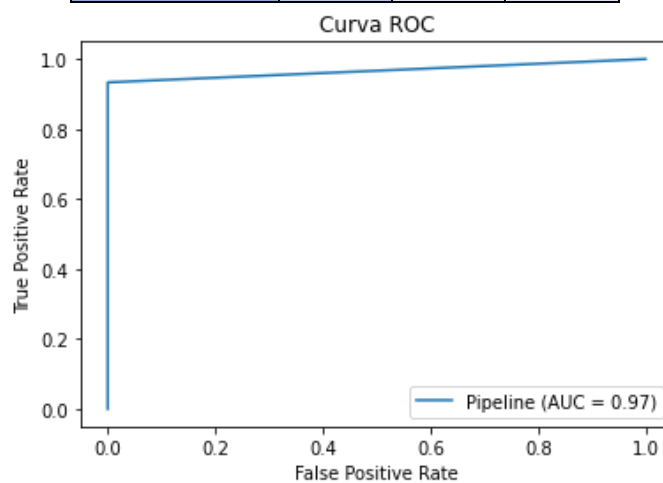
$F1_{\text{entrenamiento}} = 0.97$   
 $F1_{\text{generalización}} = 0.96$   
 $\text{Exactitud (Accuracy)} = 0.96$

En la tabla siguiente se muestran los resultados de las métricas locales:

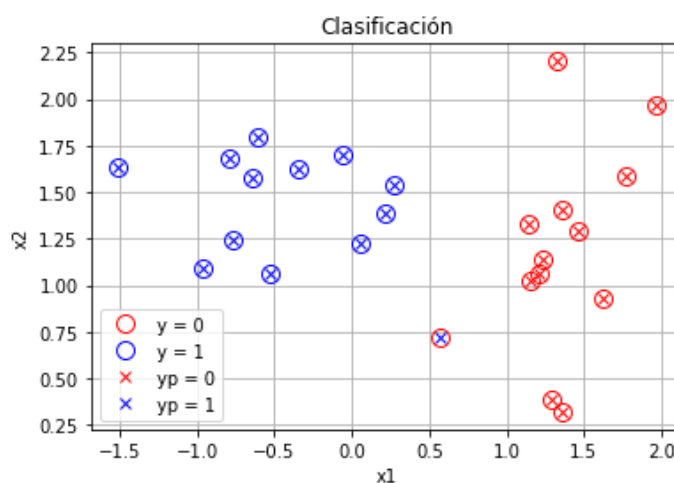
	Precisión	Sensibilidad	Coeficiente F1
0	0.91	1	0.95
1	1	0.93	0.97

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	10	0
	1	1	14



- K vecinos más cercanos



Los resultados de las métricas globales son:

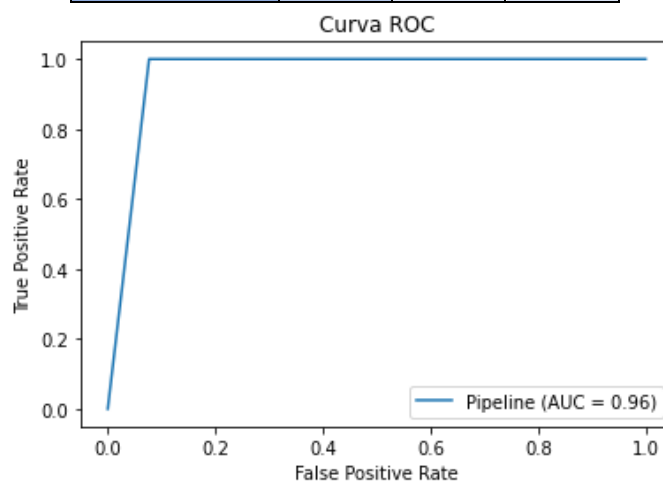
$F1_{\text{entrenamiento}} = 0.96$   
 $F1_{\text{generalización}} = 0.96$   
 $\text{Exactitud (Accuracy)} = 0.96$

En la tabla siguiente se muestran los resultados de las métricas locales:

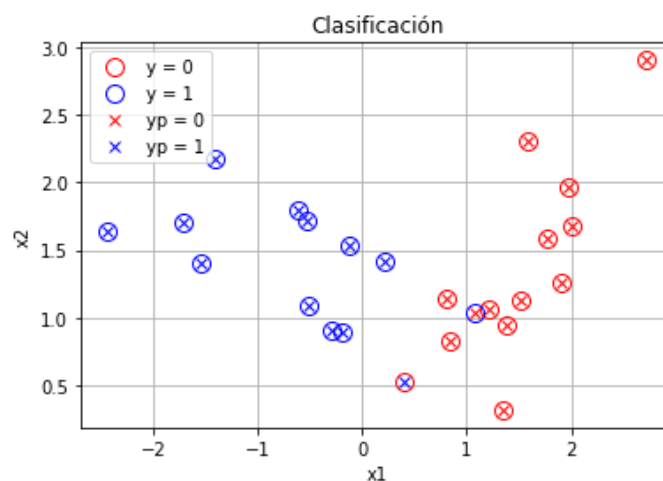
	Precisión	Sensibilidad	Coeficiente F1
0	1	0.92	0.96
1	0.92	1	0.96

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	12	1
	1	0	12



- Máquinas de soporte vectorial



Los resultados de las métricas globales son:

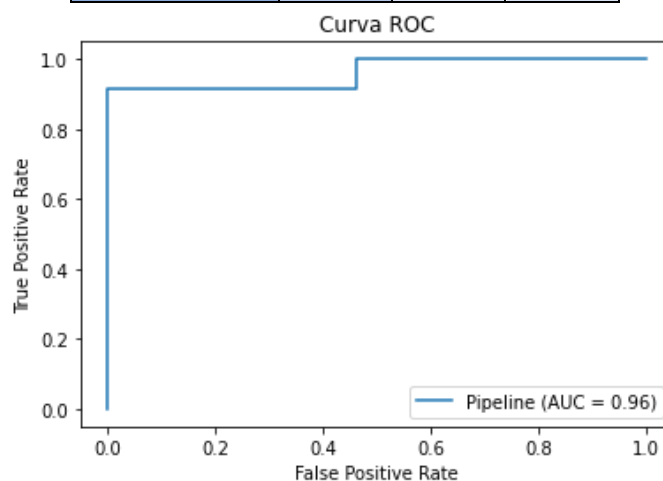
$F1_{\text{entrenamiento}} = 0.92$   
 $F1_{\text{generalización}} = 0.92$   
 $\text{Exactitud (Accuracy)} = 0.92$

En la tabla siguiente se muestran los resultados de las métricas locales:

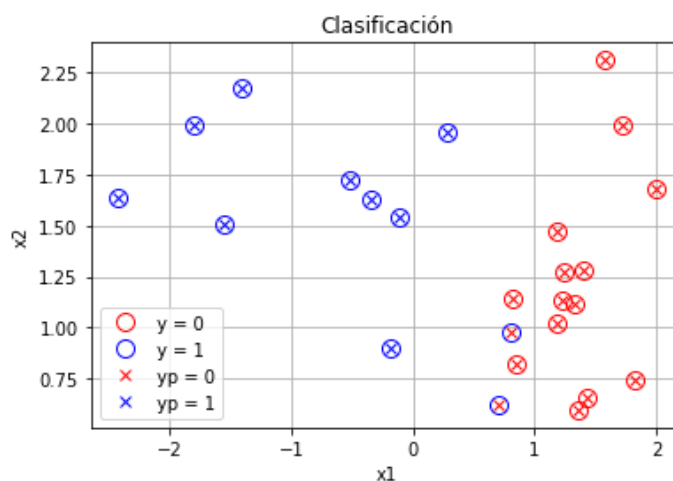
	Precisión	Sensibilidad	Coeficiente F1
0	0.92	0.92	0.92
1	0.92	0.92	0.92

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	12	1
	1	1	11



#### ○ Perceptrón Multicapa



Los resultados de las métricas globales son:

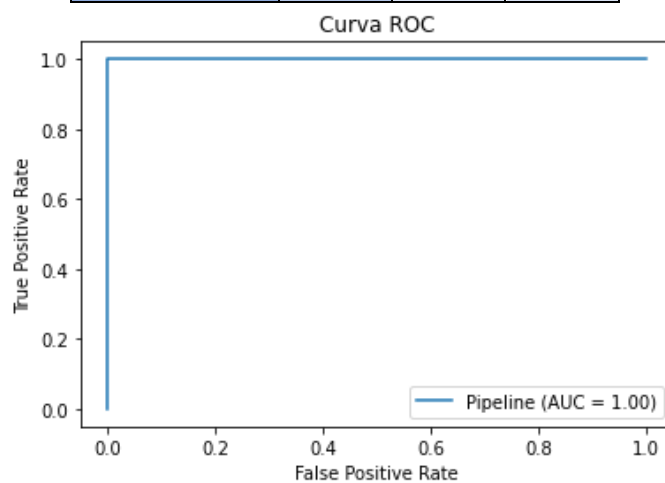
$F1_{\text{entrenamiento}} = 0.92$   
 $F1_{\text{generalización}} = 0.92$   
 $\text{Exactitud (Accuracy)} = 0.92$

En la tabla siguiente se muestran los resultados de las métricas locales:

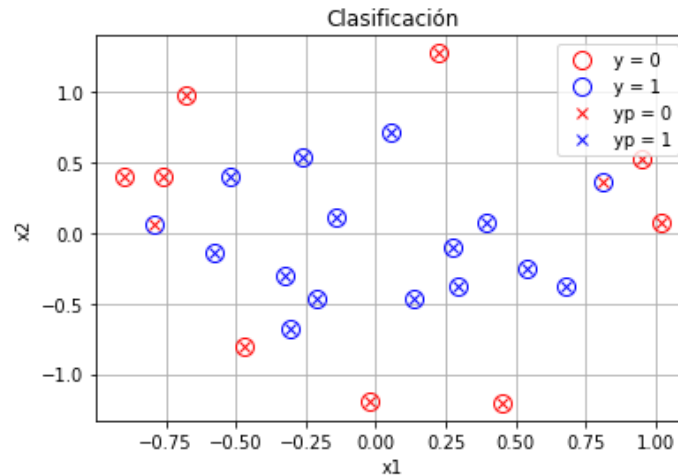
	Precisión	Sensibilidad	Coeficiente F1
0	0.88	1	0.93
1	1	0.82	0.90

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	14	0
	1	2	9



- Para el archivo adjunto “[df\\_clasificacion\\_2.csv](#)” se muestran a continuación los mejores resultados obtenidos para cada modelo de clasificación.
  - Árboles de decisión



Los resultados de las métricas globales son:

$F1_{entrenamiento} = 0.9333$

$F1_{generalización} = 0.92$

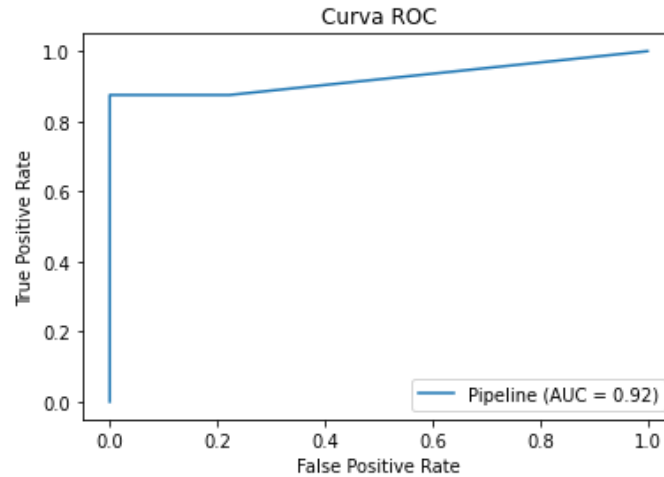
$Exactitud (Accuracy) = 0.92$

En la tabla siguiente se muestran los resultados de las métricas locales:

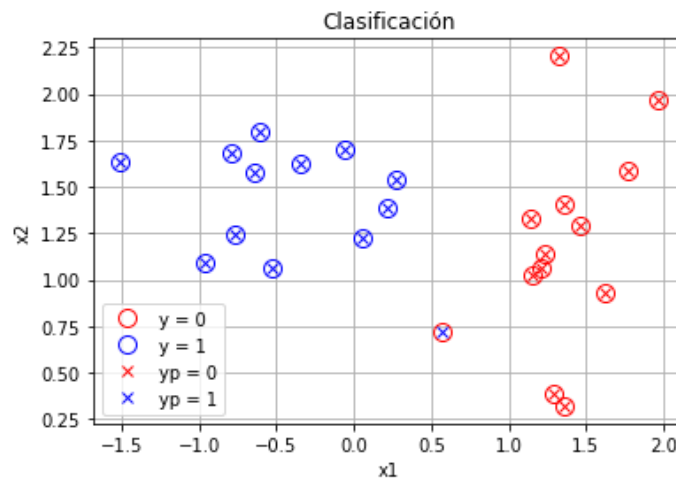
	Precisión	Sensibilidad	Coeficiente F1
0	0.82	1	0.90
1	1	0.88	0.93

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	9	0
	1	2	14



- K vecinos más cercanos



Los resultados de las métricas globales son:

$F1_{entrenamiento} = 0.91$

$F1_{generalización} = 0.92$

$Exactitud (Accuracy) = 0.92$

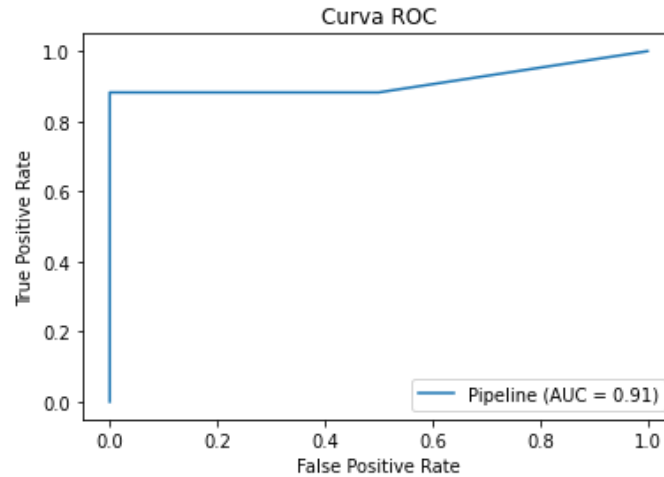
En la tabla siguiente se muestran los resultados de las métricas locales:

	Precisión	Sensibilidad	Coeficiente F1
0	0.81	1	0.89
1	1	0.88	0.94

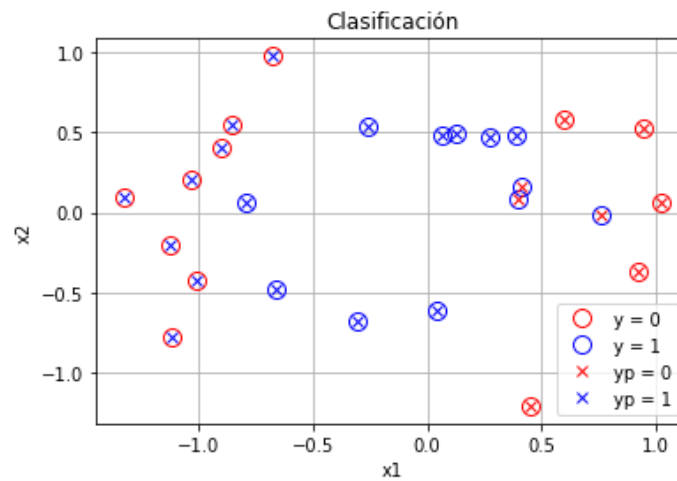
La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	8	0
	1	2	15





- Máquinas de soporte vectorial



Los resultados de las métricas globales son:

$F1_{entrenamiento} = 0.88$

$F1_{generalización} = 0.88$

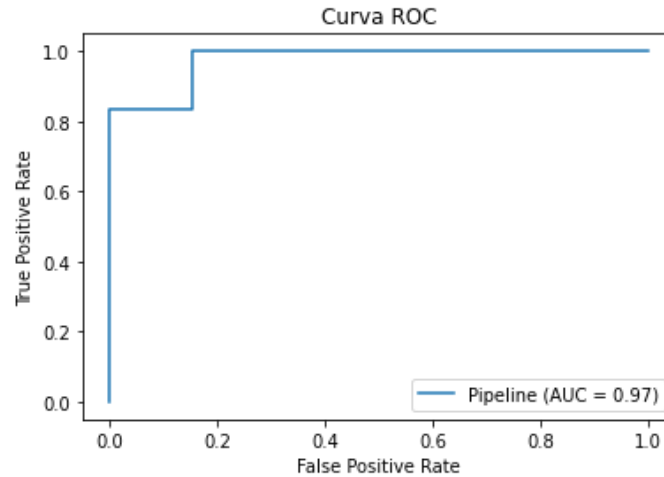
$Exactitud (Accuracy) = 0.88$

En la tabla siguiente se muestran los resultados de las métricas locales:

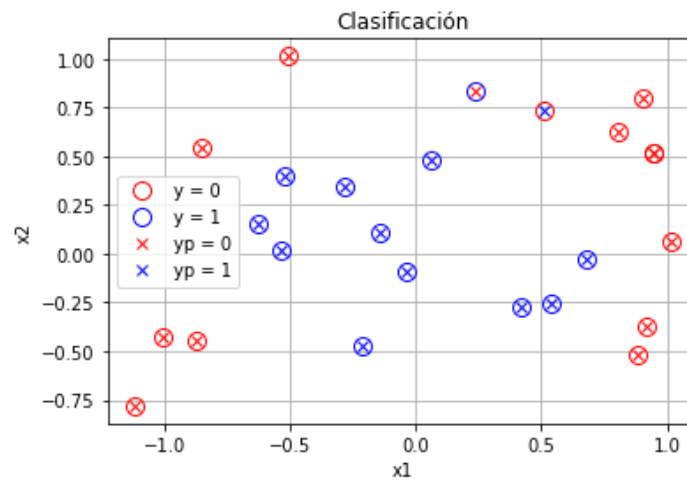
	Precisión	Sensibilidad	Coeficiente F1
0	0.86	0.92	0.89
1	0.91	0.83	0.87

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	12	1
	1	2	10



○ Perceptrón Multicapa



Los resultados de las métricas globales son:

$$F1 \text{ entrenamiento} = 0.92$$

$$F1 \text{ generalización} = 0.92$$

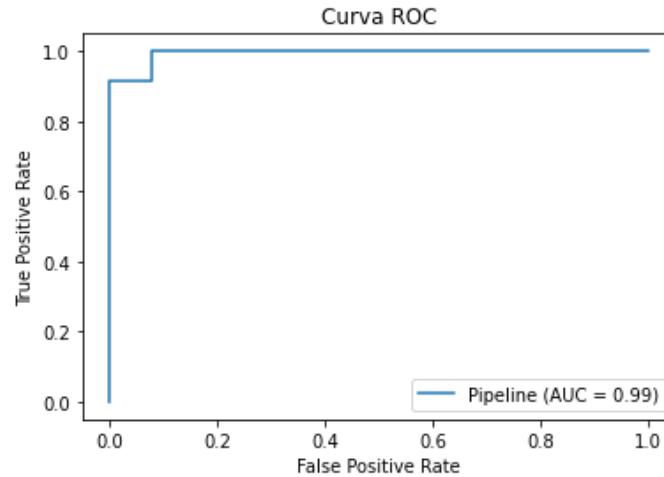
$$\text{Exactitud (Accuracy)} = 0.92$$

En la tabla siguiente se muestran los resultados de las métricas locales:

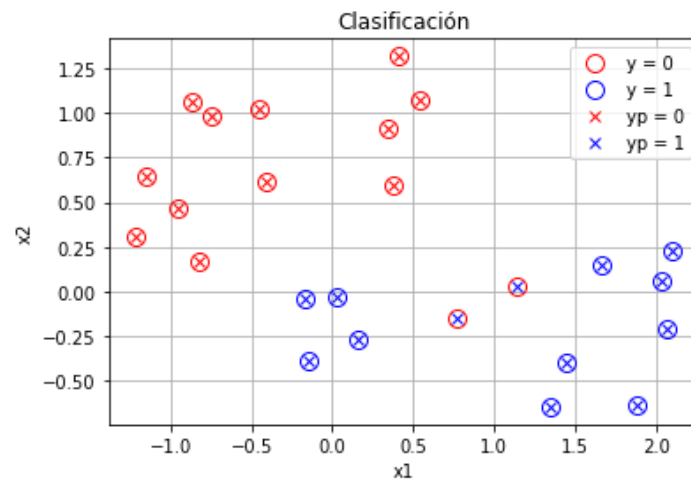
	Precisión	Sensibilidad	Coeficiente F1
0	0.92	0.92	0.92
1	0.92	0.92	0.92

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	12	1
	1	1	11



- Para el archivo adjunto [“df\\_clasificacion\\_3.csv”](#) se muestran a continuación los mejores resultados obtenidos para cada modelo de clasificación.
  - Árboles de decisión



Los resultados de las métricas globales son:

$F1$  entrenamiento = 0.92

$F1$  generalización = 0.92

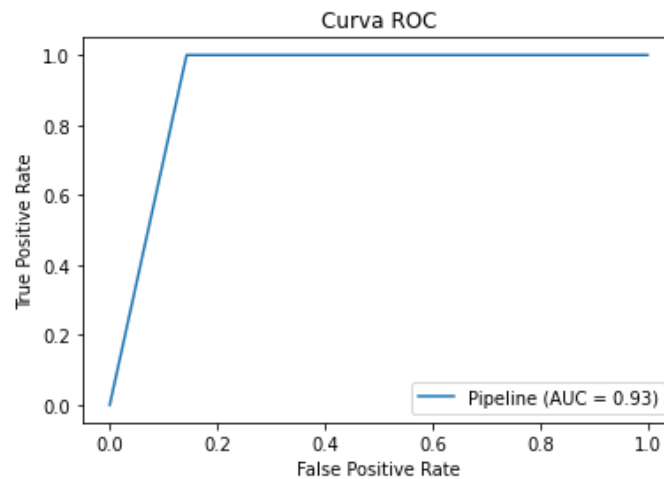
Exactitud (Accuracy) = 0.92

En la tabla siguiente se muestran los resultados de las métricas locales:

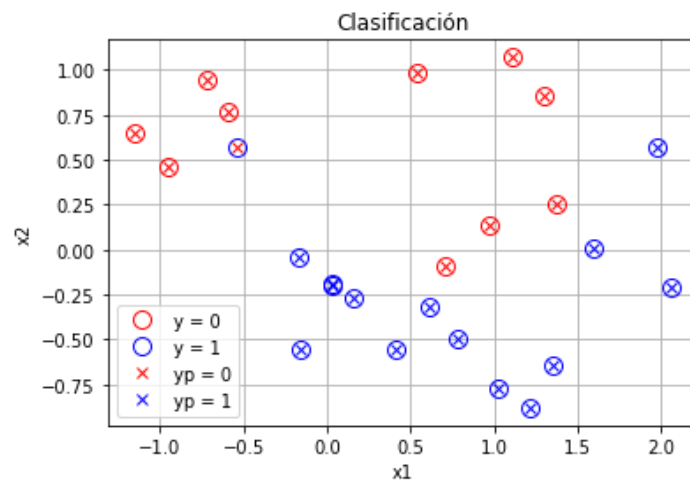
	Precisión	Sensibilidad	Coeficiente F1
0	1	0.86	0.92
1	0.85	1	0.92

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	12	2
	1	0	11



- K vecinos más cercanos



Los resultados de las métricas globales son:

$F1_{\text{entrenamiento}} = 0.96$

$F1_{\text{generalización}} = 0.96$

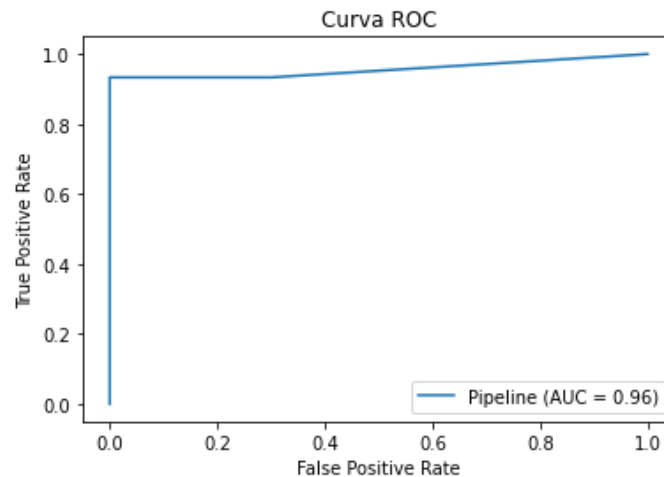
$\text{Exactitud (Accuracy)} = 0.96$

En la tabla siguiente se muestran los resultados de las métricas locales:

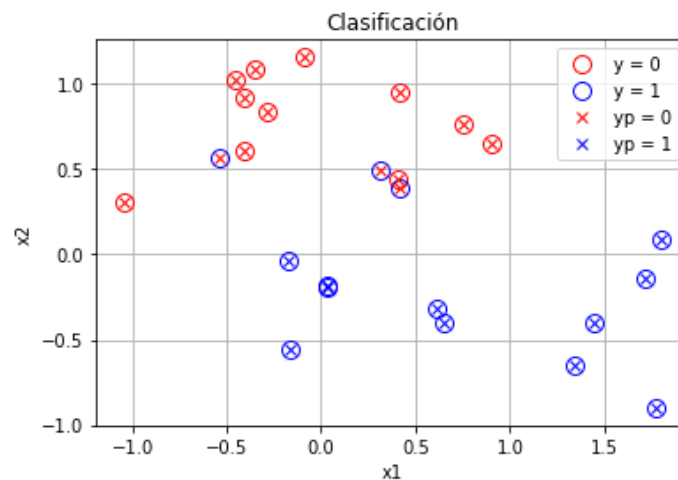
	Precisión	Sensibilidad	Coeficiente F1
0	0.91	1	0.95
1	1	0.93	0.97

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	10	0
	1	1	14



- Máquinas de soporte vectorial



Los resultados de las métricas globales son:

$F1_{\text{entrenamiento}} = 0.88$

$F1_{\text{generalización}} = 0.88$

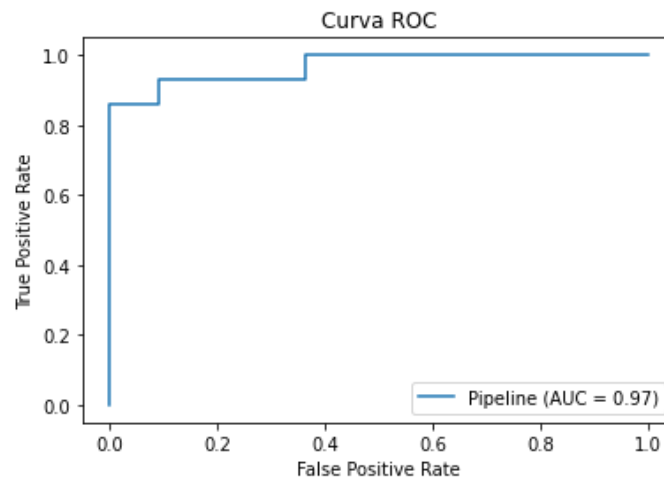
$\text{Exactitud (Accuracy)} = 0.88$

En la tabla siguiente se muestran los resultados de las métricas locales:

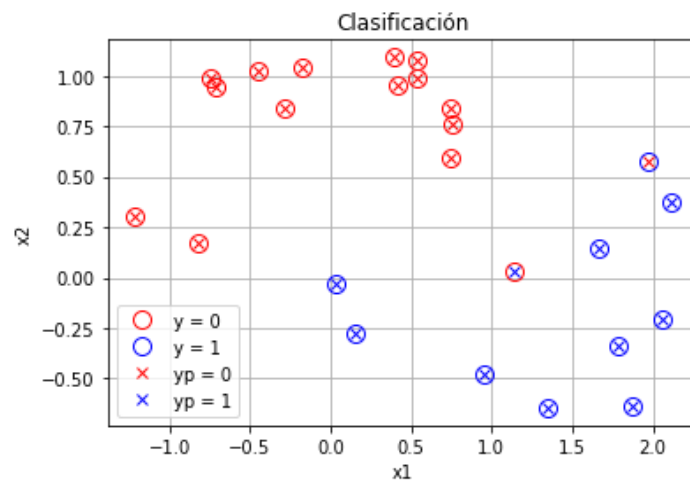
	Precisión	Sensibilidad	Coeficiente F1
0	0.79	1	0.88
1	1	0.79	0.88

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	11	1
	1	3	11



#### ○ Perceptrón Multicapa



Los resultados de las métricas globales son:

$F1_{entrenamiento} = 0.92$

$F1_{generalización} = 0.92$

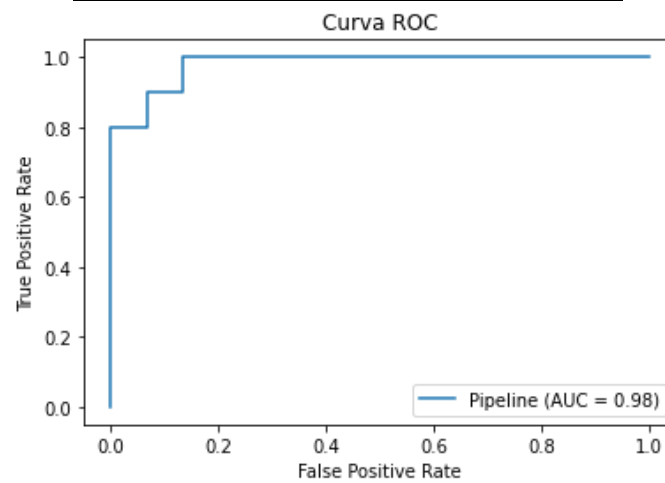
$Exactitud (Accuracy) = 0.92$

En la tabla siguiente se muestran los resultados de las métricas locales:

	Precisión	Sensibilidad	Coeficiente F1
0	0.93	0.93	0.93
1	0.90	0.90	0.90

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	14	1
	1	1	9



## Parte 2

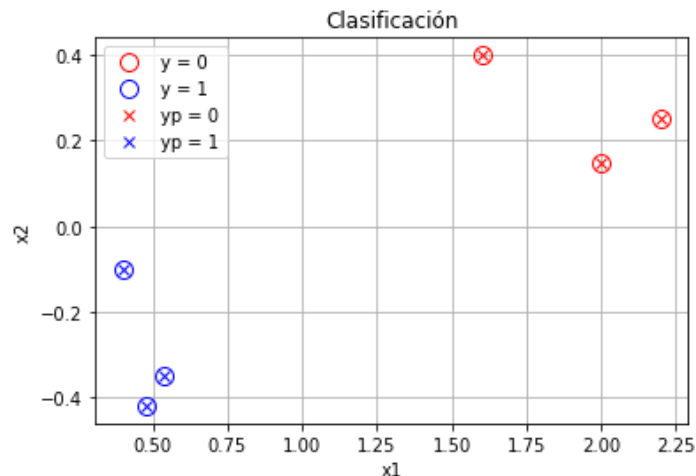
- En esta sección se planteó un problema de clasificación binaria y se resolvió usando las herramientas vistas en la parte 1.

En un terrario se tienen 2 especies de la familia de insectos Formicidae (Hormigas), con sensores se toman datos a cada individuo y se obtienen pares de datos, uno referente al color del individuo y otro referente al tipo de feromona que secreta. Aclarar que cada especie está relacionada a un rango de color y un tipo de feromona. Los datos recabados se muestran en la siguiente tabla:

$x_1$	1.2	1.3	1	1.5	1.2	1.6	1.4	0.1	0.4	0.95	0.8	1.7
$x_2$	0.08	-0.3	-0.08	-0.2	-0.1	0.4	0.2	-0.47	-0.1	-0.13	-0.4	0.08
$\hat{y}$	1	0	0	1	0	0	0	1	1	1	1	0

$x_1$	0.48	2.2	2.4	2	2.3	1.65	0.63	0.9	0.54	0.14	1.5	0.27
$x_2$	-0.42	0.25	-0.3	0.15	0.5	0.33	-0.37	-0.32	-0.35	-0.21	-0.45	-0.17
$\hat{y}$	1	0	0	0	0	0	1	1	1	1	1	1

- Arboles de decisión



Los resultados de las métricas globales son:

$$F1 \text{ entrenamiento} = 0.84$$

$$F1 \text{ generalización} = 0.8333$$

$$\text{Exactitud (Accuracy)} = 0.83$$

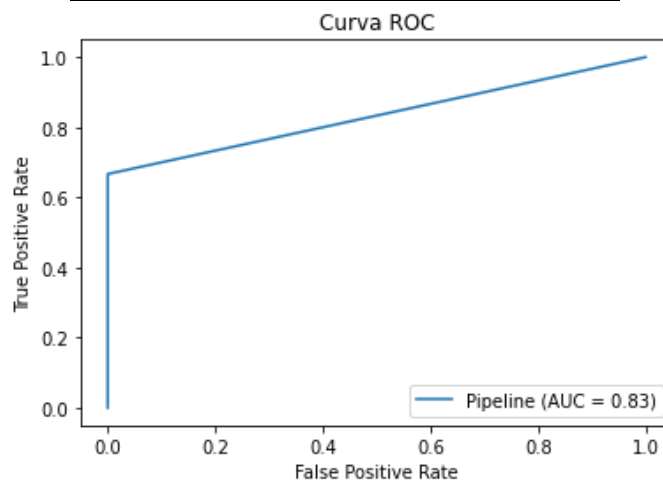
En la tabla siguiente se muestran los resultados de las métricas locales:

	Precisión	Sensibilidad	Coeficiente F1
0	0.75	1	0.86
1	1	0.67	0.80

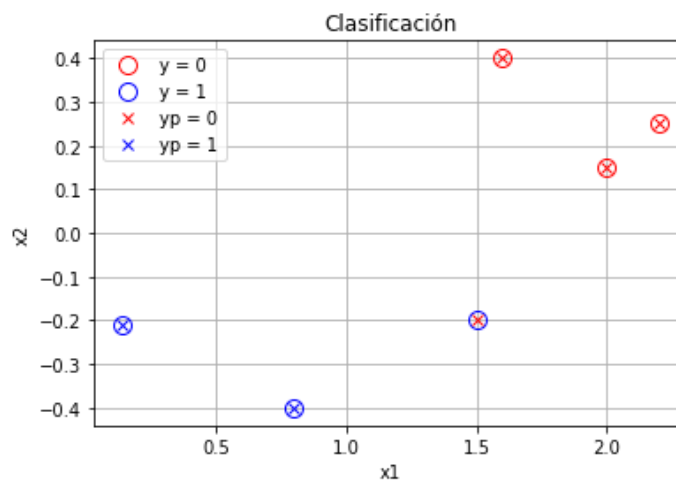


La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	3	0
	1	1	2



- K vecinos más cercanos



Los resultados de las métricas globales son:

$F1_{entrenamiento} = 0.8333$

$F1_{generalización} = 0.8333$

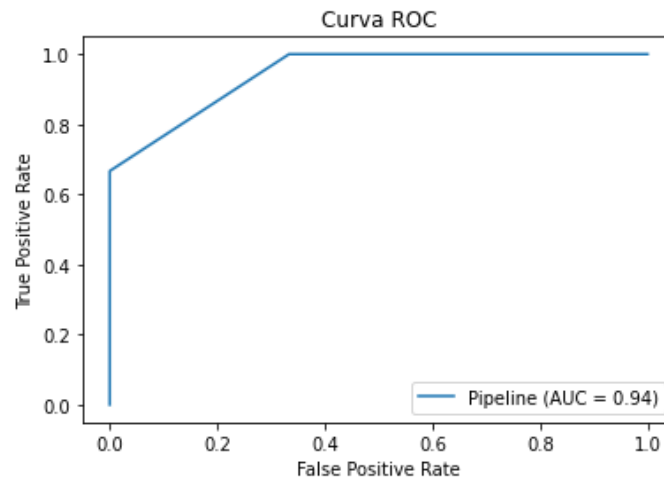
$Exactitud (Accuracy) = 0.83$

En la tabla siguiente se muestran los resultados de las métricas locales:

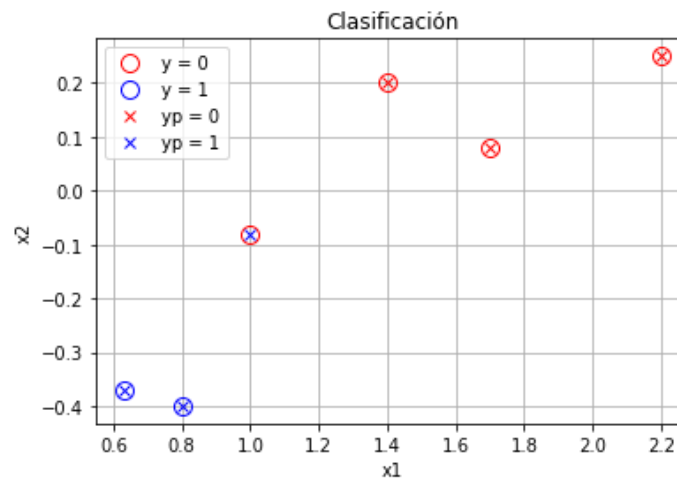
	Precisión	Sensibilidad	Coeficiente F1
0	0.75	1	0.86
1	1	0.67	0.80

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	3	0
	1	1	2



- Máquinas de soporte vectorial



Los resultados de las métricas globales son:

$F1_{entrenamiento} = 0.83$

$F1_{generalización} = 0.83$

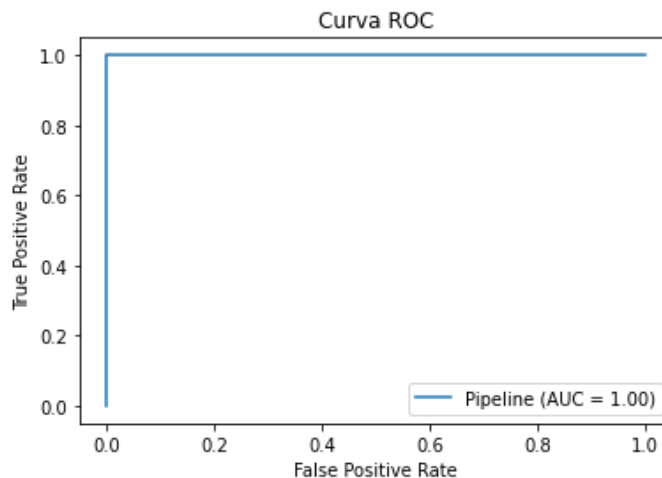
$Exactitud (Accuracy) = 0.83$

En la tabla siguiente se muestran los resultados de las métricas locales:

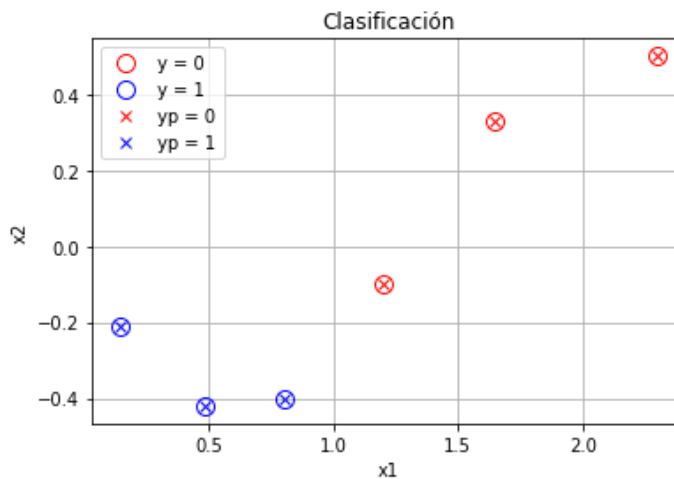
	Precisión	Sensibilidad	Coeficiente F1
0	1	0.75	0.86
1	0.67	1	0.80

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	3	1
	1	0	2



#### ○ Perceptrón Multicapa



Los resultados de las métricas globales son:

$$F1_{\text{entrenamiento}} = 1$$

$$F1_{\text{generalización}} = 1$$

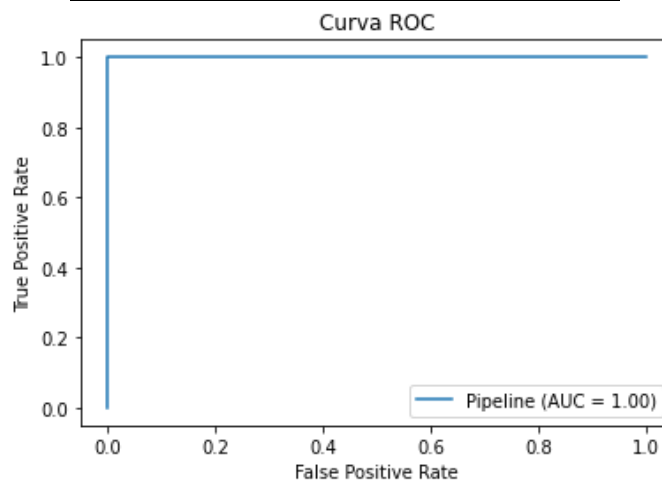
$$\text{Exactitud (Accuracy)} = 1$$

En la tabla siguiente se muestran los resultados de las métricas locales:

	Precisión	Sensibilidad	Coeficiente F1
0	1	1	1
1	1	1	1

La matriz de confusión y curva ROC generadas se muestran enseguida:

		Predicciones	
		0	1
Valores verdaderos	0	3	0
	1	0	3



## Conclusión

Al realizar la actividad se notó que el modelo de regresión de árboles de decisiones y K vecinos más cercanos fueron los más difícil de configurar en cuanto a sus parámetros con el fin de obtener una relativa buena generalización. Mientras que los modelo que menos ajustes necesitaron fueron el modelo de máquinas de soporte vectorial (SVR) y el perceptrón multicapa (MLP). En la parte 2 de la actividad se usó un ejemplo con menor ruido, por lo que fueron más fácil de configurar todos los modelos, y se obtuvieron mejores resultados.