

SISTEMAS INTELIGENTES

Lecture 04: Algoritmos de Clasificación

Dr. Edwin Valencia Castillo
Departamento de Sistemas
Facultad de Ingeniería
Universidad Nacional de Cajamarca
2024

Algoritmos de Clasificación

En este tema veremos los algoritmos de clasificación, los cuales son un tipo de algoritmos de aprendizaje supervisado donde la variable a predecir es categórica.

Objetivo: obtener la clase más probable para cada una de sus instancias.

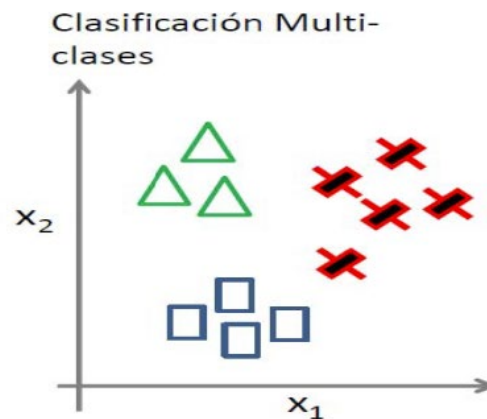
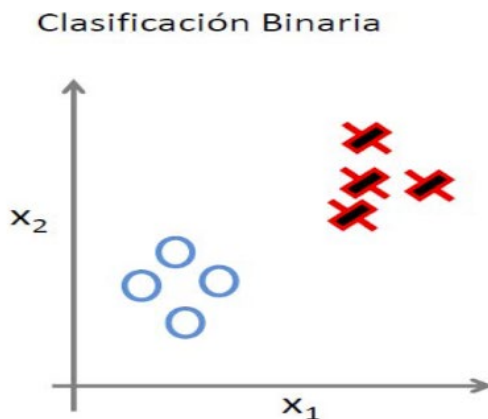
La variable a predecir es categórica

Se pueden utilizar para predecir o estimar la probabilidad de pertenencia de una clase de entre 2 posibles (Clasificación binaria) ó

Predecir o estimar la probabilidad de una clase de entre varias posibles (Clasificación multi-clase)

Problema de Clasificación

- ✓ Partiendo de unos **datos de entrenamiento**, con los valores de las variables X_1, \dots, X_p (**características**) y el de la variable categórica Y (**respuesta**) correspondiente a cada observación, se pretende separar los datos según los valores de Y (de modo que se pueda determinar el valor de Y a partir de los valores de X_1, \dots, X_p)
- ✓ Cuando la variable respuesta Y solo toma dos valores distintos, se dice clasificación binaria.
- ✓ **Objetivo:** Encontrar un **modelo de clasificación**, es decir una función F tal que $F(X)=Y$. Para ello se intenta separar los datos en regiones mediante una **frontera de decisión** (si existen hiperplanos que separan los datos, según los valores de Y , se dice que hay frontera lineal).



Clasificador: Aproximación sistemática a un modelo, construido usando un Algoritmo de aprendizaje

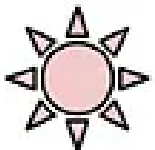


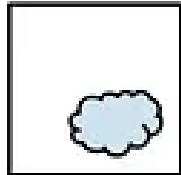
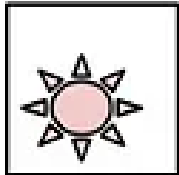
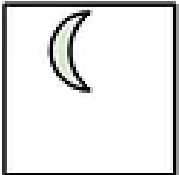
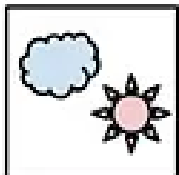
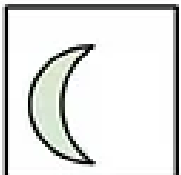
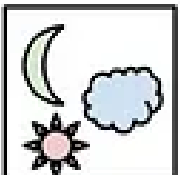
Ejemplos de clasificación binaria

- ✓ **Diagnóstico médico** : Identificar si un paciente tiene una enfermedad o no basándose en los resultados de pruebas diagnósticas.
- ✓ **Calificación crediticia** : determinar si es probable que un solicitante de crédito incumpla con el pago de un préstamo.
- ✓ **Marketing** : Predecir la pérdida de clientes o la probabilidad de que un cliente compre un producto.
- ✓ **Detección de spam** : clasifica los correos electrónicos como spam o no spam.
- ✓ **Control de calidad** : evaluar la probabilidad de que un producto pase los estándares de calidad.

Ejemplos de clasificación multiclase

- ✓ Un algoritmo que busque predecir en qué país de doce posibles un cliente va a realizar la siguiente reserva de un vuelo.
- ✓ La clasificación mediante características extraídas de un conjunto de imágenes de fruta, donde cada imagen puede ser de una naranja, una manzana o una pera.

Diferencia entre multiclase y multietiqueta

Multi-Class		Multi-Label	
C = 3	Samples	Samples	
  	  	  	
	Labels (t)	Labels (t)	
	$[0\ 0\ 1]$ $[1\ 0\ 0]$ $[0\ 1\ 0]$	$[1\ 0\ 1]$ $[0\ 1\ 0]$ $[1\ 1\ 1]$	

Fuente: <https://medium.com/analytics-vidhya/multi-label-classification-using-fastai-a-shallow-dive-into-fastai-data-block-api-54ea57b2c78b>

Evaluando el rendimiento

- ✓ **Cualquier evaluación corre el riesgo de ser subjetiva.**
- ✓ **Es necesario establecer los métodos de evaluación más objetivos en cualquier escenario.**
- ✓ **Los algoritmos de los diferentes modelos pueden tener fortalezas y debilidades y es necesario disponer de la posibilidad de evaluarlos en las mismas condiciones**

Métricas de evaluación

La mejor métrica de rendimiento de un algoritmo de clasificación es ver si el clasificador tiene éxito para su propósito.

Para evaluar un clasificador se pueden utilizar los valores predichos de las clases, los valores reales de las clases o bien la probabilidad estimada de la predicción.

En la práctica lo habitual es mantener dos vectores de datos. Por un lado, un vector que contiene la verdad o los valores observados y otro vector que contiene los valores predichos o estimados. Es importante resaltar que ambos vectores deben tener el mismo tamaño y los datos en el mismo orden.

En este tipo de algoritmos de aprendizaje supervisado la clase verdadera se conoce de antemano y esta variable se utiliza para evaluar el algoritmo en el conjunto de test comparándola con el resultado predicho.

Es decir, los valores predichos se obtienen por medio de la aplicación de un modelo sobre el conjunto de test, el cual ha sido entrenado previamente con datos de entrenamiento.

Métricas de evaluación

- ✓ Por ejemplo, un mensaje de correo electrónico puede ser predicho como spam con una probabilidad de 0,9 o de 0,59, etc.
- ✓ De forma general si dos algoritmos de clasificación diferentes producen los mismos errores, pero uno de ellos es más capaz de tener en cuenta la incertidumbre, sería un mejor modelo.
- ✓ Por tanto, para evaluar correctamente el resultado de un clasificador es necesario considerar el valor de probabilidad obtenido en lugar de utilizar únicamente la clase más probable.

En resumen, el objetivo es obtener modelos que tienen mucha confianza en las predicciones correctas y sean temerosos en las predicciones dudosas. Este **balance entre confianza y prudencia** es la clave de la evaluación de modelos.

¡¡¡¡Veamos un caso!!!!

- ✓ Consideremos una tarea para clasificar si una persona está embarazada o no . Si la prueba de embarazo es positiva (+ve), entonces la persona está embarazada. Por otro lado, si la prueba de embarazo es negativa (-ve), entonces la persona no está embarazada.
- ✓ Ahora, consideremos la clasificación anterior (embarazada o no embarazada) realizada por un algoritmo de aprendizaje automático.
- ✓ El resultado del algoritmo de aprendizaje automático se puede asignar a una de las siguientes categorías.

¡¡¡¡Veamos un caso!!!!

Una persona que en realidad no está embarazada (negativa) y que se clasifica como no embarazada (negativa). Esto se llama VERDADERO NEGATIVO (TN)



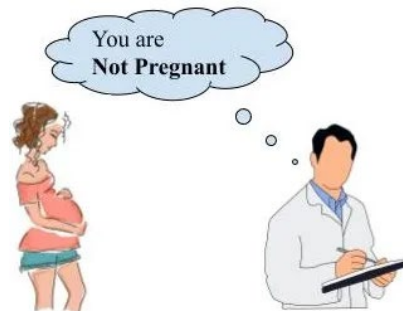
TRUE NEGATIVE

Una persona que en realidad no está embarazada (negativa) y que se clasifica como embarazada (positiva). Esto se llama FALSO POSITIVO (FP)



FALSE POSITIVE

Una persona que está realmente embarazada (positiva) y se clasifica como no embarazada (negativa). Esto se llama FALSO NEGATIVO (FN)



FALSE NEGATIVE

Una persona que está realmente embarazada (positiva) y clasificada como embarazada (positiva). Esto se llama VERDADERO POSITIVO (TP)



TRUE POSITIVE

¡¡¡¡Veamos un caso!!!!

- ✓ Lo que deseamos es VERDADERO POSITIVO y VERDADERO NEGATIVO , pero debido a las clasificaciones erróneas, también podemos terminar en FALSO POSITIVO y FALSO NEGATIVO. Por lo tanto, existe una confusión al clasificar si una persona está embarazada o no.
- ✓ Esto se debe a que ningún algoritmo de aprendizaje automático es perfecto.
- ✓ Pronto describiremos esta confusión al clasificar los datos en una matriz llamada matriz de confusión.

Métrica de evaluación: Matriz de confusión

- ✓ Ahora volveremos al ejemplo anterior de clasificar a 100 personas (que incluye 40 mujeres embarazadas y los 60 restantes son mujeres no embarazadas y hombres con barriga abultada) como embarazadas o no embarazadas.
- ✓ De 40 mujeres embarazadas, 30 mujeres embarazadas se clasifican correctamente y las 10 mujeres embarazadas restantes se clasifican como no embarazadas por el algoritmo de aprendizaje automático.
- ✓ Por otro lado, de 60 personas en la categoría de no embarazadas, 55 se clasifican como no embarazadas y los 5 restantes se clasifican como embarazadas.

Métrica de evaluación: Matriz de confusión

✓ En este caso, $TN = 55$, $FP = 5$, $FN = 10$, $TP = 30$. La matriz de confusión es la siguiente.

		PREDICTED LABEL	
		NEGATIVE	POSITIVE
TRUE LABEL	NEGATIVE	55 TRUE NEGATIVE	5 FALSE POSITIVE
	POSITIVE	10 FALSE NEGATIVE	30 TRUE POSITIVE

¿Cuál es la exactitud del modelo de aprendizaje automático para esta tarea de clasificación?

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

- ✓ La exactitud representa el número de instancias de datos clasificadas correctamente sobre el número total de instancias de datos.
 - En este ejemplo, $Accuracy = (55 + 30)/(55 + 5 + 30 + 10) = 0,85$ y en porcentaje la exactitud será del 85%.
- ✓ ¿Es la exactitud la mejor medida?
 - La exactitud puede no ser una buena medida si el conjunto de datos no está equilibrado (tanto las clases negativas como las positivas tienen una cantidad diferente de instancias de datos).
 - Explicaremos esto con un ejemplo.

¿Es la exactitud la mejor medida?

Consideremos el siguiente escenario: hay 90 personas sanas (negativas) y 10 personas que padecen alguna enfermedad (positivas). Ahora digamos que nuestro modelo de aprendizaje automático clasificó perfectamente a las 90 personas como sanas, pero también clasificó a las personas no sanas como sanas. ¿Qué sucederá en este escenario? Veamos la matriz de confusión y descubramos la precisión.

		PREDICTED LABEL	
		NEGATIVE	POSITIVE
TRUE LABEL	NEGATIVE	90 TRUE NEGATIVE	0 FALSE POSITIVE
	POSITIVE	10 FALSE NEGATIVE	0 TRUE POSITIVE

La exactitud en este caso será $(90 + 0)/(100) = 0,9$ y en porcentaje la exactitud es del 90 %.

¿Hay algo sospechoso?

La exactitud, en este caso, es del 90 %, pero este modelo es muy deficiente porque las 10 personas que no están sanas se clasifican como sanas. Con este ejemplo, lo que estamos tratando de decir es que la exactitud no es una buena métrica cuando el conjunto de datos no está equilibrado. El uso de la exactitud en tales escenarios puede dar lugar a una interpretación errónea de los resultados.

Precisión (valor predictivo positivo)

$$Precision = \frac{TP}{TP + FP}$$

✓ ¿Qué significa precisión?

- Lo ideal sería que la precisión fuera 1 (alta) para que un clasificador fuera bueno. La precisión se convierte en 1 solo cuando el numerador y el denominador son iguales, es decir, $TP = TP + FP$, lo que también significa que FP es cero. A medida que FP aumenta, el valor del denominador se vuelve mayor que el numerador y el valor de la precisión disminuye (lo cual no queremos).
- Entonces, en el ejemplo del embarazo, $precision = 30/(30+ 5) = 0,857$

Recall – sensibilidad o Tasa de verdaderos positivos

$$\text{Recall} = \frac{TP}{TP + FN}$$

- ✓ Lo ideal sería que el valor de Recall fuera 1 (alto) para que el clasificador fuera bueno. El valor de Recall se convierte en 1 solo cuando el numerador y el denominador son iguales, es decir, $TP = TP + FN$, lo que también significa que FN es cero. A medida que FN aumenta, el valor del denominador se vuelve mayor que el numerador y el valor de recuperación disminuye (lo cual no queremos).
- ✓ Entonces, en el ejemplo del embarazo, veamos cuál será Recall.
 - $\text{Recall} = 30 / (30 + 10) = 0.75$

F1 Score

$$F1 \text{ Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- ✓ Lo ideal es que en un buen clasificador tanto la precisión como recall sean iguales, lo que también significa que FP y FN son cero. Por lo tanto, necesitamos una métrica que tenga en cuenta tanto la precisión como recall. F1 Score es una métrica que tiene en cuenta tanto la precisión como recall.
- ✓ F1 Score se convierte en 1 solo cuando la precisión y recall son ambos 1. F1 Score se vuelve alto solo cuando tanto la precisión como recall son altos. F1 Score es la media armónica de la precisión y recall y es una mejor medida que la exactitud .
 - En el ejemplo del embarazo, $F1 \text{ Score} = 2 * (0,857 * 0,75) / (0,857 + 0,75) = 0,799$.

Métrica de evaluación: Matriz de confusión – Resumen

Una persona que en realidad no está embarazada (negativa) y que se clasifica como no embarazada (negativa). Esto se llama VERDADERO NEGATIVO (TN)

Se trata de las clasificaciones correctas de las instancias que corresponden a la clase negativa



TRUE NEGATIVE

Una persona que en realidad no está embarazada (negativa) y que se clasifica como embarazada (positiva). Esto se llama FALSO POSITIVO (FP)

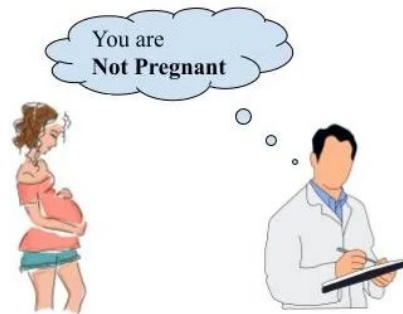
Se trata de las clasificaciones de la clase negativa que han sido incorrectamente clasificadas como clase positiva



FALSE POSITIVE

Una persona que está realmente embarazada (positiva) y se clasifica como no embarazada (negativa). Esto se llama FALSO NEGATIVO (FN)

Se trata de las clasificaciones de la clase positiva que han sido incorrectamente clasificadas como clase negativa



FALSE NEGATIVE













Una persona que está realmente embarazada (positiva) y clasificada como embarazada (positiva). Esto se llama VERDADERO POSITIVO (TP)

Se trata de las clasificaciones correctas de las instancias que corresponden a la clase positiva

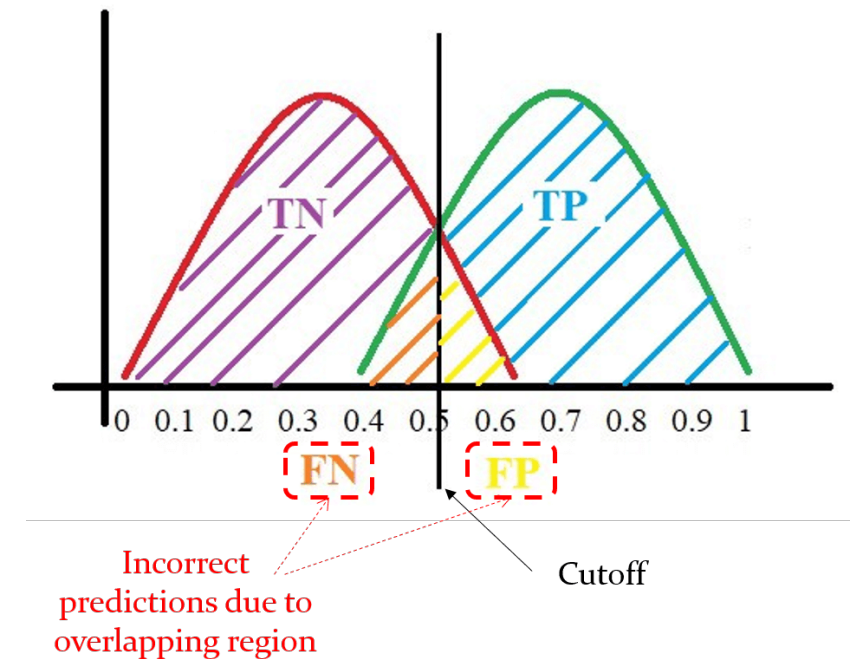


TRUE POSITIVE

Métrica de evaluación: Matriz de confusión – Resumen

		Two Classes				Three Classes		
		Predicted Class				Predicted Class		
		A	B			A	B	C
Actual Class	A			Actual Class	A			
	B				B			
						C		

Las instancias clasificadas correctamente caen en la diagonal de la matriz. Los valores fuera de la diagonal indican las instancias clasificadas incorrectamente, se trata de predicciones incorrectas. Las métricas de rendimiento se basan en la cantidad de instancias que caen dentro de la diagonal y fuera.



Métrica de evaluación: curvas ROC, AUC

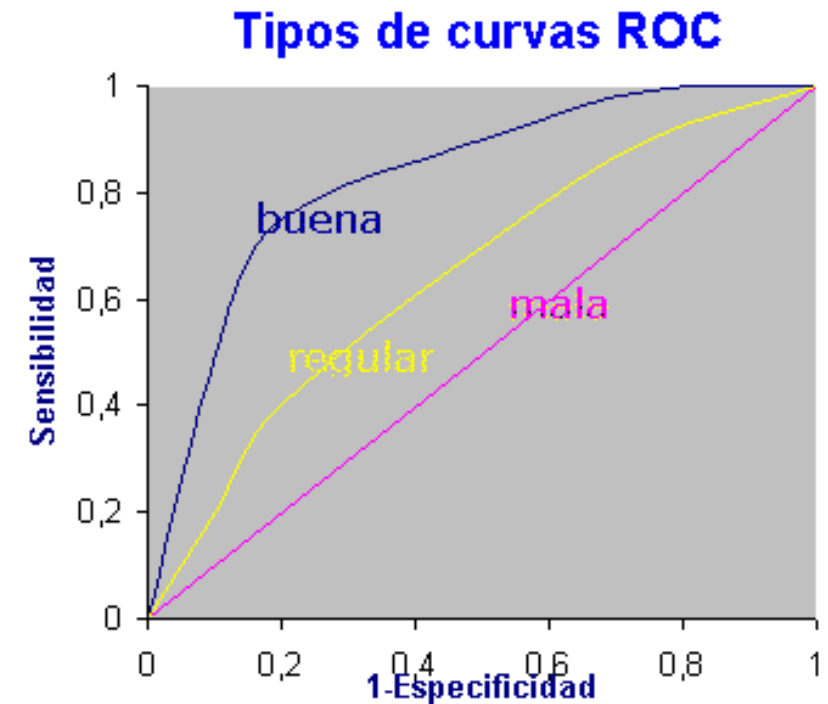
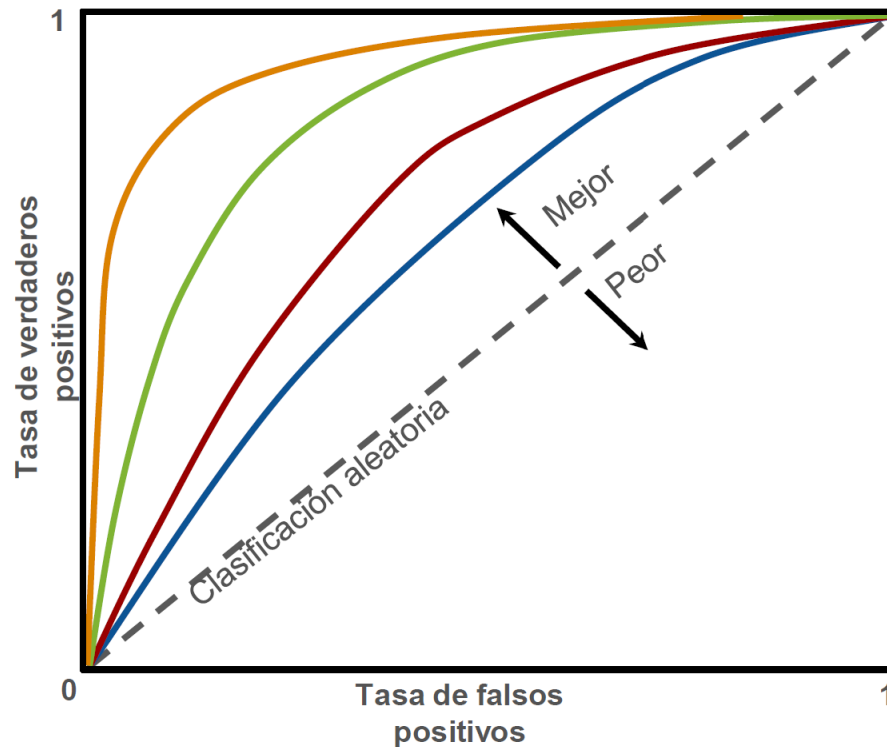
- ✓ **ROC: Receiver Operating Characteristic o Característica Operativa del Receptor**, es una gráfica utilizada para evaluar el rendimiento de un modelo de clasificación binaria. La curva ROC muestra la relación entre la Tasa de Verdaderos Positivos (True Positive Rate, TPR) y la Tasa de Falsos Positivos (False Positive Rate, FPR) a diferentes umbrales de decisión.
- ✓ **Aquí hay un desglose más detallado:**
 - **True Positive Rate (TPR) o Sensibilidad:** Es la proporción de verdaderos positivos (casos correctamente identificados como positivos) entre todos los casos positivos reales. Se calcula como $TPR = TP / (TP + FN)$.
 - **False Positive Rate (FPR) o Especificidad:** Es la proporción de falsos positivos (casos incorrectamente identificados como positivos) entre todos los casos negativos reales. Se calcula como $FPR = FP / (FP + TN)$.
 - **La curva ROC:** Se construye trazando la TPR en el eje Y y la FPR en el eje X para diferentes umbrales de decisión. Un modelo ideal se representaría con un punto en la esquina superior izquierda de la gráfica ($TPR = 1$ y $FPR = 0$), indicando una clasificación perfecta sin falsos positivos ni falsos negativos.

Métrica de evaluación: curvas ROC, AUC

- ✓ **AUC – Area Under the Curve.** En el contexto de la curva ROC, el AUC mide la capacidad del modelo para distinguir entre clases. Un AUC de 1 indica un modelo perfecto, mientras que un AUC de 0.5 indica un modelo que no tiene capacidad de discriminación mejor que el azar.
- ✓ **La interpretación del AUC en la curva ROC es la siguiente:**
 - **AUC = 0.5:** El modelo no tiene capacidad de discriminación; es equivalente a adivinar al azar.
 - **$0.5 < \text{AUC} < 0.7$:** El modelo tiene una capacidad de discriminación baja.
 - **$0.7 \leq \text{AUC} < 0.9$:** El modelo tiene una capacidad de discriminación moderada.
 - **$0.9 \leq \text{AUC} < 1$:** El modelo tiene una capacidad de discriminación alta.
 - **AUC = 1:** El modelo tiene una capacidad de discriminación perfecta.
- ✓ **El AUC es una métrica popular porque es independiente del umbral de decisión y proporciona una medida agregada del rendimiento del modelo a través de todos los posibles umbrales de clasificación.**

La curva ROC

- ✓ El principal objetivo de la curva ROC es determinar qué modelo es mejor dado un conjunto de modelos que sirven para una clasificación.

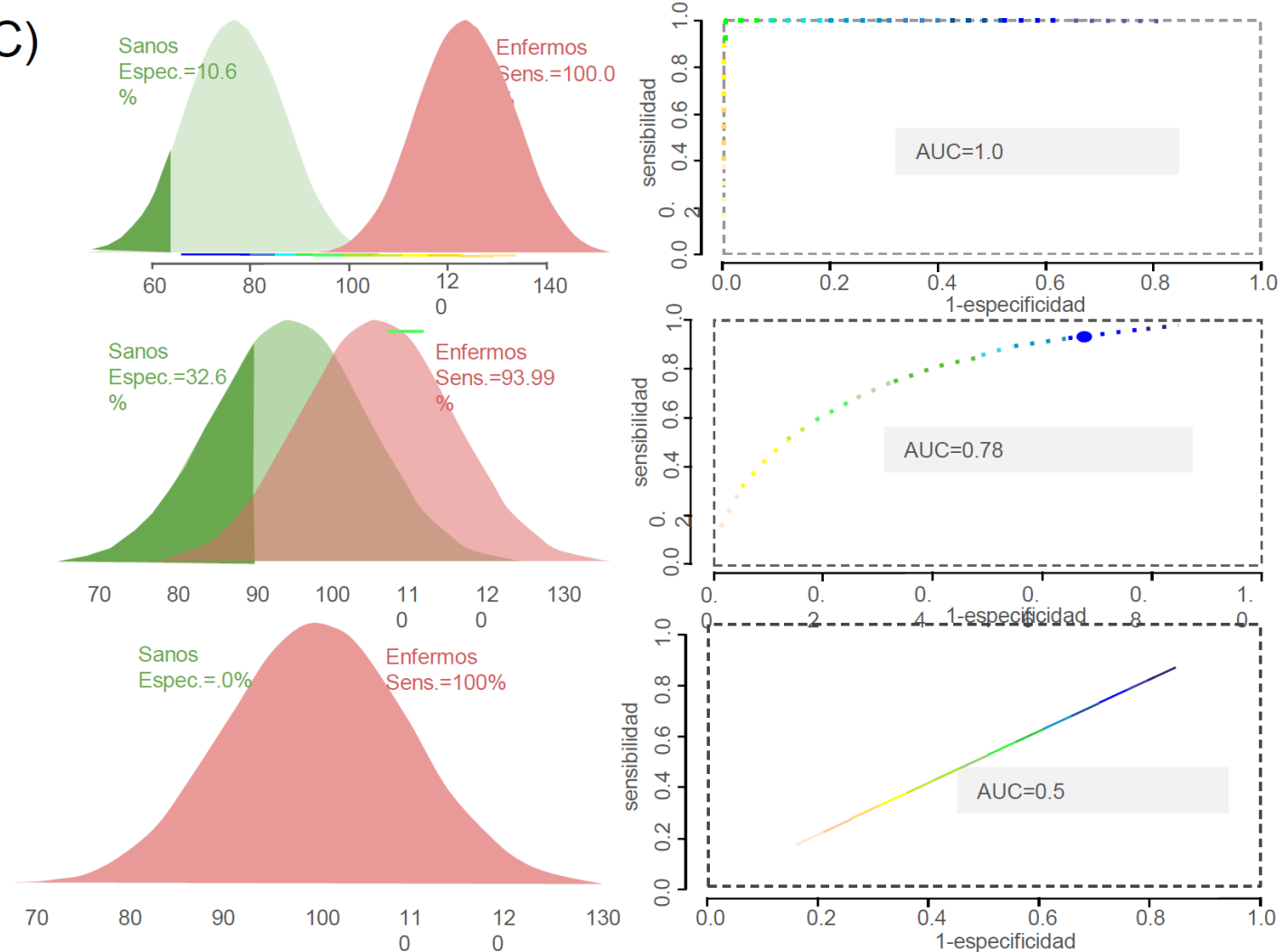


<https://www.newtechdojo.com/auc-roc-curve-visually-explained/>

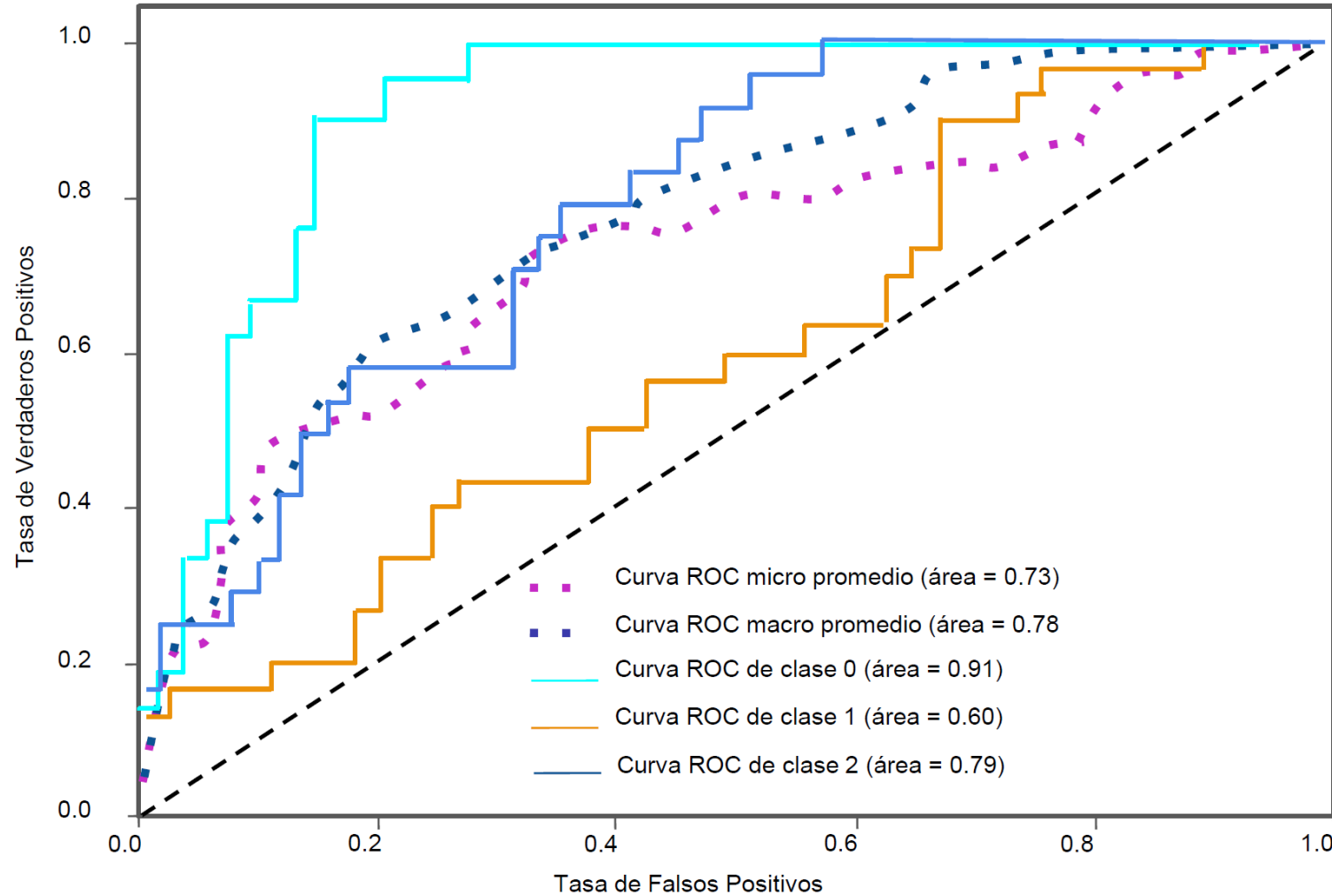
La curva ROC

Área bajo la curva (AUC)

- 0.97-1: Excelente
- 0.9-0.97: Muy bueno
- 0.75-0.9: Bueno
- 0.6-0.75: Regular
- 0.5-0.6: Malo



Curva ROC Multiclase



https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

EJERCICIO: Evaluación de un modelo de clasificación con un filtro de spam

- ✓ Supongamos que queremos clasificar nuestros correos en spam (correo no deseado) y no spam (correo deseado).
- ✓ La matriz de confusión calculada es la siguiente:

	Predicted negative (non-spam)	Predicted positive (spam)
Negative class (non-spam)	264	14
Positive class (spam)	22	158

- ✓ Realice un análisis de los resultados de las métricas obtenidas

PREGUNTAS??

Dr. Edwin Valencia Castillo
Departamento de Sistemas
Facultad de Ingeniería
Universidad Nacional de Cajamarca
2024