
Clase 066 — Curva ROC y AUC

Parte: 1 — Machine Learning Clásico · Fuente: Géron, cap. 3 § The ROC Curve. · Duración estimada: 50 min.

Clase 066 — Curva ROC y AUC

Parte: 1 — Machine Learning Clásico · Fuente: Géron, cap. 3 § The ROC Curve. · Duración estimada: 50 min.

Objetivo

Que el alumno entienda qué mide la curva ROC, calcule el AUC con scikit-learn y sepa decidir cuándo ROC es la métrica adecuada y cuándo conviene usar Precision-Recall — sobre todo en datasets desbalanceados, donde ROC tiende a mentir.

Resultados de aprendizaje

Al finalizar la clase, el alumno podrá:

1. Construir la curva ROC con `roc_curve` graficando TPR vs FPR a distintos umbrales.
2. Calcular el AUC con `roc_auc_score` e interpretar el valor (0.5 = azar, 1.0 = perfecto).
3. Comparar dos clasificadores superponiendo sus curvas ROC y eligiendo por AUC.
4. Decidir ROC vs PR según la prevalencia de la clase positiva.
5. Evitar la trampa del desbalance: reconocer cuándo un AUC alto esconde mala precisión.

Temas

#	Tema	Por qué importa
1	TPR (recall) y FPR	Los dos ejes de ROC; entender qué se gana
2	Curva ROC con <code>roc_curve</code>	Visualiza el trade-off completo, no un pun
3	AUC con <code>roc_auc_score</code>	Resumen escalar e independiente del umbral
4	Diagonal de azar y modelo perfecto	Referencias visuales obligatorias.
5	ROC vs Precision-Recall	En clases desbalanceadas, ROC pinta lindo
6	Comparación de modelos	Cómo elegir entre clasificadores con AUC +

Definiciones y características

TPR (True Positive Rate / Recall / Sensibilidad)

: Proporción de positivos reales correctamente detectados. $TPR = TP / (TP + FN)$. Es el eje Y de ROC.

FPR (False Positive Rate)

: Proporción de negativos reales clasificados erróneamente como positivos. $FPR = FP / (FP + TN) = 1 - \text{especificidad}$. Es el eje X de ROC.

Curva ROC (Receiver Operating Characteristic)

: Gráfico de TPR vs FPR barriendo todos los umbrales de decisión. Cada punto es un umbral. Cuanto más cerca del vértice superior izquierdo, mejor.

AUC (Area Under the Curve)

: Área bajo la curva ROC. Probabilidad de que el modelo asigne mayor score a un positivo aleatorio que a un

negativo aleatorio. 0.5 = azar, 1.0 = perfecto, <0.5 = peor que tirar moneda (invertí las etiquetas).

Diagonal de azar

: Línea $y = x$. Representa un clasificador que adivina al azar. Toda curva ROC útil va por encima de esta diagonal.

Curva Precision-Recall (PR)

: Alternativa a ROC para clases desbalanceadas. Grafica precision vs recall. No incluye TN, por lo que es insensible al inflado por la clase mayoritaria.

ROC vs PR — regla práctica

: Usá ROC cuando las clases están balanceadas o te importan ambas por igual. Usá PR cuando la clase positiva es rara (fraude, enfermedad, clicks) y los falsos positivos en una clase abundante distorsionan el FPR.

Dataset / recursos

- `sklearn.datasets.fetch_openml('mnist_784')` con el clasificador binario "es un 5" (Géron cap. 3), naturalmente desbalanceado (~10% positivos).
- Opcional: dataset sintético desbalanceado vía `make_classification(weights=[0.99, 0.01])` para ver el contraste ROC vs PR en extremo.

Ejercicios

1. Scores y curva ROC. Entrená un `SGDClassifier` sobre "es un 5", obtené scores con `cross_val_predict(..., method='decision_function')` y graficá la curva ROC con `roc_curve`.
2. AUC. Calculá `roc_auc_score(y_true, y_scores)`. Interpretá el valor en una línea.
3. Comparar dos modelos. Entrená un `RandomForestClassifier` (usá `predict_proba`, columna 1) y superponé ambas curvas ROC en un mismo plot. Elegí el mejor por AUC.
4. ROC vs PR en desbalance. Generá `make_classification(weights=[0.99, 0.01], n_samples=10_000)`, entrená un modelo decente y graficá lado a lado curva ROC y curva PR (`precision_recall_curve`). Mostrá cómo ROC sigue "linda" mientras PR revela la pobreza real.
5. Punto operativo. Sobre la curva ROC del ejercicio 1, encontrá el umbral que maximiza TPR - FPR (índice de Youden) y reportá precision y recall en ese punto.

Homework verificable

Notebook con dataset "es un 5" de MNIST: (a) entrenar SGD y RandomForest; (b) graficar ambas ROC superpuestas con AUC en la leyenda; (c) graficar las dos curvas PR con `average_precision_score` en la leyenda; (d) tabla final con AUC y AP por modelo; (e) párrafo de 3-4 líneas eligiendo el ganador y justificando con qué métrica decidiste y por qué (pista: prevalencia ≈ 10%).

Criterio de aceptación: El notebook corre end-to-end. Ambas curvas en un mismo eje. La conclusión menciona explícitamente el desbalance y por qué PR/AP puede ser preferible a ROC/AUC.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
AUC = 0.99 pero la precision real es 0.10	Clase muy desbalanceada — el FPR sigue baj
roc_auc_score lanza ValueError: y should b	Le pasaste predict_proba entero (matriz N×
Curva ROC "escalonada" rara con predict	Estás pasando etiquetas duras (0/1) en lug
AUC < 0.5	El modelo está invertido o pasaste la colu
Comparo modelos solo por AUC y elijo mal	AUC promedia sobre todos los umbrales, inc

Preguntas frecuentes

¿ROC o Precision-Recall?

Si la clase positiva es rara (<10-20%) o te importa específicamente cómo te va con los positivos, usá PR. ROC incluye TN en el denominador del FPR y con muchos TN el FPR queda artificialmente bajo aunque la precision sea malísima. En clases balanceadas o cuando importan ambas clases por igual, ROC está bien.

¿Qué AUC se considera "bueno"?

Depende del dominio. Regla gruesa: 0.5 azar, 0.7 aceptable, 0.8 bueno, 0.9+ excelente, 1.0 sospechá data leakage. En medicina o fraude a veces 0.95 es lo mínimo aceptable.

¿Puedo usar AUC con multiclase?

Sí: `roc_auc_score(..., multi_class='ovr')` (one-vs-rest) o 'ovo' (one-vs-one). Pero para multiclase suele ser más informativo mirar la matriz de confusión y métricas por clase.

¿decision_function o predict_proba?

Da igual para ROC/AUC — ROC depende del orden de los scores, no de su escala. `decision_function` devuelve scores no calibrados (puede ser negativo), `predict_proba` devuelve probabilidades en [0, 1]. Si tu modelo no tiene `predict_proba` (como `SGDClassifier` default), usá `decision_function`.

¿Cómo elijo el umbral final si AUC es independiente del umbral?

AUC sirve para comparar modelos. Para deployar tenés que elegir umbral según el costo de FP vs FN en tu dominio. Métodos: índice de Youden ($\max(\text{TPR} - \text{FPR})$), restricción tipo "recall \geq 0.95 y maximizá precision", o análisis de costo explícito.

Referencias

- Géron, cap. 3 § The ROC Curve y § Precision/Recall Trade-off.
- scikit-learn — `roc_curve`
- scikit-learn — `roc_auc_score`
- scikit-learn — Precision-Recall vs ROC
- Saito & Rehmsmeier (2015), The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets.

Siguiente clase

Clase 067 — Clasificación multiclase, multilabel, multioutput

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb