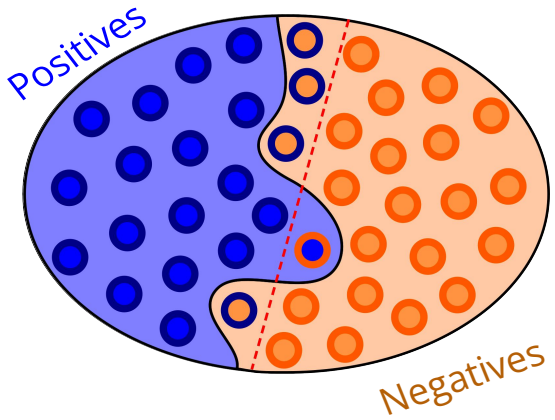
Métricas de desempeño

Performance measurements



42 17 1 20 4

Total samples: $N = \mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}$

Accuracy: $\mathrm{ACC} = \dfrac{(\mathrm{TP} + \mathrm{TN})}{N}$

¿cuánto se acerca a los valores reales? exactitud

Sensitivity/Recall: $\mathrm{TPR} = \dfrac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}$

tasa de enfermos correctamente detectados

Specificity: $\mathrm{TNR} = \dfrac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})}$

tasa de sanos correctamente clasificados

Precision: $\mathrm{PPV} = \dfrac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})}$

de los clasificados positivos ¿cuántos son los realmente positivos?

False positive rate: $\mathrm{FPR} = \dfrac{\mathrm{FP}}{(\mathrm{FP} + \mathrm{TN})}$

probabilidad de una falsa alarma.

False discovery rate: $\mathrm{FDR} = \dfrac{\mathrm{FP}}{(\mathrm{FP} + \mathrm{TP})}$

F-score: $\mathrm{F} = \dfrac{2\mathrm{TP}}{(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN})}$

Media armónica entre TPR y PPV.

TP: true positives
success

TN: true negatives
correct rejection

FP: false positives
type I error (false alarm)

FN: false negatives
type II error

system login

disease detection

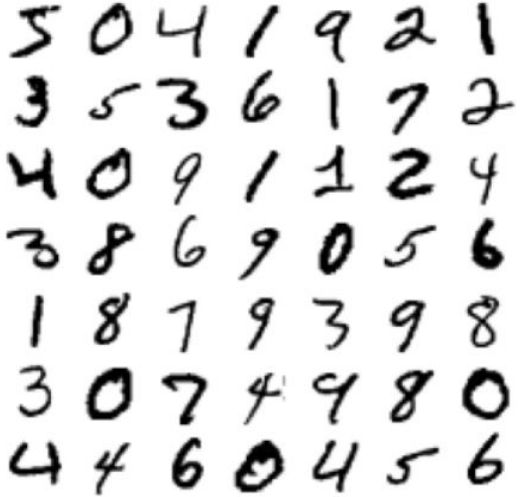
Confusion matrix

Real labels \ Predicted labels	Positive (sick)	Negative (healthy)
Positive (sick)	TP	FN
Negative (healthy)	FP	TN

Ver más métricas en: https://en.wikipedia.org/wiki/Precision_and_recall

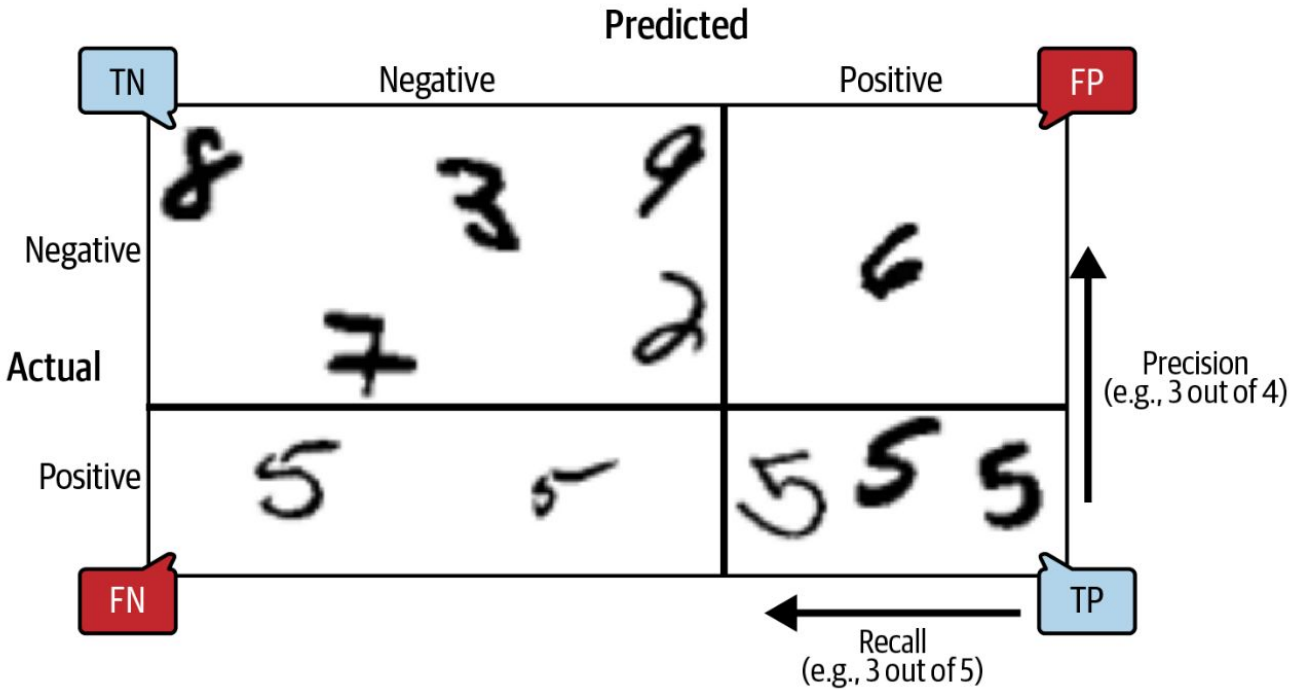
Ejemplo con base MNIST y clasificador lineal

- Base MNIST
  - Author: Yann LeCun, Corinna Cortes, Christopher J.C. Burges
  - Source: [MNIST Website](http://yann.lecun.com/exdb/mnist/)
  - Base de imágenes de dígitos manuscritos
    - 784 características (imágenes de 28x28 píxeles)
  - Accesible en sklearn mediante la función: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_openml.html#sklearn.datasets.fetch_openml
  - 70000 instancias
    - 60000 train
    - 10000 test

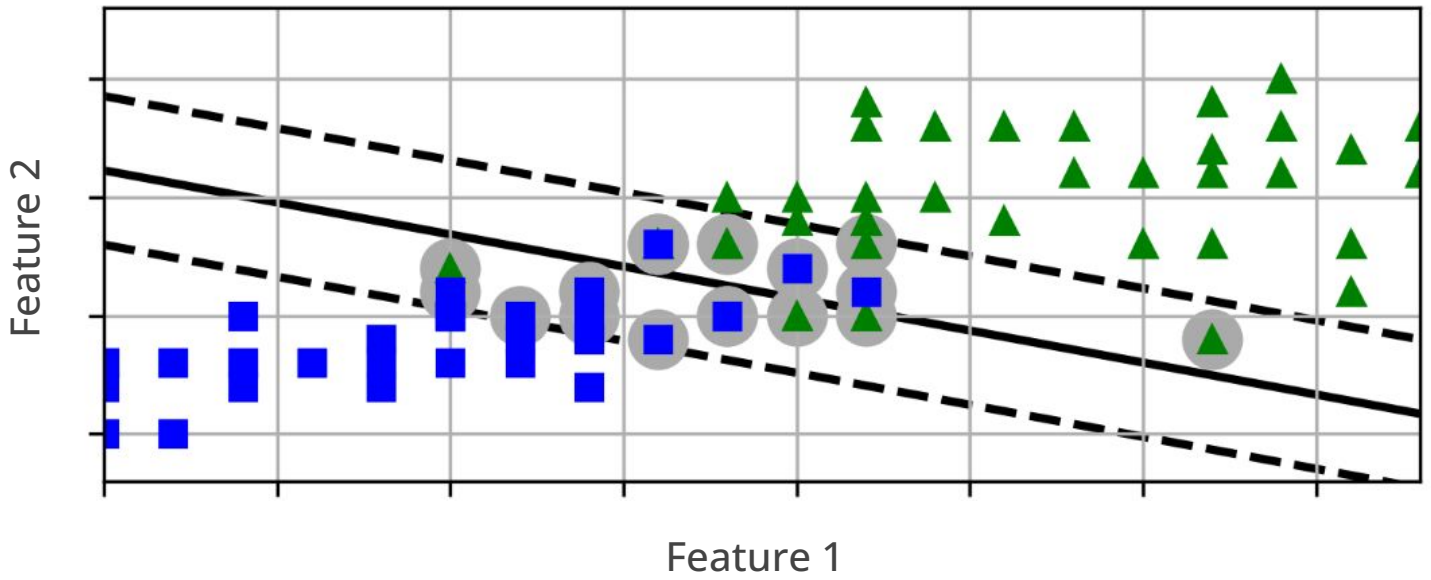


- SGDClassifier
  https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html

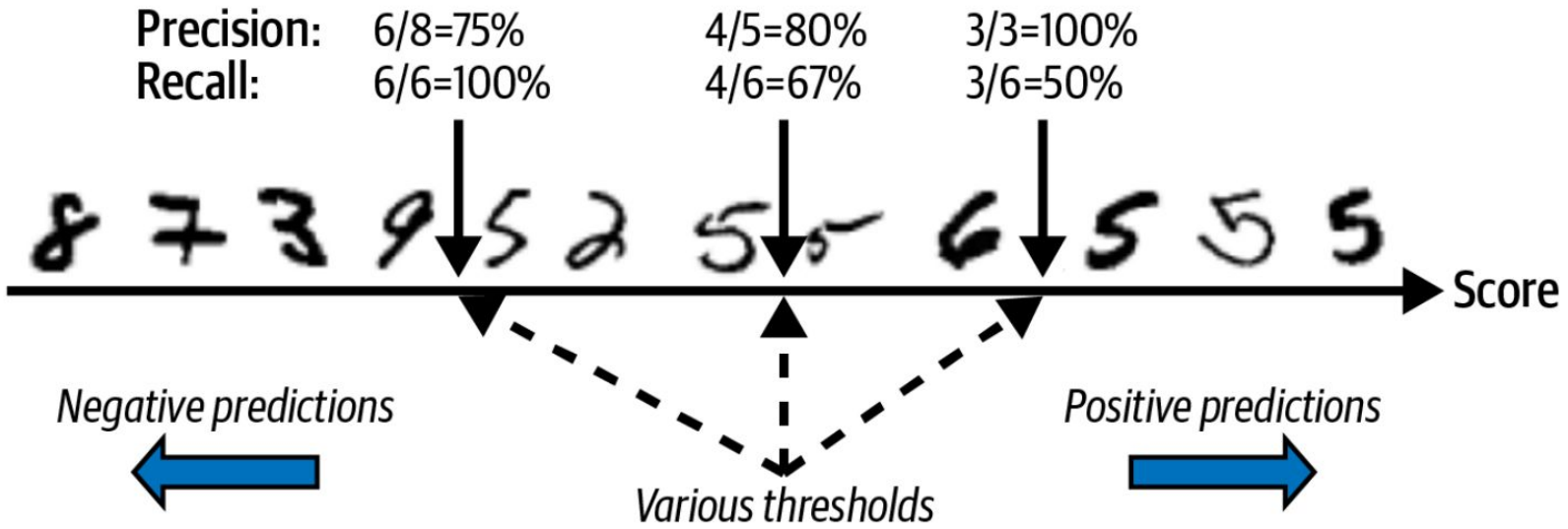
Matriz de confusión



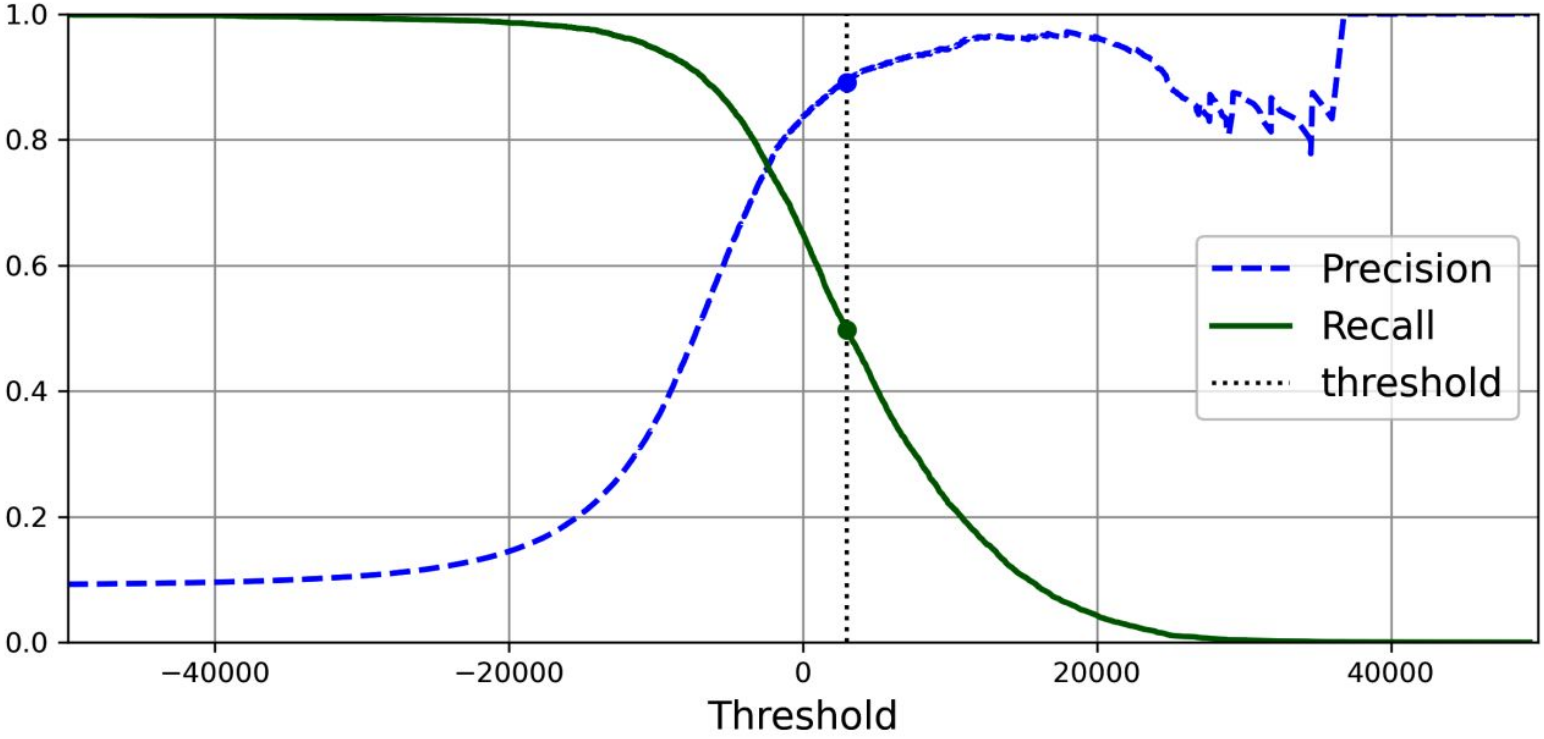
Score de confianza. Distancia al hiperplano



Compromiso precision/recall



Compromiso precision/recall

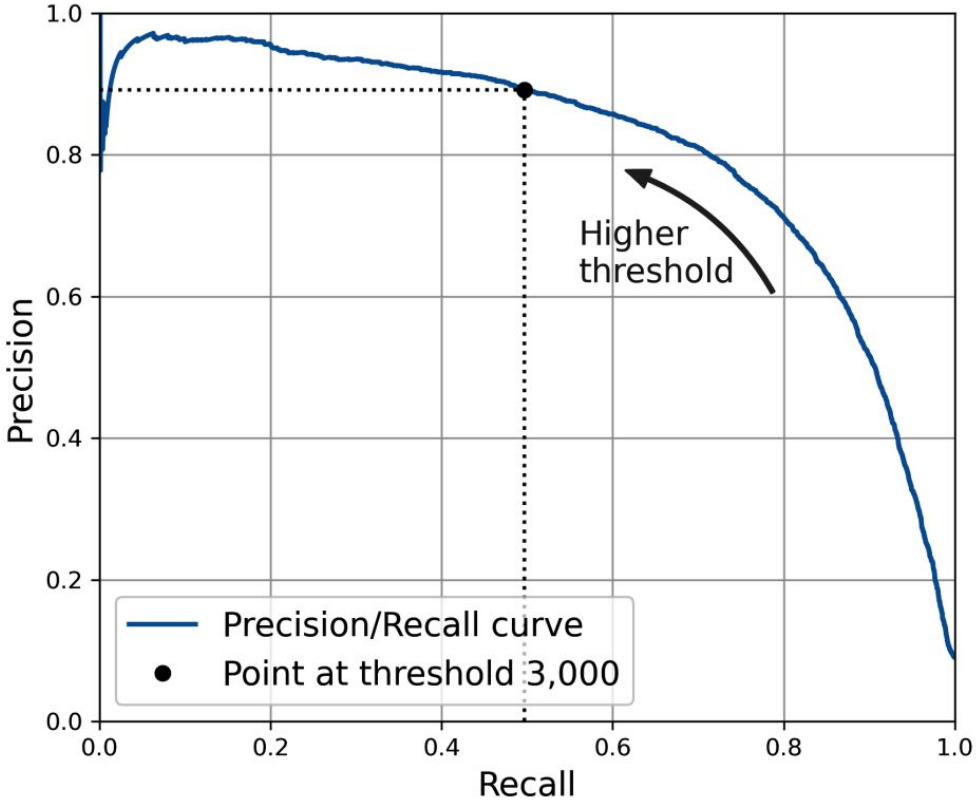


F1-score

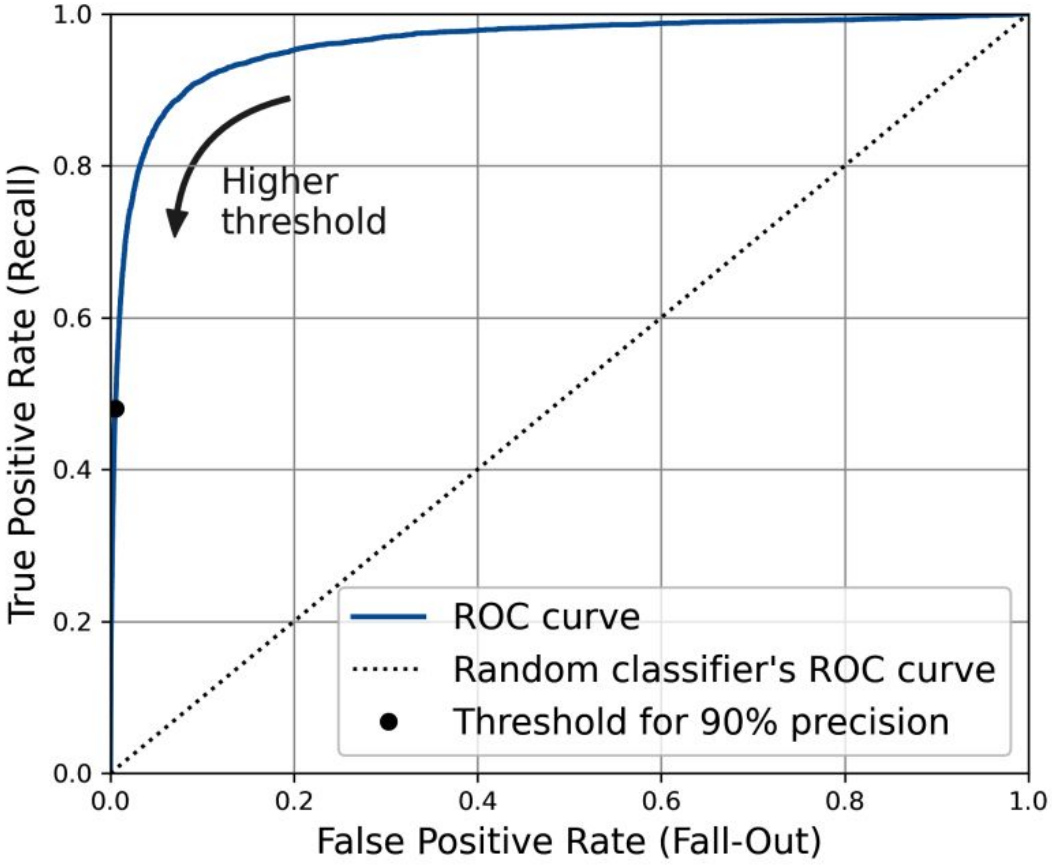
- Un score que combina precision y recall
- Tener un solo score permite comparar dos clasificadores
- Es la media armónica de precision y recall
  - Da importancia a los valores bajos
  - Un F1 alto se da sólo si precision y recall son ambos altos

$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{FN + FP}{2}}$$

Curva precision vs recall

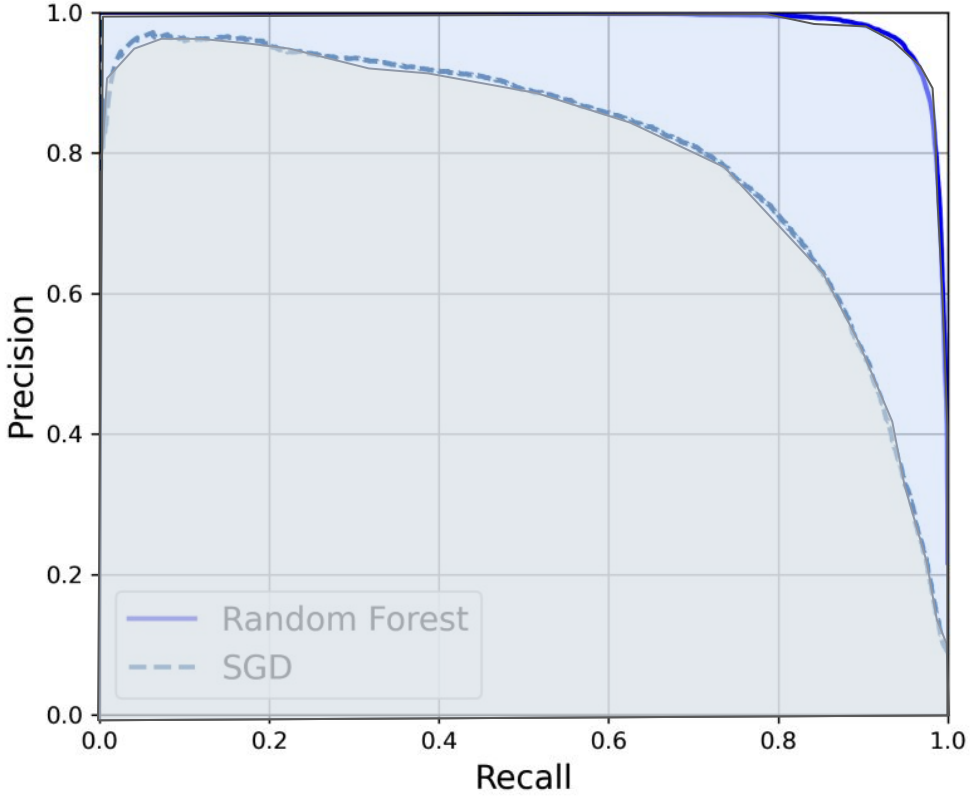


Curva ROC - Receiver Operating Characteristic

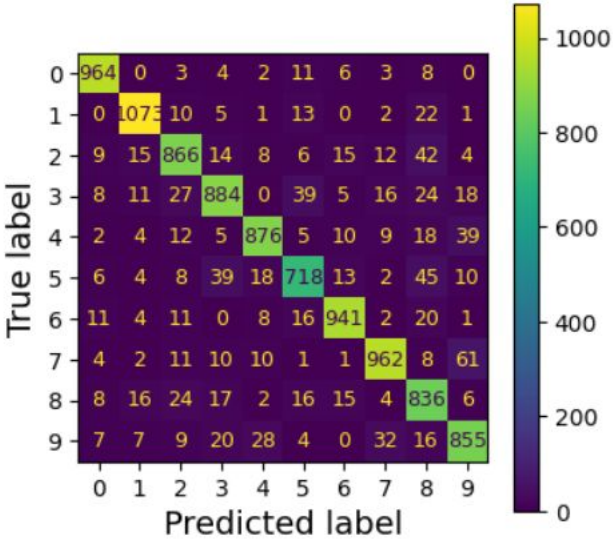


Curva precision vs recall. Comparación de algoritmos

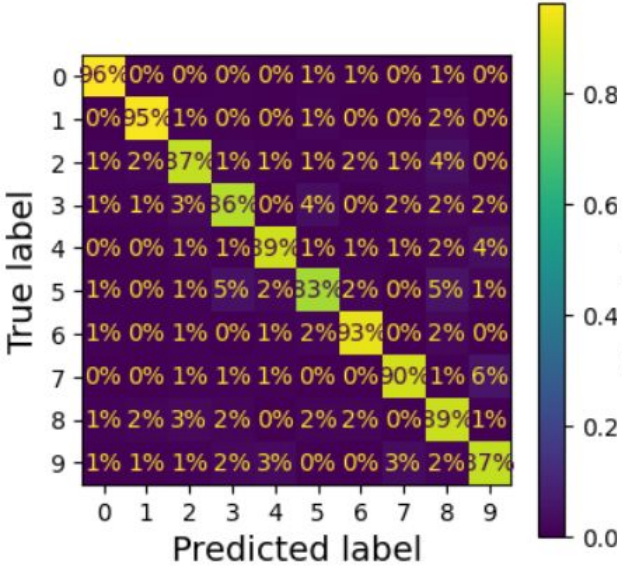
PR-AUC
Area under curve



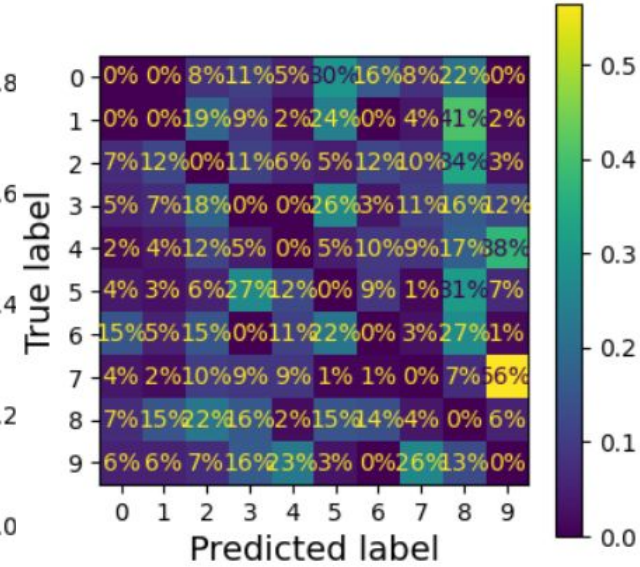
Matriz de confusión (multiclase)



Normalizada



Normalizada por filas




FACULTAD DE
INGENIERÍA
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY