

Algoritmos de Aprendizaje Automático y sus aplicaciones

Trabajo de Fin de Grado

Javier Díaz Bustamante Ussia

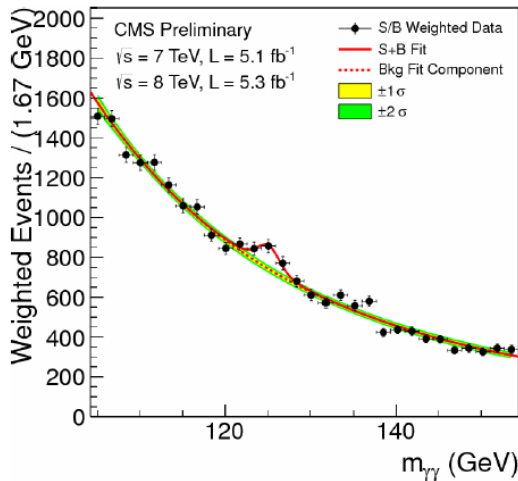
Universidad Complutense de Madrid

16 de septiembre de 2015

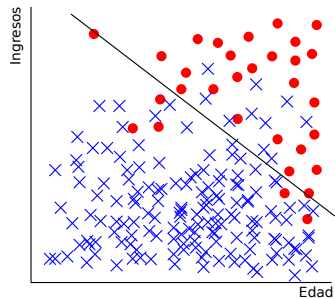
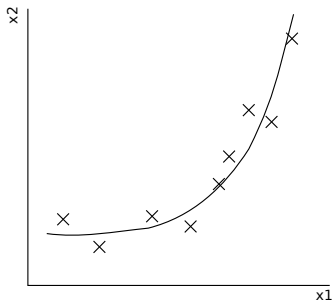
Aplicaciones del Machine Learning

- Motores de búsqueda
- Reconocimiento de escritura y voz
- Reconocimiento facial
- Sistemas de recomendaciones
- Detección de mensajes SPAM
- Detección de fraudes en transacciones con tarjetas de crédito
- Conducción autónoma
- Clasificación de secuencias de ADN
- Búsqueda del bosón de Higgs
- ...

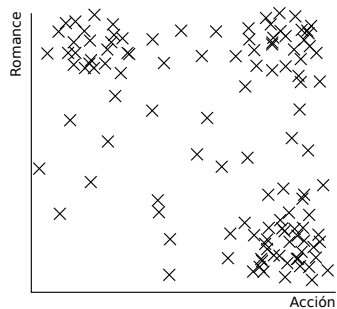
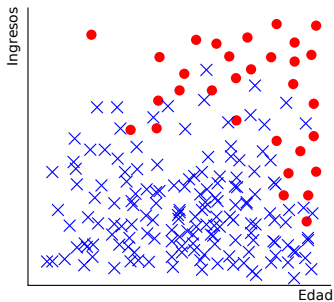
¿Cómo ayuda a buscar el Higgs?



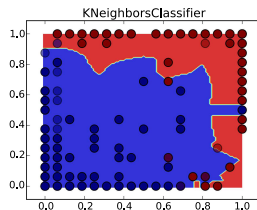
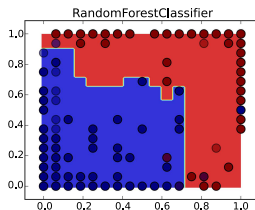
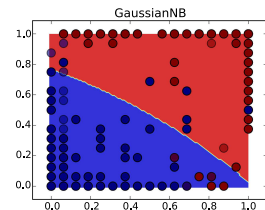
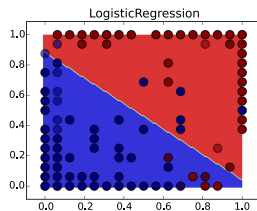
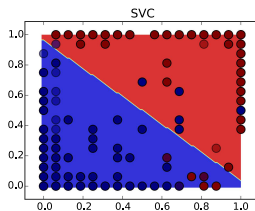
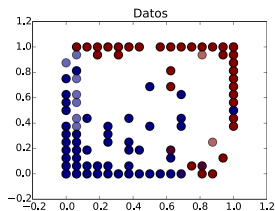
Clasificación vs. Regresión



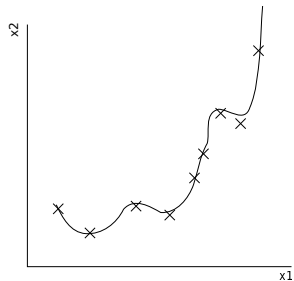
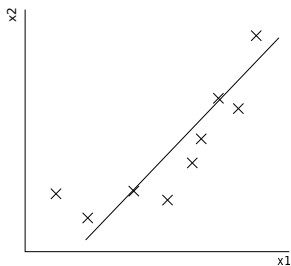
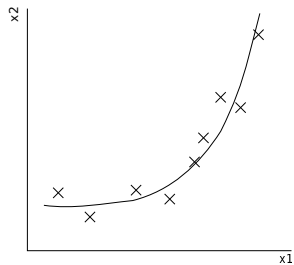
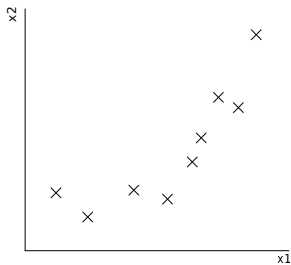
Supervisado o no supervisado



Superficies de decisión



Sesgo y varianza



Métricas

	0 real	1 real
0 predicho	Verdadero negativo (TN)	Falso negativo (FN)
1 predicho	Falso positivo (FP)	Verdadero positivo (TP)

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

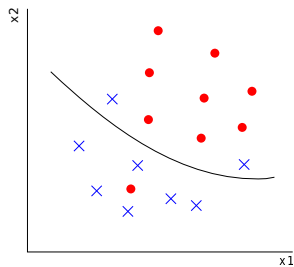
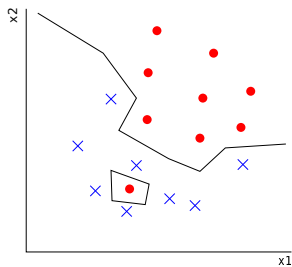
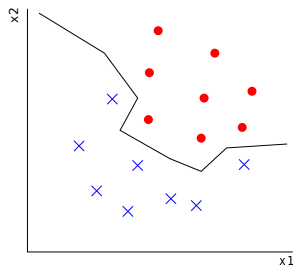
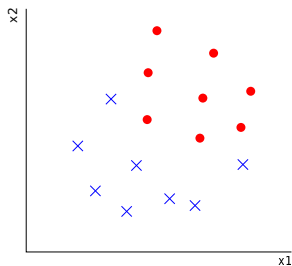
$$\text{Specificity} = \frac{TN}{FP + TN}$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

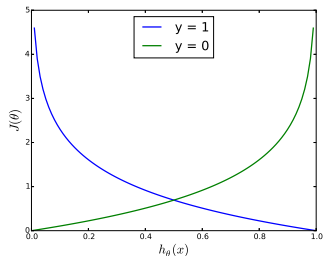
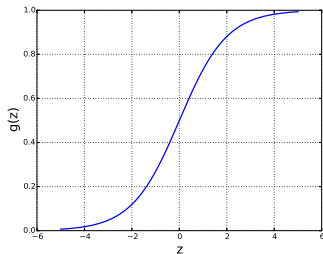
$$\text{Recall} = \frac{TP}{TP + FN}$$

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

K-Nearest Neighbors



Logistic Regression



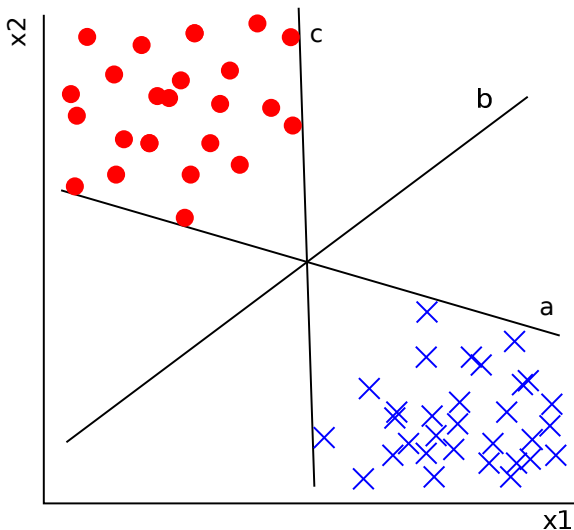
Naïve Bayes

$$p(C_I|\mathbf{x}) = \frac{p(C_I)p(\mathbf{x}|C_I)}{p(\mathbf{x})}$$

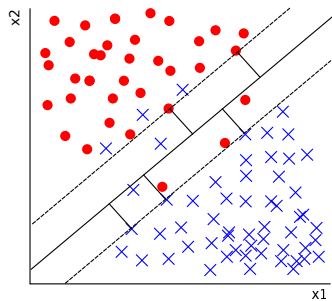
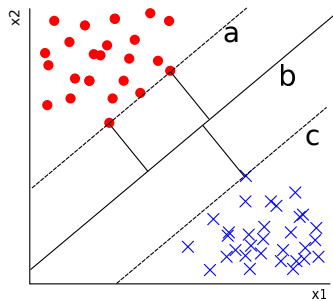
$$p(\mathbf{x}|C) = \prod_{j=1}^n p(x_j|C)$$

$$k = \max_I p(C_I|\mathbf{x})$$

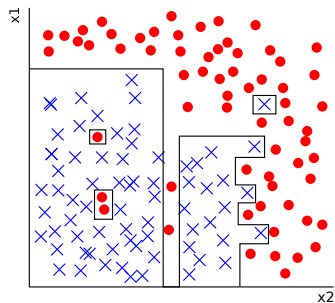
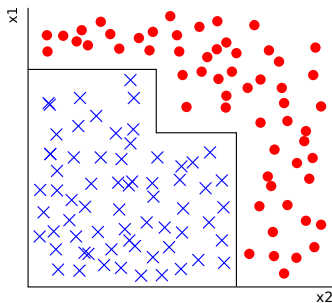
Support Vector Machines



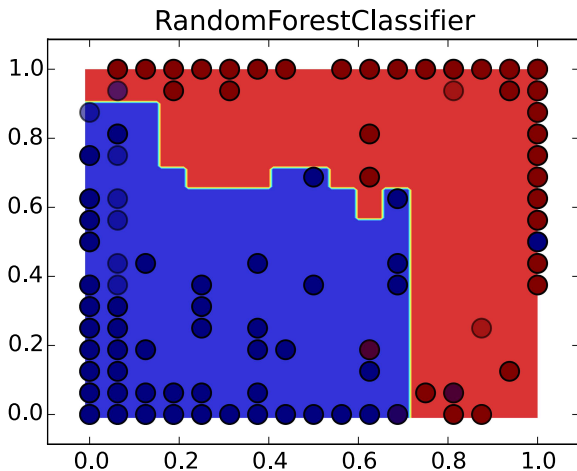
Support Vector Machines



Random Forest, árboles de decisión



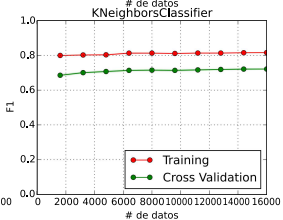
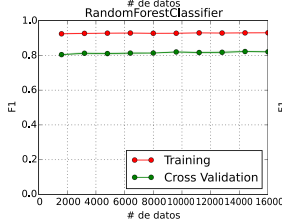
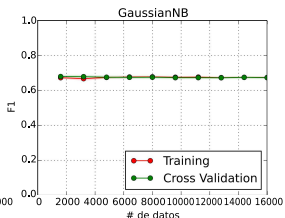
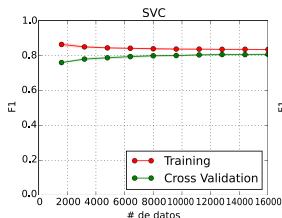
Random Forest, ensemble methods



Resultados, el Higgs

Algoritmo	Accuracy	Precision	Recall	F1-Score
Gaussian NB	0.6858	0.7367	0.8144	0.7736
SVM	0.8129	0.8413	0.8826	0.8614
Logistic Regression	0.7412	0.7810	0.8608	0.8190
Random Forest	0.8252	0.8450	0.9000	0.8716
KNN	0.7249	0.7855	0.8015	0.7934

Resultados, el Higgs



Resultados

El código utilizado para los cálculos de este trabajo ha sido realizado en Python, con ayuda del paquete de Inteligencia Artificial `scikit-learn`. Todos los cálculos han sido llevados a cabo por un ordenador con sistema operativo Windows 7 de 64 bits, un procesador Intel® Core™ i3 a 1.7 GHz, con 4 GB de memoria RAM.

Todo el código, datos y gráficas vienen recogidos en el CD con material suplementario entregado junto a la memoria del trabajo.