

¿Cómo aprenden las máquinas?

La rama que nos interesa de la inteligencia artificial postula que la forma de imitar la inteligencia humana por parte de una máquina se puede lograr a través del aprendizaje.

Formalmente Tom Mitchell en su libro “**Machine Learning**” define al aprendizaje automático como:

Se dice que un programa de computadora aprende de la experiencia E respecto a una tarea T y una medida de desempeño P , si el desempeño medido con P en una tarea T , mejora con la experiencia E .

Importante

- ▶ El ML es un mecanismo inductivo de búsqueda de conocimiento (**Sesgo inductivo**).
- ▶ En la práctica, muchas veces el entendimiento es mucho más importante que la exactitud (**Navaja de Ockham**).

Pluralitas non est ponenda sine necessitate

¿Cómo aprenden las máquinas?

En una clasificación de dos clases los errores posibles son:

- ▶ Clasificar una clase positiva como negativa.
- ▶ Clasificar una clase negativa como positiva.

Matriz de confusión

Clase Real	Clase Predicha	
	Positiva	Negativa
Positiva	TP	FN
Negativa	FP	TN

Medición de errores

Precision

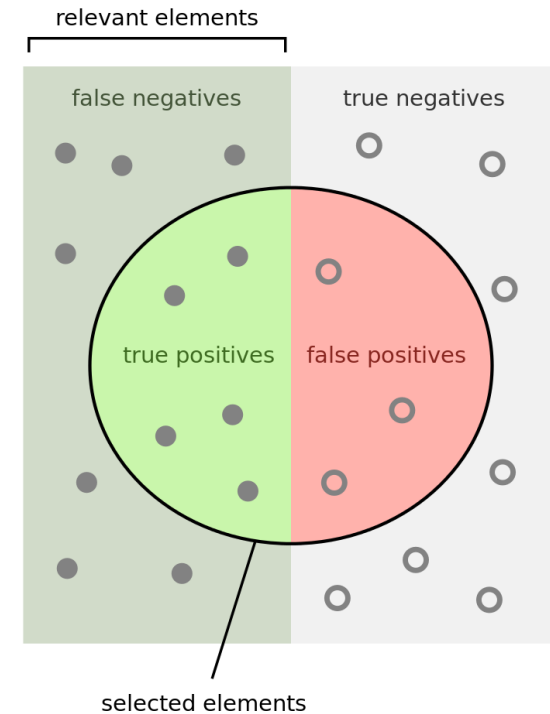
$$Precision = \frac{TP}{TP + FP}$$

Recall

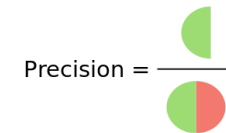
$$Recall = \frac{TP}{TP + FN}$$

Tasa de Falsos-Positivos (FPR)

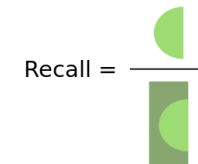
$$FPR = \frac{FP}{FP + TN}$$



How many selected items are relevant?



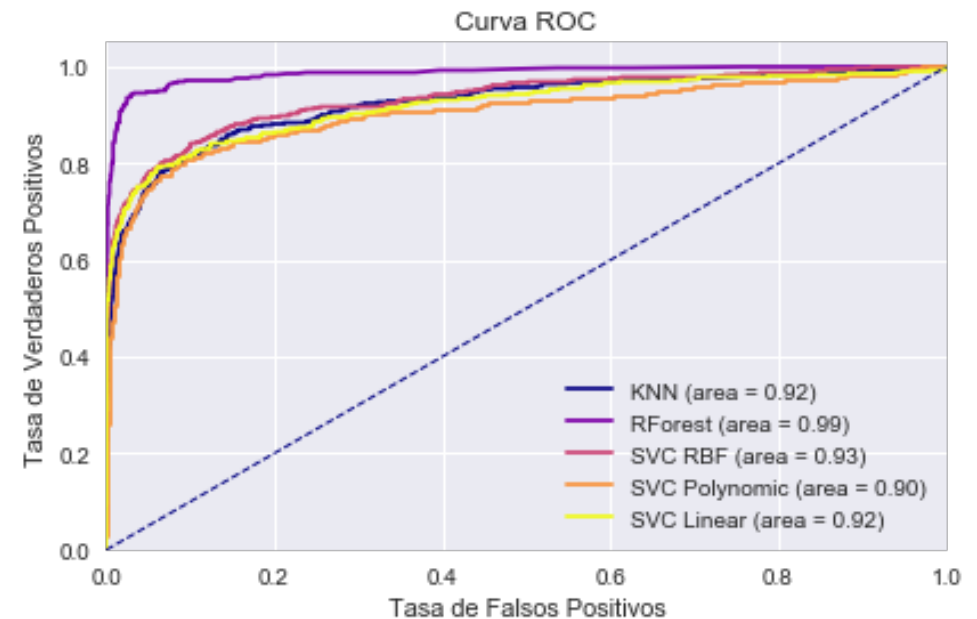
How many relevant items are selected?



Medición de errores

Curva ROC

- ▶ Algunos clasificadores entregan un puntaje que representa con qué probabilidad cada ejemplo es miembro de una clase.
- ▶ Puede ser utilizado para producir muchos clasificadores variando el umbral de pertenencia de una clase a la otra.
- ▶ Si se grafican el *Recall* y el *FPR* correspondientes para cada variación, lo que se obtiene es una potente herramienta de evaluación llamada “Curva Característica Operativa del Receptor” (*ROC*)



Consideraciones Importantes

Sobreajuste

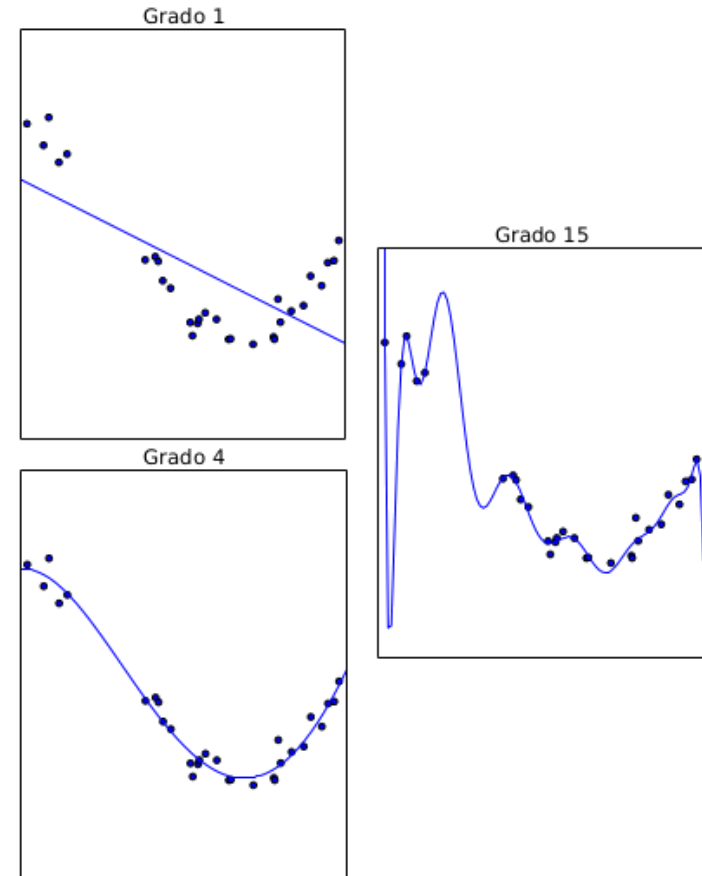
Los patrones explicados en el entrenamiento disminuya la predicción de datos nunca vistos. Se percibe como una estimación optimista.

Se soluciona dividiendo el conjunto de entrenamiento en: **Entrenamiento, Prueba, Validación.**

Desbalance de clases

Sucede cuando la distribución de etiquetas de entrenamiento no es uniforme.

Soluciones: - Sobre-muestro aleatorio. - Sub-muestro aleatorio



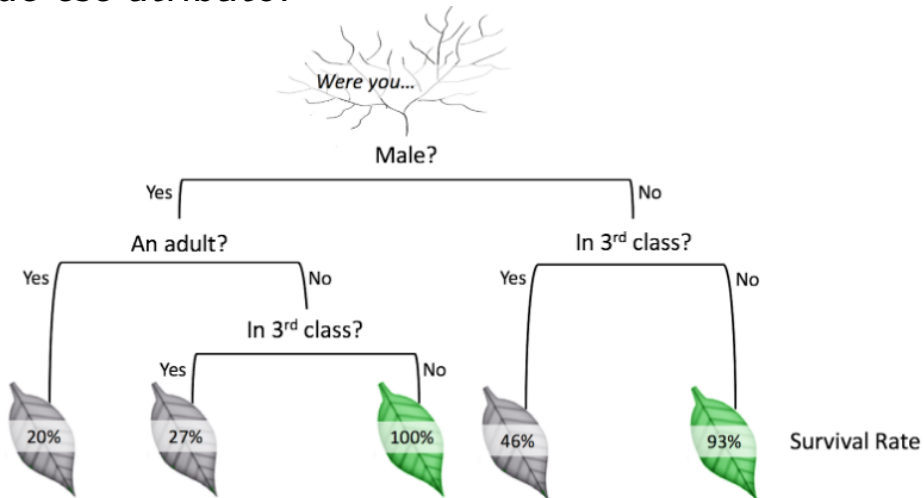
Selección de características

- ▶ Su utilidad es la de obtener un subconjunto de características más relevantes del conjunto completo de características, según una función de criterio determinada.
- ▶ Dentro de estos métodos se destaca el llamado “*Recursive Feature Elimination*” o “Eliminación recursiva de características” (*RFE*)
- ▶ Opera de la siguiente manera:
 1. Se calcula la importancia de todas las características utilizando un clasificador.
 2. Se eliminan las características con menor puntaje.
 3. Se repiten los pasos anteriores hasta que el número de características sea el deseado o se cumpla alguna condición de parada dada.
- ▶ Las condiciones de corte pueden ser varias, como por ejemplo que una métrica dada (*precision*, *recall* o *AUC*) disminuya o se les asigne una cota mínima o máxima.

Métodos de aprendizaje automático

Árboles de decisión (DT)

Los **DT** clasifican instancias ordenándolas en un árbol partiendo desde la raíz hasta las hojas. Cada nodo especifica un atributo, y cada rama descendente de ese nodo corresponde a un valor de ese atributo.



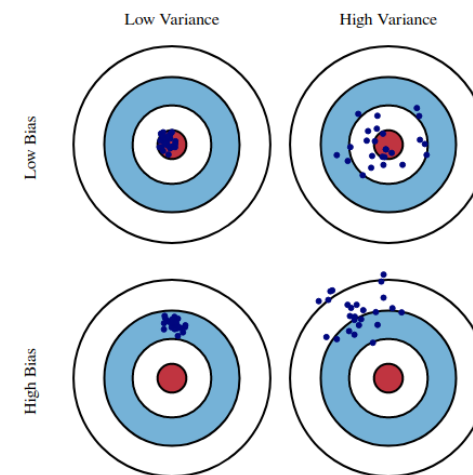
- ▶ Son fáciles de entrenar.
- ▶ Son fáciles de interpretar.
- ▶ Suele no alcanzar para capturar complejidad.

Random Forest (RF)

Utiliza un conjunto de árboles de decisión, entrenados sobre un sub-conjunto aleatorio de datos lo cual los hace especializarse y tener una muy buena predicción cuando trabajan en conjunto. Estos árboles votan la clase de un ejemplo dado.

Hiper-parámetros

- ▶ Cantidad de árboles.
- ▶ Cantidad de variables por predictor.



Métodos de aprendizaje automático

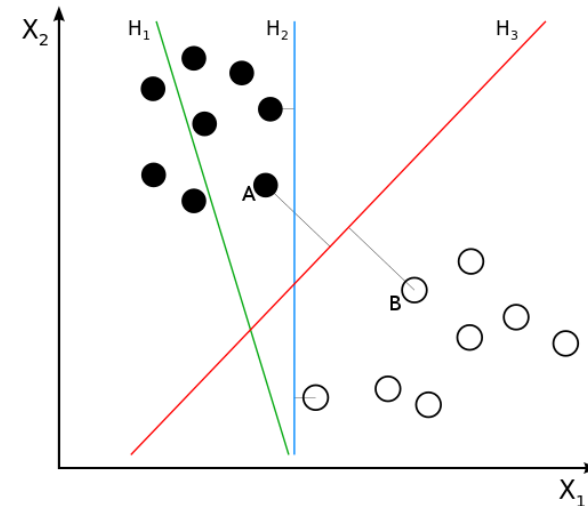
Máquinas de vectores de soporte (SVM)

El modelo representa a los ejemplos como puntos en un espacio vectorial, donde clases están divididas por una superficie con un claro margen.

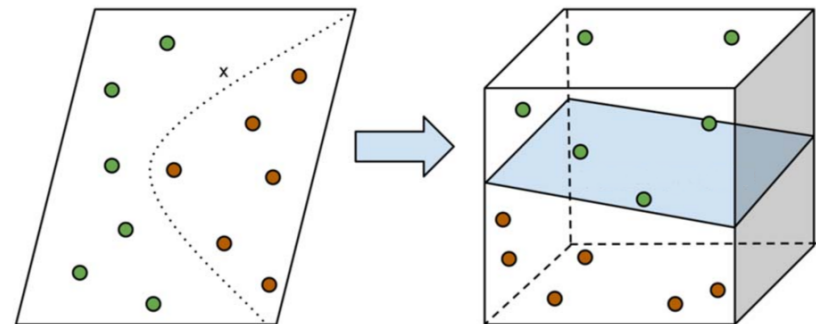
El ejemplo:

- ▶ H_1 no separa linealmente las clases.
- ▶ H_3 es el hiper-plano con mayor distancia a los puntos más cercanos y por lo tanto el que mejor generaliza.

En la práctica se utiliza el “margen débil”, que soporta una tolerancia a puntos que quedan del lado incorrecto del hiper-plano.



Kernel



Métodos de aprendizaje automático

K-vecinos más cercanos (KNN)

La clase asignada a una observación es la que con más frecuencia se encuentre dentro de los K vecinos más próximos del conjunto de entrenamiento.

Es un tipo de algoritmo conocido como “basado en instancias” donde la función sólo es aproximada localmente y toda la computación del resultado se retrasa al momento de la clasificación.

