

# Aprendizaje Automático – Guía de Ejercicios \*

Departamento de Computación – FCEyN  
Universidad de Buenos Aires

Primer cuatrimestre 2024  
Versión: 26 de marzo de 2024

## 1. Repaso probabilidad y estadística

Para resolver los siguientes ejercicios recomendamos leer el **Capítulo 6** del libro Mathematics for Machine Learning de Deisenroth, Faisal y Soon Ong. <https://mml-book.github.io/book/mml-book.pdf>

**Ejercicio 1.1.** Explique por qué los siguientes eventos son independientes de a pares pero no independientes entre todos. Dadas 2 monedas,

- (a) la primera moneda es cara;
- (b) la segunda moneda cara;
- (c) las dos monedas son iguales.

**Ejercicio 1.2.** Demostrar el teorema de Probabilidad Total: dados una partición  $\{A_i\}_{i=1}^n$  del espacio muestral tal que  $P(A_i) > 0$  para todo  $i$ , y un evento  $B$ :

$$P(B) = \sum_{i=1}^n P(B | A_i) \cdot P(A_i)$$

**Ejercicio 1.3.**

- (a) Sugerencia, mirar este video: [https://www.youtube.com/watch?v=HZGCoVF3YvM&t=57s\(3blue1brown Bayes\)](https://www.youtube.com/watch?v=HZGCoVF3YvM&t=57s(3blue1brown Bayes))
- (b) Demostrar el teorema de Bayes: dados dos eventos  $A$  y  $B$  tal que  $P(B) > 0$ ,

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

- (c) Un local vende dos marcas de televisores, A y B. El 40 % de los televisores vendidos son de la marca A y un 30 % de ellos tienen un defecto. Por otro lado, el 20 % de los televisores vendidos son de la marca B y el 10 % tienen un defecto. Si un televisor tiene un defecto, ¿cuál es la probabilidad de que sea de la marca A?
- (d) Supongamos que tienes dos máquinas, A y B, que producen tornillos. La longitud de los tornillos producidos por cada máquina sigue una distribución normal. Se sabe que la máquina A produce tornillos con una longitud media de 10 cm y una desviación estándar de 0.5 cm, mientras que la máquina B produce tornillos con una longitud media de 10.5 cm y una desviación estándar de 0.7 cm.

Ahora, supongamos que se selecciona un tornillo al azar de la producción combinada de ambas máquinas y se encuentra que tiene una longitud de 10.2 cm. ¿Cuál es la probabilidad de que este tornillo provenga de la máquina B?

**Ejercicio 1.4.**

- (a) Explicar con tus palabras qué es la media y qué es el desvío estándar.
- (b) En una fábrica de producción de caramelos, se mide la longitud de los caramelos producidos. Si la longitud media es de 5 cm y un desvío de 0,05 cm<sup>2</sup>. Siendo que tener un caramelo de más de 5.05 cm o menos de 4.95 cm se considera defectuoso, ¿qué significa esto en términos de la calidad de los caramelos producidos por esa fábrica?. ¿Hicieron alguna suposición sobre la distribución?

---

\* Algunos ejercicios fueron adaptados de los libros “Machine Learning”, de Tom Mitchell (McGraw-Hill, 1997); “Pattern Recognition and Machine Learning”, de Christopher Bishop (Springer, 2006); y “An Introduction to Statistical Learning”, de James, Witten, Hastie & Tibshirani (Springer, 2015).

### Ejercicio 1.5.

Has realizado un experimento en el que lanzaste una moneda 10 veces y la secuencia observada de resultados fue HHHHTHTTHHT.

1. Suponiendo que la moneda es justa (es decir, la probabilidad de que salga cara ( $P(H)$ ) = 0,5 y la probabilidad de que salga cruz ( $P(T)$ ) = 0,5), calcule la probabilidad de observar la secuencia dada.
2. Es más probable que la moneda esté sesgada hacia la cara en 70 % ó que la moneda este balanceada dados los datos observados. (HHHTHTTHHT).
3. Imaginen ahora que queremos estimar la carga de la moneda lo mejor posible dados los datos de la tirada. Plantear los pasos a seguir para encontrar este valor. Tip: están calculando máxima verosimilitud.
4. Calcularla :)
5. Sugerencia, ver: <https://www.youtube.com/watch?v=Dn6b9fCIUpM&t=195s> (statquest)

## 2. Introducción

**Ejercicio 2.1.** Revisar y completar el notebook `notebook_1_herramientas.ipynb`.

**Ejercicio 2.2.** Describir para los siguientes problemas si se trata de aprendizaje supervisado o aprendizaje no supervisado. Especificar qué medida de performance y de un ejemplo de una base de datos que permita encarar el problema.

- (a) detección de discurso de odio en tweets;
- (b) recomendación de películas;
- (c) diagnóstico de tumores por imágenes;
- (d) autocompletar textos;
- (e) segmentación comercial de clientes;
- (f) autenticación biométrica (ej: huellas dactilares);
- (g) detección de fraude en tarjetas de crédito.

**Ejercicio 2.3.** Determinar para los siguientes problemas de aprendizaje supervisado si se trata de problemas de clasificación o de regresión. Para cada caso, indique un ejemplo de instancia (el valor de sus atributos) junto a una etiqueta posible especificando el tipo de cada valor.

- (a) Dado un tweet, determinar si habla en contra o a favor de un candidato presidencial.
- (b) Predecir cuánto gastará una empresa en luz el próximo semestre.
- (c) Dado un tweet, predecir la probabilidad de que hable en contra o a favor de un candidato.
- (d) Predecir a qué distancia de la facultad vive una persona.
- (e) Predecir si se gastará más o menos que \$50.000 por mes de luz el próximo semestre.
- (f) Predecir la probabilidad de que se gaste más o menos que \$50.000 por mes de luz el próximo semestre.
- (g) Predecir la nota que tendrá un alumno en un examen cuya nota puede ser 0, 1, 2, ..., 10
- (h) Predecir la nota que tendrá un alumno en un examen cuya nota puede ser "A", "R" o "I".
- (i) Predecir dónde vive una persona.
- (j) Predecir la próxima palabra a autocompletar dadas las oraciones anteriores.
- (k) Predecir el valor que tomará el dolar en los próximos diez días.

**Ejercicio 2.4.** Sea un problema de clasificación en el cual cada instancia tiene 2 atributos numéricos (coordenadas  $x$  e  $y$ ) y pertenece a una de dos clases posibles (blanco o negro).

Se tienen tres tipos de hipótesis ilustrados en la Figura que representan (a) rectas, (b) líneas verticales (hasta 30 líneas), (c) elipses (tantas como se quiera).

Para cada uno de ellos, se pide:

- Describir el espacio de hipótesis  $H$ ;
- Identificar los parámetros de la hipótesis (el conjunto de valores que permiten describir una hipótesis en concreto,  $\theta$ ).<sup>1</sup>

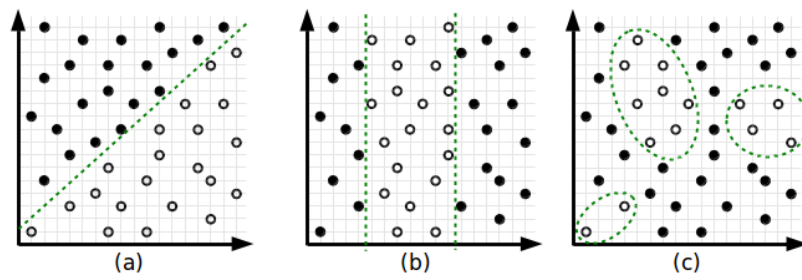


Figura 1: Tipos de Hipótesis

**Ejercicio 2.5.** Completar el notebook `notebook_2_titanic.ipynb` .

<sup>1</sup>“Machine Learning”, de Tom Mitchell (McGraw-Hill, 1997);

### 3. Árboles de decisión

Cielo	Temperatura	Humedad	Viento	¿Salgo? (clase a predecir)
Sol	Calor	Alta	Débil	No
Sol	Calor	Alta	Fuerte	No
Nublado	Calor	Alta	Débil	Sí
Lluvia	Templado	Alta	Débil	Sí
Lluvia	Frío	Normal	Débil	Sí
Lluvia	Frío	Normal	Fuerte	No
Nublado	Frío	Normal	Fuerte	Sí
Sol	Templado	Alta	Débil	No
Sol	Frío	Normal	Débil	Sí
Lluvia	Templado	Normal	Débil	Sí
Sol	Templado	Normal	Fuerte	Sí
Nublado	Templado	Alta	Fuerte	Sí
Nublado	Calor	Normal	Débil	Sí
Lluvia	Templado	Alta	Fuerte	No

Tabla 1: Salgo a caminar

**Ejercicio 3.1.** Hacer en papel y lápiz un árbol de decisión correspondiente a entrenar con los datos de la Tabla *Salgo a caminar*. Utilizar el criterio “Gini Gain” para calcular el feature que mejor separa cada decisión. Armar luego tres ejemplos posibles de instancias nuevas (no existentes en la tabla) y usar el árbol para predecir la clase de salida. Calcular además la importancia de cada atributo (decrecimiento total en la impureza Gini).

**Ejercicio 3.2.** ¿Cómo cambiaría el árbol si restringiéramos su altura a 2 niveles?

(a) ¿Cuál sería el resultado de las predicciones del ejercicio anterior?

(b) ¿Algunas de las instancias existentes en la tabla, serían clasificadas incorrectamente?

**Ejercicio 3.3.** ¿Quién es quién?

El quién es quién es un juego (buscar en Google) en el que hay que adivinar el nombre del personaje del rival. Para ello se hacen preguntas que son respondidas por sí o por no.

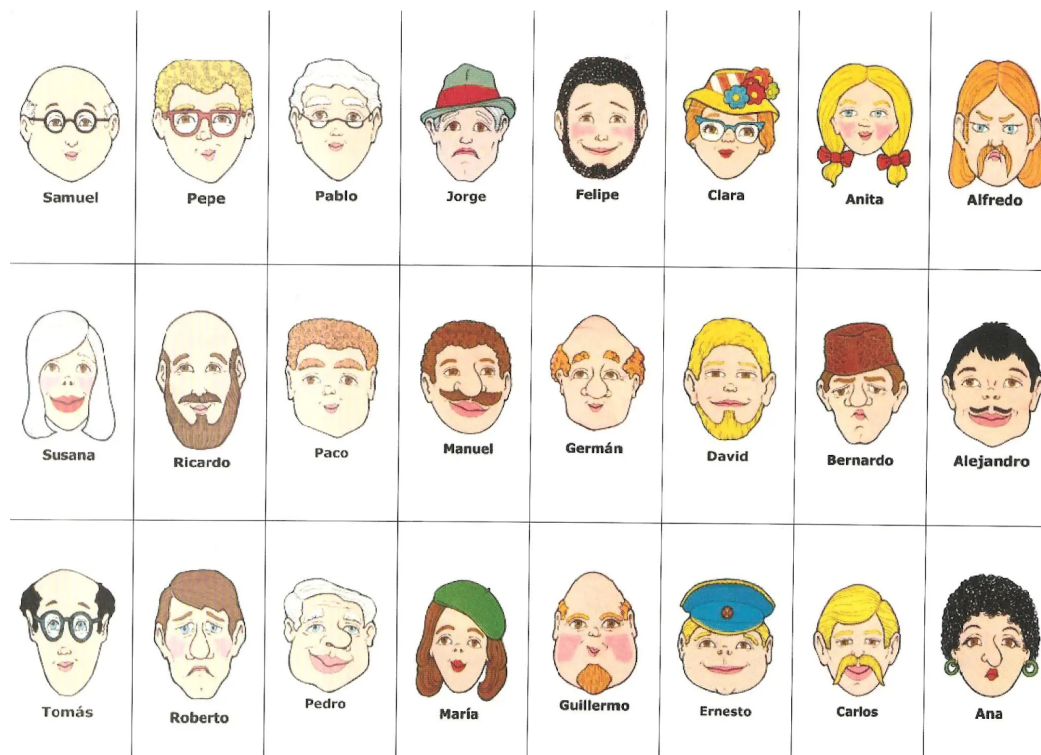


Figura 2: Juego de ¿quién es quién?

Tomemos la libertad de jugar únicamente a adivinar el sexo biológico (masculino/femenino).

- (a) Si preguntamos si el personaje es calvo... ¿Cuál es la ganancia gini al separar por este atributo?
- (b) Ordenar por valor de ganancia gini qué conviene como primera pregunta: i) ¿Tiene calvicie? ii) ¿Tiene cabello rubio? iii) ¿Tiene vello facial? (bigote y/o barba) iv) ¿Tiene sombrero?

**Ejercicio 3.4.** En la Figura *Cortes en el espacio de atributos* puede verse diversas regiones en el espacio de atributos.

- Determinar cuáles de ellas pueden haber sido generadas por árboles de decisión.
- Para las que lo sean, mostrar un árbol que hubiese generado estas regiones (suponer ejes  $x_1$  y  $x_2$ )

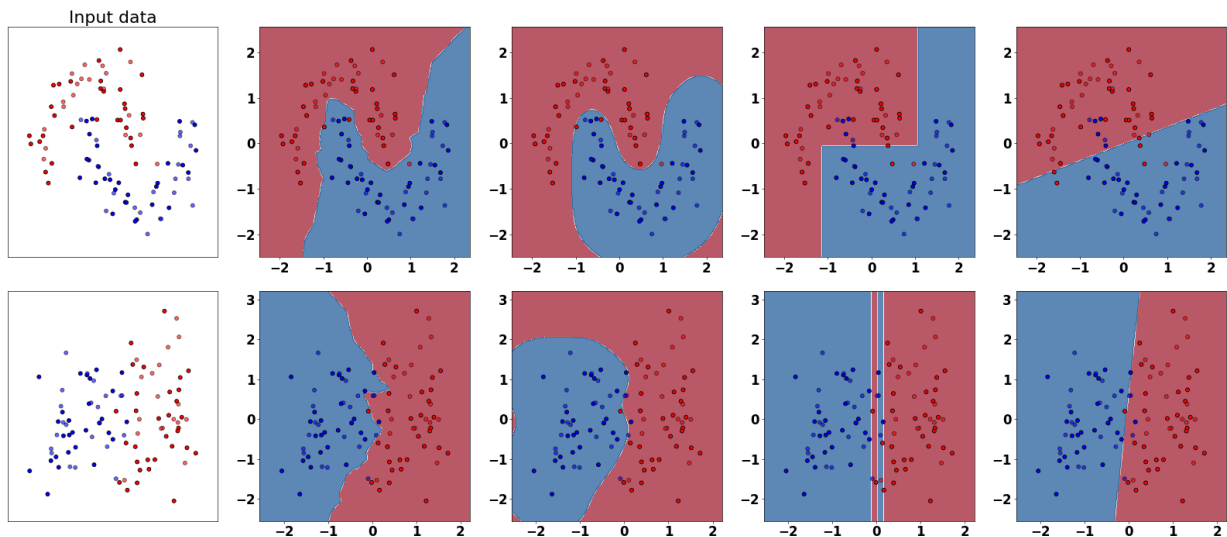
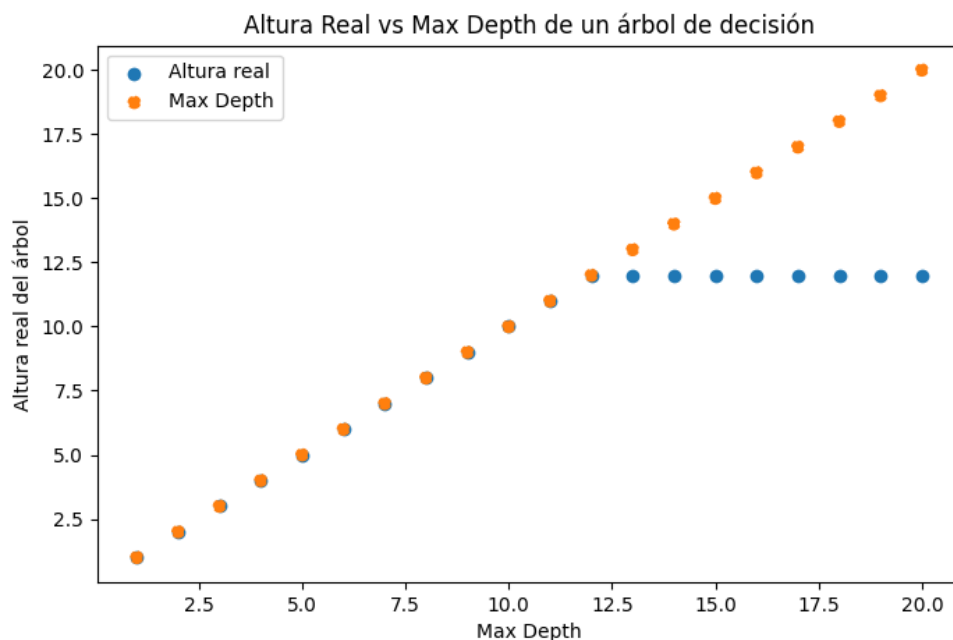


Figura 3: Cortes en el espacio de atributos

**Ejercicio 3.5.** Se entrenan 20 árboles distintos variando el parámetro  $max\_depth$  desde 1 hasta 20 inclusive. Graficamos, para cada uno de estos árboles, el valor de  $max\_depth$  y el valor de la altura real de estos. Responder:



- ¿Por qué sucede que, a partir de  $max\_depth$  igual a 12, la altura real del árbol es constante?
- Dado  $n$  instancias, ¿cuál sería el valor máximo de  $max\_depth$  que tiene sentido utilizar?

**Ejercicio 3.6.** Dado el algoritmo de construcción de árboles visto en clase para atributos continuos: Sea  $S$  una muestra de instancias con atributos  $A$ . Para construir un árbol de decisión ejecutamos:

(1) Mientras no se cumpla un criterio de detención:

- (I) Creamos `nodo_actual`, un nodo que representa una decisión de corte.
- (II) Elegimos el par  $a \in A, c \in R$  entre los posibles pares  $\langle \text{atributo}, \text{corte} \rangle$ , que mejor divida a  $S$  para `nodo_actual` según  $\Delta M(S, \langle a, c \rangle)$   
( $\Delta M$  es una función que devuelve cuánto gana si dividido a  $S$  en dos utilizando  $\langle a, c \rangle$  y según la medida  $M$ . Por ejemplo  $\Delta M$  podría ser `GananciaDeInformación`).
- (III) Crear dos hijos de `nodo_actual`.
- (IV) Dividir las instancias de  $S$  en los nuevos nodos, según  $\langle a, c \rangle$ :

$$S_{\leq} \leftarrow \{x | x \in S \wedge x[a] \leq c\}$$

$$S_{>} \leftarrow \{x | x \in S \wedge x[a] > c\}$$

(v) Repetir para cada hijo.

(2) El valor asignado a cada región resultante (a cada hoja) será el de la clase mayoritaria de las instancias que pertenezcan a esta región.

Se pide:

- (a) Escribir el pseudocódigo (puede ser similar a python) para el paso (b) (elegir el mejor par  $a, c$  entre los posibles pares). Es decir, definir `mejor_corte(S, A, ΔM)`, en donde  $S$  representa la muestra,  $A$  un conjunto de atributos y  $\Delta M$  la función que computa la ganancia de una división. Puede suponer dadas funciones tales como  $S[a]$  que devuelve la columna de valores para el atributo  $a$ . Es importante que esta función no evalúe dos veces cortes que devuelven exactamente las mismas regiones para un mismo atributo.
- (b) Introducir los cambios necesarios en el algoritmo general (y si fuera necesario, en la función que definieron para `mejor_corte`) que permita medir la importancia de cada atributo. La importancia deberá ser un valor numérico entre 0 y 1.

**Ejercicio 3.7.** Determinar cuáles de las siguientes son afirmaciones verdaderas:

- (a) El objetivo de construir un árbol de decisión es crear el árbol de menor tamaño posible en el cual las hojas contengan valores de una sola clase.
- (b) Los algoritmos de construcción vistos (CART, ID3, etc) exploran todos los posibles árboles y se quedan con el que mejor separa a las instancias.
- (c) La pureza describe qué tan cerca está un nodo de contener instancias de una sola clase.
- (d) Un atributo puede aparecer sólo una vez en cada rama del árbol (llamamos rama a un camino directo desde una hoja hasta la raíz).
- (e) Un par (atributo, corte) puede aparecer sólo una vez en cada rama del árbol (llamamos rama a un camino directo desde una hoja hasta la raíz).
- (f) Para cada nueva instancia, un árbol permite predecir la clase a la que pertenece. Por otra parte, para predecir **la probabilidad** de pertenecer a una clase u otra, es necesario modificar el algoritmo de creación de árboles.

**Ejercicio 3.8.** Resolver el notebook `notebook_3_arboles_de_decision_sklearn.ipynb`.

**Ejercicio 3.9.** Preguntas conceptuales para discutir:

- (a) ¿Cuál el sesgo inductivo del algoritmo que construye el árbol de decisión?
- (b) ¿Qué sucede cuando dos atributos empatan en ganancia de información? ¿Esta decisión es parte del sesgo inductivo?
- (c) ¿Cómo se comporta la ganancia de información en comparación a impureza Gini cuando se comparan atributos con gran cantidad de valores distintos? Por ejemplo, si el atributo  $x_1$  tiene dos valores posibles (true y false) y el atributo  $x_2$  tiene 40 valores distintos, ¿es justo usar ganancia de información para elegir entre ellos? ¿Qué desventajas tiene? ¿Cómo se podría mitigar?
- (d) Dado un atributo  $a$  continuo en la cual queremos encontrar el mejor corte  $c$  según alguno de los criterios considerados (Gini o Entropy), ¿cuál es la complejidad de peor caso si se asume que tenemos  $n$  instancias en nuestro conjunto de datos?

**Ejercicio 3.10.** Completar el notebook `notebook_4_implementacion_arbol.ipynb`. Este notebook contiene una implementación parcial de un algoritmo de creación de árboles de decisión.

**Ejercicio 3.11.** Dado el algoritmo de construcción de árboles para atributos continuos (ver algoritmo en el ejercicio 3.6) ¿Qué cambios son necesarios para que funcione para problemas de regresión? Para pensar:

- ¿Pueden las medidas vistas en clase (CER, Gini, Entropía) ser utilizadas en este caso?
- ¿Qué cambiarían en el criterio de detención?
- ¿Cómo cambiarían el valor asignado a los nodos hojas?