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) Segmentación de imágenes en los objetos que contiene;

(c) Detección de fraude en tarjetas de crédito.

(d) Agrupar voces según qué tan parecido suenan.

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 los posibles valores que puedan tomar las etiquetas, especificando detalladamente **el tipo de datos (computacional)** que le corresponde. Tip 1: Imaginar la base de datos (es decir, qué le hubieran pedido a anotadores expertos que completen). Warning: chatGPT no hace bien este ejercicio
Ejemplo:

- **Problema**: Dado un tweet, determinar si habla a favor de un candidato presidencial.
- **Etiquetas posibles**: `Sí`, `No`.
- **Tipo etiquetas**: `bool`
- **Respuesta**: Clasificación

(a) Predecir cuánto gastará una empresa en luz el próximo semestre.

(b) Dado un tweet, predecir la probabilidad de que hable en contra o a favor de un candidato.

(c) Predecir a qué distancia de la facultad vive una persona.

(d) Predecir si se gastará más o menos que $50.000 por mes de luz el próximo semestre.

(e) Predecir la probabilidad de que se gaste más o menos que $50.000 por mes de luz el próximo semestre.

(f) Predecir la nota que tendrá un alumno en un examen cuya nota puede ser $0, 1, 2, \ldots, 10$

(g) Predecir la nota que tendrá un alumno en un examen cuya nota puede ser “A”, “R” o “I”.

(h) Predecir dónde vive una persona.

(i) Predecir la próxima palabra a autocompletar dadas las oraciones anteriores.

(j) Predecir, dada una imagen, qué subconjunto de los siguientes elementos aparece: { pelota, niños, cielo, bicicleta }

Revisar el punto (d): ¿Qué responderían si en la base de datos tenemos etiquetas $\mathbb{R}$? ¿Y si tuviéramos etiquetas binarias?

Ejercicio 2.4. Sea un problema de clasificación en el cual cada instancia tiene 2 atributos numéricos y pertenece a una de dos clases posibles (blanco o negro). Se tienen tres tipos de hipótesis ilustrados en Fig 1 que representan (a) rectas y la dirección en la cual se clasifica a una instancia como blanco, (b) dos rectas sin inclinación (ya sea horizontales o verticales, entre las cuales las instancias son clasificadas como blancos, (c) 3 elipses (que delimitan instancias blancas). Para cada uno de ellos, se pide:

- Describir el espacio de hipótesis H;
- Identificar la cantidad de parámetros mínimos para describir una hipótesis de esta forma.

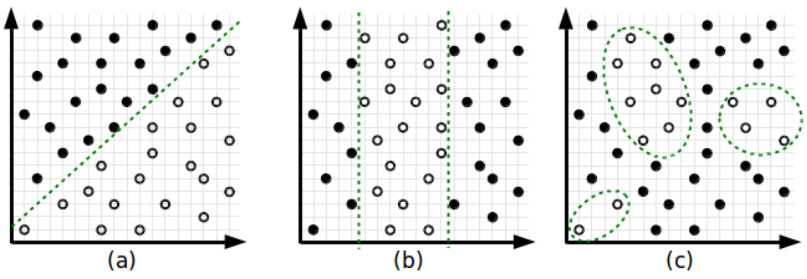

Figura 1: Tipos de Hipótesis

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

Ejercicio 2.6. Queremos construir un clasificador a partir de un conjunto de datos en el plano (x_1, x_2) con dos clases posibles: $\{A, B\}$.

(a) Supongamos que nuestro sesgo inductivo es que las hipótesis permitidas son siempre **rectas que dividen el plano** en dos regiones (una para la clase A y otra para la clase B). Supongamos también que los puntos son linealmente separables (es decir, hay una recta que hace que todas las instancias de una clase queden de un lado y el resto del otro). Escribí un algoritmo que, dados ejemplos etiquetados, elija una de estas rectas. Si es posible, un pseudocódigo (o python) para estar seguros que no hay ambigüedades.

(b) Definí/dibujá un ejemplo de un conjunto de puntos etiquetados con clases A y B y proponé una recta que los separe según tu regla.

(c) ¿Pensás que tu algoritmo encontró la recta "que mejor separa". ¿Si sí, con respecto a qué métrica de separación?

(d) Cambiarías algo en tu algoritmo si quitaramos la supoción de que los puntos son linealmente separables.

Ejercicio 2.7. Consideremos un problema de clasificación binaria con dos atributos discretos: **Color** $\in$ {rojo, verde, azul, amarillo}; **Tamaño** $\in$ {chico, grande}. Las clases pueden ser **A** o **B**.

(a) ¿Cuál es el tamaño del **espacio de instancias** X?

(b) Suponiendo clasificación binaria, ¿cuál es el tamaño del **espacio de hipótesis** (es decir, todas las posibles funciones objetivo $f^* : X \to \{A, B\}$)?

(c) Supongamos ahora un **sesgo inductivo**: nuestro algoritmo sólo puede construir hipótesis de la forma "*Si Color = rojo entonces A; en otro caso B*" o bien la regla inversa "*Si Color = rojo entonces B; en otro caso A*". ¿Cuántas hipótesis distintas permite este sesgo inductivo?

(d) Elegí un **nuevo sesgo inductivo** posible. Escribí claramente cómo serían las hipótesis permitidas bajo tu sesgo. ¿Cuántas hipótesis posibles habría en tu espacio de hipótesis?

(e) Diseñá, con tus palabras, un **algoritmo** sencillo que, dado un conjunto de ejemplos etiquetados, elija una hipótesis dentro del espacio delimitado por tu sesgo inductivo.

Ejercicio 2.8. Entropía (ver `https://www.youtube.com/watch?v=YtebGVx-Fxw`)
Se tiene un dado equilibrado con 6 resultados equiprobables.

(a) ¿Cuál es la sorpresa de obtener un 5?

(b) Calculá la entropía del dado.

(c) Calcularla ahora considerando un dado cargado con probas: $p(1) = 0.5;\ p(2) = p(3) = p(4) = p(5) = p(6) = 0.1$

(d) ¿Dirías que ahora hay menos o más incertidumbre en el resultado que arrojará el dado?

(e) ¿Cuál sería la sorpresa de obtener un "ninguno" (por ej cuando el dado quedó inclinado)? Considerar $p(ninguno)$ como un valor infinitesimal que se le resta al resto de las probabilidades. Justificar.

Ejercicio 2.9. Entropía de una región
Sean 12 instancias en un plano bidimensional (x_1, x_2). Cada ejemplo está etiquetado como Clase A o Clase B.

- En la región $x_1 < 0$: 5 ejemplos (3 son clase A, 2 son clase B).
- En la región $x_1 \geq 0$: 7 ejemplos (2 son clase A, 5 son clase B).

(a) Calculá la entropía de todo el conjunto de datos (sin dividir).

(b) Calculá la entropía de cada región.

(c) ¿Dirías que la división por x_1 reduce o aumenta la incertidumbre? Pensá en este ejemplo: si tiro una nueva instancia al azar en el conjunto original (sin dividir) y la predigo según la clase mayoritaria, ¿la probabilidad de equivocarme es mayor o menor que si primero miro en qué región cae (usando x_1) y luego aplico la clase mayoritaria de esa región?

(d) Calculá la entropía promedio (promedio pesado) después de dividir en función de x_1 <0. ¿Es esta nueva entropía menor o mayor a la que tenía la región sin dividir?

(e) Considerá ahora dividir el espacio en función de $x_2 < 0$.

- En la región $x_2 < 0$: 6 ejemplos (3 son clase A, 3 son clase B).
- En la región $x_2 \geq 0$: 6 ejemplos (2 son clase A, 4 son clase B).

¿Dirías que este es un mejor o peor corte que el anterior para reducir la incertidumbre?