2020 - I semester
INFO257 - Inteligencia Artificial
Unidad 1 - Aprendizaje Supervisada

Prueba teórica

Nombre: Sebastián Lara Barría

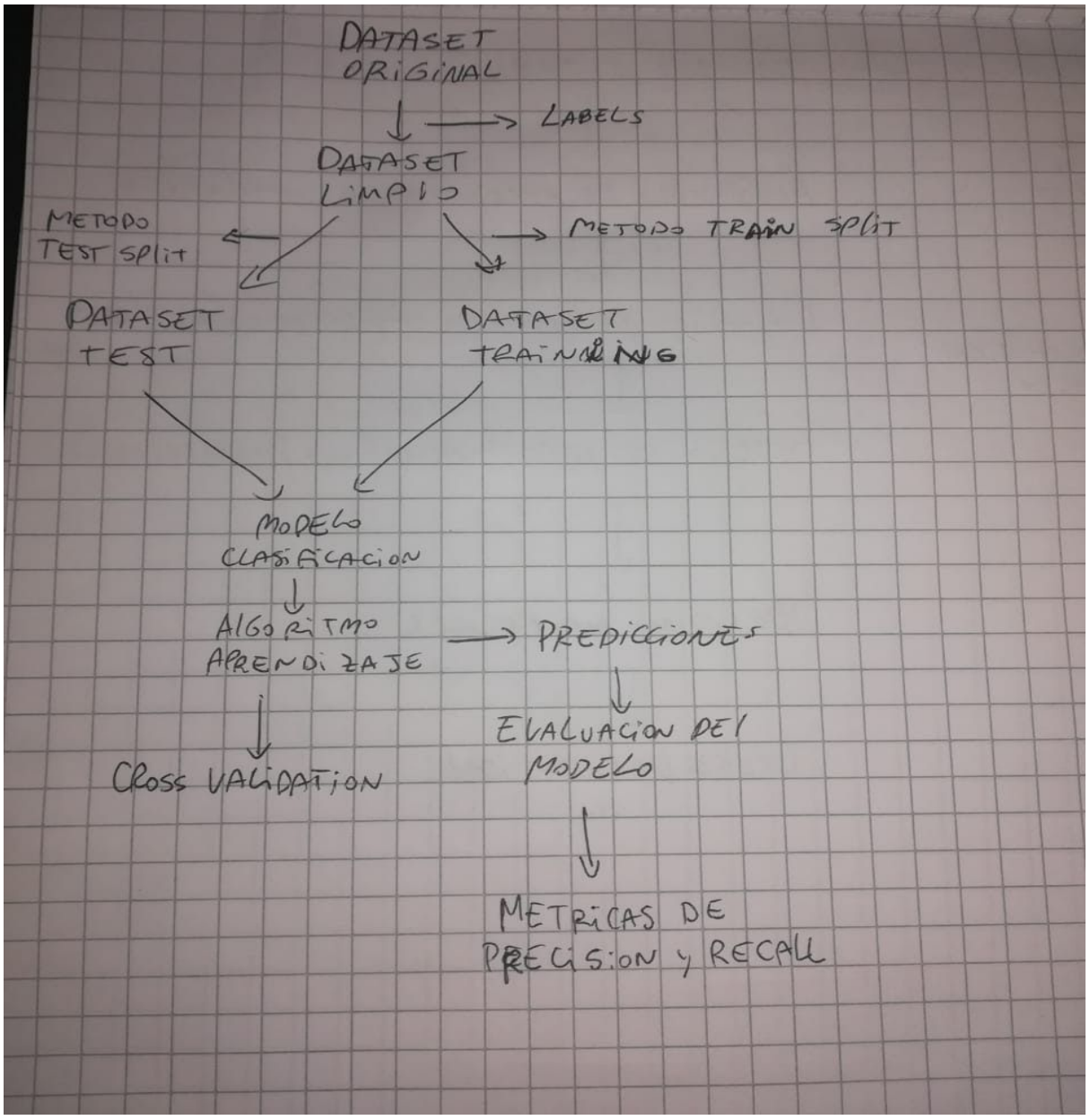
Repositorio GitHub *(tarea práctica)***:**
https://github.com/SLaraB/INFO257_2020-master_SebastianLara

Al inicio de la prueba:

1. Hacer una copia de este documento
2. Nombrar su archivo Google Doc "IA_unidad1_2020_*NombreApellido*"
3. Indicar su nombre y la direccion del repositorio GitHub en la cual se encontrará su tare práctica
4. Enviarme el enlace de su archivo Google Doc por mensaje privado Slack
5. Responder directamente en su archivo **hasta las 15.30hrs**.

Preguntas abiertas:

1) Proponer un diagrama que representa el workflow que seguir para resolver un problema de clasificación utilizando una metodología de aprendizaje supervisado. Su diagrama posicionará al menos los conceptos siguientes: dataset de entrenamiento, dataset de test, algoritmo de aprendizaje, modelo de clasificación, labels, features (características), predicciones, evaluación del modelo, métricas de precisión y recall, cross-validación o método de train/test split.



2) Supongamos que un modelo de clasificación permite obtener la matríz de confusión siguiente:

	A	B	C
Predicción (A)	94	16	10
Predicción (B)	21	113	16
Predicción (C)	4	4	92

- Definir los conceptos de Precision y Recall
  -**Precision**: con esta métrica podemos calcular la calidad del modelo de aprendizaje supervisado en términos de clasificación

  - **Recall**: A su vez este concepto, nos permite saber la cantidad que el modelo de aprendizaje supervisado es capaz de identificar.

- Calcular la Precision y el Recall
- **Precision** = True Positive / (True Positive + False Positive)
- **Recall** = True Positive / (True Positive + False Negative)

  - **Precision A =** 94/120 = 0.78333
  - **Precision B =** 113/150= 0.75333
  - **Precision C =** 92/ 100 = 0.92
  - **Recall A =** 94/119 = 0.78992
  - **Recall B =** 113/133= 0.84962
  - **Recall C =** 92/118= 0.77966

- ¿Por qué se considera que la métrica de Precisión mide el "ruido" generado por el modelo, y el Recall mide el "silencio" generado por el modelo?

Se debe a que precision suma los falsos positivos, o sea, los casos casos cuando fue negativo pero fue predecido positivo, entonces estos resultados afectan el modelo de manera negativa.

Por su parte recall, suma los falsos negativos, es decir, los casos cuando fue positivo pero fue predecido negativo, entonces estos resultados no afectan el modelo.

3) ¿Qué aprenden los algoritmos de Regresión Lineal, Regresión Logística y Árbol de decisión? (*Responder indicando el o los conceptos aprendidos por cada uno de los algoritmos*)

-**Regresión Lineal** aprende a través de datos ya sabidos donde tenemos 2 variables, la variable "y" que es la variable escalar dependiente y la cual queremos predecir dados los datos que obtenemos gracias a la "x" ,que puede ser una o varias variables. Este modelo nos permite predecir un valor como por ejemplo: cantidad de volumen de negocio, cantidad de grasa corporal, tiempo estimado para recorrer ciertos kilómetros, etc.

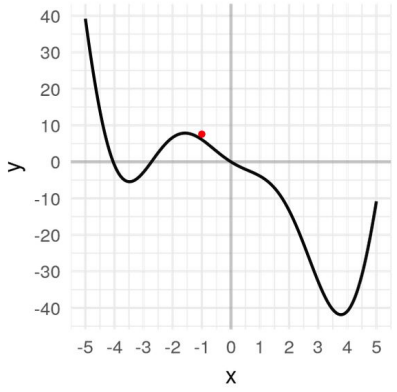
- **Regresión Logística** este modelo básicamente mide la relación entre una variable dependiente (lo que se desea predecir), con las variables independientes(conjunto de características disponibles). Es decir nos permite predecir si algo es correcto o no (True or False, 1 or 0, etc). En otras palabras nos servirá para clasificar los datos. Por ejemplo nos permite clasificar si un registro es fraude, si una persona está enferma,etc.

- **Árbol de decisión** aprende a través de una representación esquemática y así se puede clasificar o tomar decisiones de mejor manera. Ya que según ciertos parámetros ordenados por su relevancia para cada problema. Formando así un árbol binario donde cada hoja es un "filtro" y así obteniendo un camino de precisión.

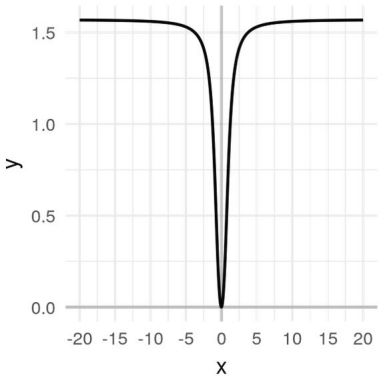
4) ¿Cómo aprende el algoritmo de Regresión Logística? (*Responder utilizando al menos 5 términos específicos*)

Este modelo primero necesita una **muestra** bastante **grande** para poder tener resultados más acertados . Luego se debe **limpiar** esta muestra para que sea lo más equilibrada posible(**eliminar ruido**) y utilizar solamente los datos que afecten directamente la salida del algoritmo. Para implementar este método se utiliza la **distribución gaussiana**, ya que este es un algoritmo lineal con una salida no lineal.

5) Supongamos que las funciones siguientes representan dos funciones de pérdida obtenida para resolver un problema de clasificación. Describir etapa por etapa cómo el algoritmo de Gradient Descent podría funcionar? ¿Qué problema podríamos tener en el caso 1 y en el caso 2?



Caso 1



Caso 2

6) Supongamos que queremos clasificar documentos textuales, cómo podríamos transformar los documentos en representaciones vectoriales? Qué podrían ser las características X de nuestras observaciones? Se podría utilizar árboles de decisión como algoritmo de aprendizaje en este contexto?

   Se puede utilizar el modelo “Bag of words” , que divide el texto en oraciones pequeñas, después de forma un alfabeto que contiene las palabras del texto y luego se crea un vector con la frecuencia en que aparece cada palabra. Si se puede utilizar árboles de decisión en algún contexto como por ejemplo para definir si un texto es spam o no, o tal vez, para determina el idioma del texto.

Preguntas con opciones múltiples *(indicar la o las respuesta(s) correcta(s))*:

7) Para una tarea de identificación de fraudes bancarios, utilizarían:
   - **a) un enfoque de aprendizaje supervisado**
   - **b) la métrica F-Score (ponderación de la Precisión y Recall)**
   - c) el cálculo del error cuadrático medio ("Mean Squared Error")
   - **d) el algoritmo de clasificación Gradient Descent**

8) Supongamos que están utilizando un algoritmo de mini-batch gradient descent para optimizar su modelo de regresión logística [1], con un tamaño de batch de 100. Sus datos de entrenamiento contienen 1.000.000 de ejemplos, y definen un número de épocas de 70. ¿Cuantás veces se van a actualizar los pesos de su modelo?

[1]https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/

   - **a) 700.000 veces**
   - b) 7.000.000.000 veces
   - c) 7.000 veces
   - d) 10.000 veces

9) ¿Cómo se puede gestionar un conjunto de datos desbalanceados?
   - **a) ponderando las clases en la función de perdida**
   - b) aumentando el volumen de datos de entrenamiento
   - **c) haciendo un sobremuestro de los datos de entrenamiento ("oversampling")**
   - **d) haciendo un submuestro de los datos de entrenamiento ("undesampling")**

10) Las técnicas "Bag-of-Words" estándares para construir representaciones vectoriales de documentos textuales tienen las limitaciones siguientes:
    - a) No permiten combinarse con algoritmos de Machine Learning (como Regresion Logistica o Árbol de decision)
    - **b) Pierden información sobre la relación entre las palabras de un mismo texto**
    - **c) Pierden información sobre la relación semántica entre las palabras**
    - d) Generan representaciones vectoriales con un pequeño número de características