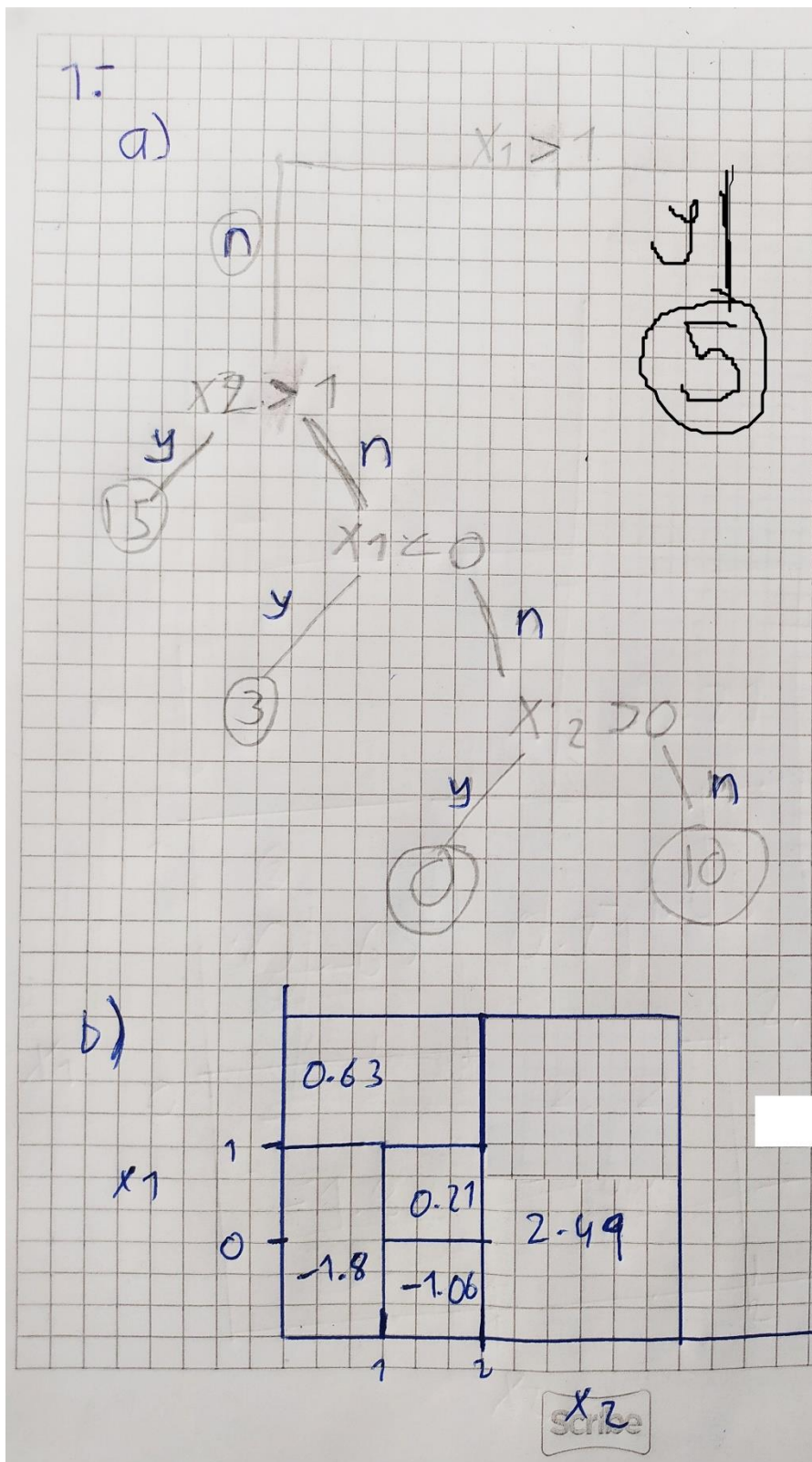
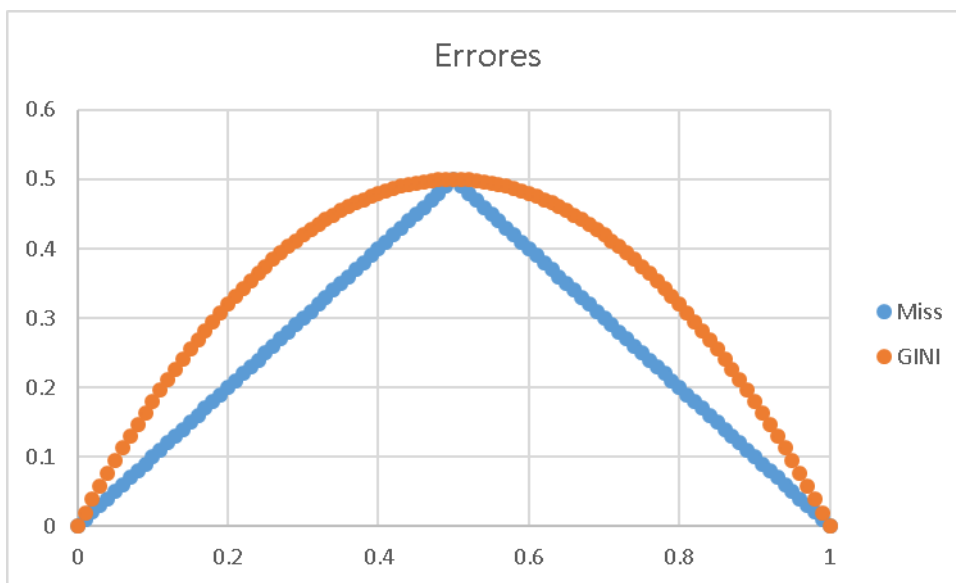


Pregunta 1



Pregunta 2



Pregunta 3

Hard Voting= 6/10 clasificado como positivo

Soft Voting= 45%

Pregunta 4

La variable más importante es X2

Pregunta 5

La característica de estas dos técnicas es que son clasificadas como de conjunto y se expresan en un grupo de árboles grandes como el operador los defina.

En el caso de boosting se comienza con una clasificación en los datos y se evalúa su desempeño así sucesivamente y comienza a tener en cuentas las clasificaciones anteriores, después de hacer esto vuelve a hacer una clasificación, pero ahora agregando peso para los casos más difíciles y tener una mejor clasificación, el proceso se puede repetir hasta tener un resultado deseado.

En el caso contrario el baggin el resultado de las clasificaciones se obtiene realizando un promedio de las respuestas de los datos

Pregunta 6

Este algoritmo entra en la clasificación de aprendizaje supervisado, puede ser utilizado para clasificación y regresión, en este algoritmo se crea un "bosque" con una n cantidad de "árboles" que están definidas por el usuario, estos árboles se pueden visualizar como un conjunto de reglas para ir subdividiendo el problema y llegar así a un caso más sencillo

Pregunta 7

```
[7] y_true = [A, A, B, A, B, A, A, B, B, A]
     y_pred = [A, B ,B, B ,B, B, A, A, B, A]
     confusion_matrix(y_true, y_pred)
```

```
↳ array([[3, 3],
         [1, 3]])
```

```
[11] from sklearn.metrics import precision_score, recall_score, f1_score
```

```
[9] precision_score(y_true, y_pred)
```

```
↳ 0.75
```

```
[10] recall_score(y_true, y_pred)
```

```
↳ 0.5
```

```
[12] f1_score(y_true, y_pred)
```

```
↳ 0.6
```

Pregunta 8

Kmeans trabaja fijando un centroide inicial y después busca las distancias entre los puntos, una vez que hace esto vuelve a fijar un nuevo centroide haciendo que la distancia entre el centro y los puntos aledaños se vaya minimizando cada vez más. Para este método se requiere fijar un numero inicial de clusters.

En cambio, DBSCAN funciona utilizando la densidad de los puntos que haya a su alrededor de esta manera es como va fijando sus “nucleos”. Para este método no se requiere fijar un numero inicial de clusters.

Pregunta 9

Matemáticamente deberían tender a ser mas bajas ya que la técnica dice que cada nodo se subdivide y minimiza los errores con cada paso, por lo que debería de ser menor. Así que matemáticamente debería ser siempre menor, pero dejémoslo en generalmente menor.

Pregunta 10

Los componentes son entradas, pesos, suma de los pesos, función de transferencia y una salida.

Pregunta 11

La ventaja de los algoritmos basados en arboles de decisiones es que no necesariamente los datos deberían estar normalizados y en experiencia personal no he notado diferencias al hacer predicciones con datos normalizados a no normalizarlos

Pregunta 12

Conjunto A= 0.185

Conjunto B=0.467