

# Cuestionario 3

---

## Aprendizaje Automático

David Gil Bautista

45925324M

## Cuestiones

1. Tanto "bagging" como validación-cruzada cuando se aplican sobre una muestra de datos nos permiten dar una estimación del error de un modelo ajustado a partir de dicha muestra. Enuncie las diferencias y semejanzas entre ambas técnicas. Diga cual de ellas considera que nos proporcionará una mejor estimación del error en cada caso concreto y por qué.

### Respuesta:

Tanto en "bagging" como en validación-cruzada seleccionamos varios conjuntos de datos, pero **con "bagging"** escogemos las muestras con reemplazamiento, es decir, **un elemento puede aparecer varias veces**. Sin embargo, **usando validación cruzada**, dividimos el set de datos y operamos con cada parte, por lo que **no se repiten datos en una misma división**.

Usar **bootstrap nos permite reducir la varianza** ya que al muestrear el set de datos varias veces obtendremos distintos resultados, al combinar dichos resultados reducimos la varianza. Con **la validación cruzada** usamos el conjunto de datos y para cada partición calculamos un resultado con el cual validamos la bondad de nuestro ajuste, lo que **nos permite estimar un error de salida** ( $E_{out}$ ).

2. Considere que dispone de un conjunto de datos linealmente separable. Recuerde que una vez establecido un orden sobre los datos, el algoritmo perceptron encuentra un hiperplano separador iterando sobre los datos y adaptando los pesos de acuerdo al algoritmo

### Algorithm 1 Perceptron

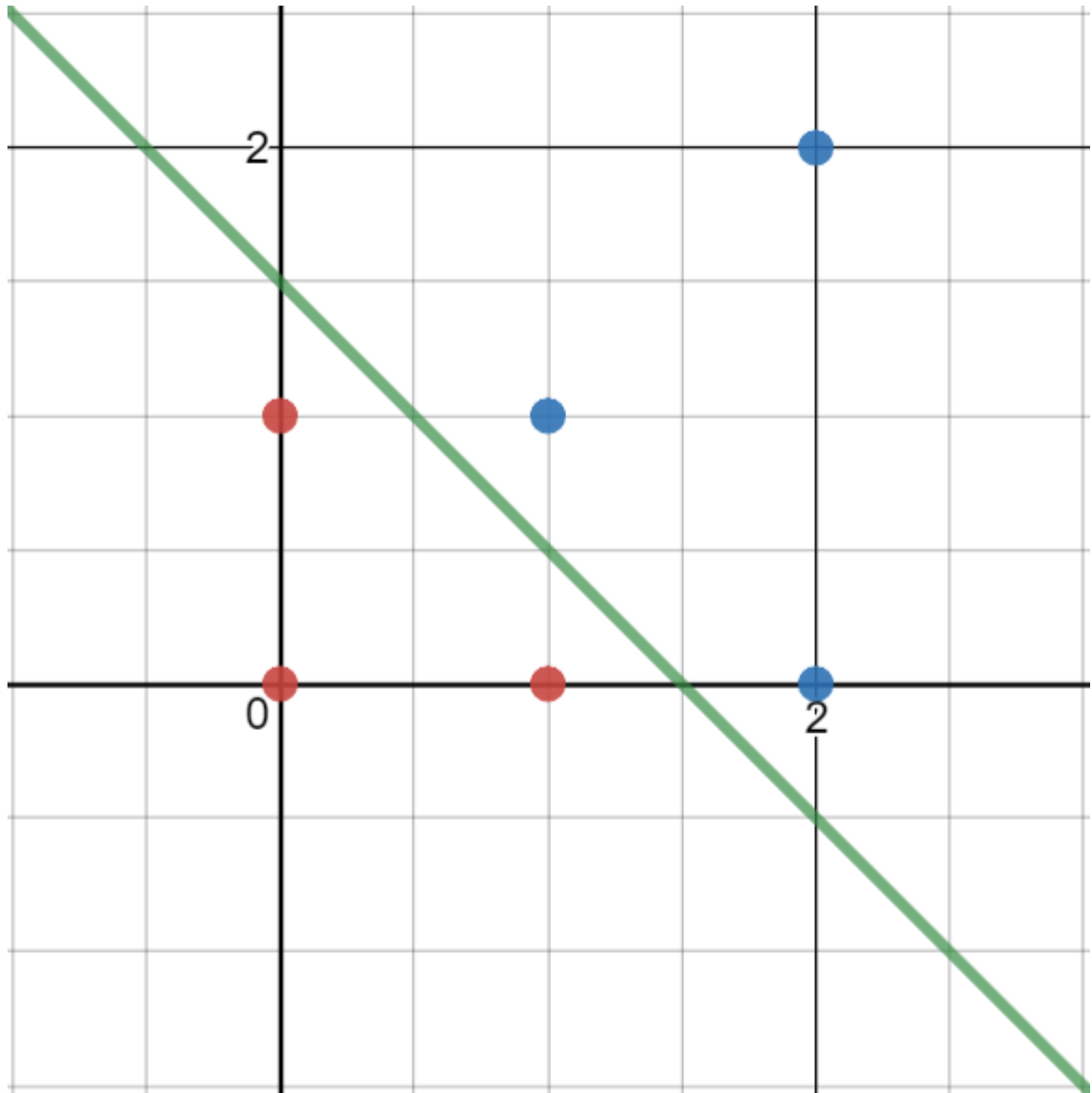
1: **Entradas:**  $(x_i, y_i), i = 1, \dots, n$ ,  $w = 0$ ,  $k = 0$  2: **repeat** 3:  $k \leftarrow (k + 1) \bmod n$  4: **if**  $\text{sign}(y_i) \neq \text{sign}(w^T x_i)$  **then** 5:  $w \leftarrow w + y_i x_i$  6: **end if** 7: **until** todos los puntos bien clasificados

Modificar este pseudo-código para adaptarlo a un algoritmo simple de SVM, considerando que en cada iteración adaptamos los pesos de acuerdo al caso peor clasificado de toda la muestra. Justificar adecuadamente/matematicamente el resultado, mostrando que al final del entrenamiento solo estaremos adaptando los vectores soporte.

### Respuesta:

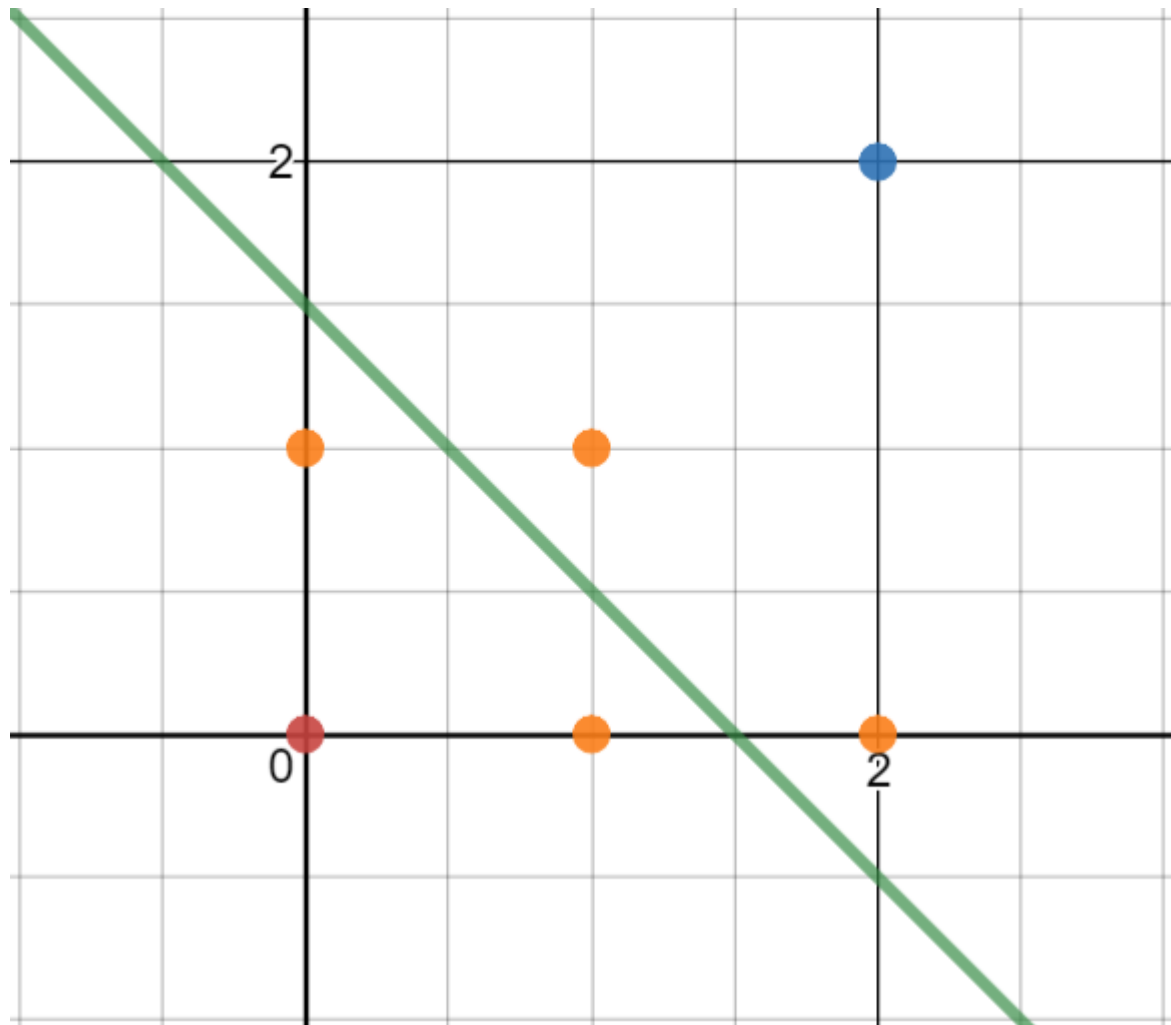
3. Considerar un modelo SVM y los siguientes datos de entrenamiento: Clase-1:  $\{(1,1),(2,2),(2,0)\}$ , Clase-2:  $\{(0,0),(1,0),(0,1)\}$

- Dibujar los puntos y construir por inspección el vector de pesos para el hiperplano óptimo y el margen óptimo.



Tras colocar los puntos podemos ver que el hiperplano que mejor los divide corresponde con la recta  $y = 1.5 - x$ , con la cual obtenemos un margen de  $\approx 0.71$  unidades

- ¿Cuáles son los vectores soporte?



Los puntos  $\{(0,1), (1,0)\}$  de la clase 2, y los puntos  $\{(2,0), (1,1)\}$  de la clase 1.

- Construir la solución en el espacio dual. Comparar la solución con la del apartado (a)

4. Una empresa está valorando cambiar su sistema de proceso de datos, para ello dispone de dos opciones, la primera es adquirir un nuevo sistema compuesto por dos sistemas idénticos al actual a 200.000 euros cada uno, y la segunda consiste en adquirir un nuevo sistema mucho mayor por 800.000 euros. Las ventas que la empresa estima que tendrá a lo largo de la vida útil de cualquiera de sus nuevos equipos es de 5.000.000 de euros en el caso de un mercado alcista, a lo que la empresa le asigna una probabilidad de que suceda del 30 %, en caso contrario, las ventas esperadas son de 3.500.000 euros. Construir el árbol de decisiones y decir que opción es la más ventajosa para la empresa.

**Respuesta:**



Con la opción 1 tendríamos una probabilidad de un 9% de obtener el máximo beneficio, con la segunda tendríamos una probabilidad de un 30%, aunque costando el doble. Por lo que para una empresa, el pagar más por tener menos riesgo a la hora de obtener beneficio podría ser crucial.

Con la primera opción, sin embargo, además de abaratar el coste del sistema, es más probable que se obtenga un poco más de beneficio, con un 30% de probabilidad se obtendrían 3.75M.

Ambas opciones son buenas, así que la elección sería de la empresa, arriesgar con menos presupuesto para intentar obtener el máximo beneficio o con un presupuesto aceptable optar a menos beneficio.

5. ¿Que algoritmos de aprendizaje no se afectan por la dimensionalidad del vector de características? Diga cuáles y por qué.

**Respuesta:**

6. Considere la siguiente aproximación al aprendizaje. Mirando los datos, parece que los datos son linealmente separables, por tanto decidimos usar un simple perceptron y obtenemos un error de entrenamiento cero con los pesos óptimos encontrados. Ahora deseamos obtener algunas conclusiones sobre generalización, por tanto miramos el valor  $d_{vc}$  de nuestro modelo y vemos que es  $d + 1$ . Usamos dicho valor de  $d_{vc}$  para obtener una cota del error de test. Argumente a favor o en contra de esta forma de proceder identificando los posibles fallos si los hubiera y en su caso cual hubiera sido la forma correcta de actuación.

**Respuesta:**

Totalmente en contra, no podemos visualizar los datos para escoger un algoritmo de aprendizaje automático (*data snooping*), por lo que todo el proceso es erróneo.

Una vez tenemos los datos probamos con varios algoritmos y podemos escoger aquel que nos proporcione una mejor solución, en vez de buscar aquel que nos de un error de 0 en el conjunto de entrenamiento.

7. Discuta pros y contras de los clasificadores SVM y Random Forest (RF). Considera que SVM por su construcción a través de un problema de optimización debería ser un mejor clasificador que RF. Justificar las respuestas.

**Respuesta:**

**Pros:** SVM ofrece un modelo que se ajusta de la mejor forma a la función objetivo, siempre llega al mínimo global. Es útil para datos linealmente separables y no linealmente separables. Funciona bien con set de datos de alta dimensión.

Random Forest crea un modelo con baja varianza y en el que al usar varias muestras de datos no sobreajusta los datos del train. Se puede observar cuáles son las características más importantes del set de datos.

**Contras:** SVM puede llegar a sobreajustar dependiendo del kernel usado y el parámetro de regularización, todo depende de escoger estos elementos correctamente. No funciona bien con muestras de datos muy grandes.

Random Forest proporciona un buen sesgo al podar ramas, lo cual implica que tengamos mayor incertidumbre.

8. ¿Cuál es a su criterio lo que permite a clasificadores como Random Forest basados en un conjunto de clasificadores simples aprender de forma más eficiente? ¿Cuáles son las mejoras que introduce frente a los clasificadores simples? ¿Es Random Forest óptimo en algún sentido? Justifique con precisión las contestaciones.

**Respuesta:**

Random Forest crea varios modelos con un conjunto de datos aleatorio usando un clasificador simple, la combinación de varios modelos simples crea un modelo final con baja varianza en el que tenemos en cuenta las características más importantes del set de datos.

Random Forest trabaja con pequeños conjuntos de características, lo cual evita el sobreajuste. Al usar "bagging" también permite que a la hora de combinar las soluciones de los árboles de decisiones tengamos un modelo que no presente varianza.

Sí, es óptimo. Random Forest es un algoritmo muy sencillo que te permite obtener un modelo en muy poco tiempo. Si quisiéramos obtener un modelo que nos garantizara encontrar la solución siempre obtendríamos un modelo con varias ramas en el que se dificultaría la comprensión del mismo.

9. En un experimento para determinar la distribución del tamaño de los peces en un lago, se decide echar una red para capturar una muestra representativa. Así se hace y se obtiene una muestra suficientemente grande de la que se pueden obtener conclusiones estadísticas sobre los peces del lago. Se obtiene la distribución de peces por tamaño y se entregan las conclusiones. Discuta si las conclusiones obtenidas servirán para el objetivo que se persigue e identifique si hay algo que lo impida.

**Respuesta:**

No basta con capturar una muestra representativa de un conjunto del cual no se conoce nada. Para que esto fuera correcto se deberían haber capturado todos los peces, y de ese conjunto separar una muestra para estudiarla.

10. Identifique dos razones de peso por las que el ajuste de un modelo de red neuronal a un conjunto de datos puede fallar o equivalentemente obtener resultados muy pobres. Justifique la importancia de las razones expuestas.

**Respuesta:**

Si escogemos unos parámetros erróneos puede que la red nunca llegue a converger. Una vez hemos establecido unos parámetros no podemos hacer un seguimiento de los cambios que realiza la red neuronal, es como una caja opaca a la que le damos unos datos, apretamos unos botones y esperamos obtener una salida sin saber lo que ocurre en el interior.

Las redes neuronales necesitan muchos más datos que cualquier otro algoritmo de aprendizaje automático. Al estar constantemente comprobando las relaciones entre las características y cambiando los pesos de estas, necesitará más datos para obtener un modelo fiable.