

## Trabajo.1: Cuestiones de Teoría

Fecha de entrega: 27 marzo 2017. Valor máximo: 9 puntos + Bonus

---

### NORMAS DE DESARROLLO Y ENTREGA DE TRABAJOS

Para este trabajo como para los demás es obligatorio presentar un informe escrito (hacerlo en pdf, MS Word). **Sin este informe se considera que el trabajo NO ha sido presentado.**

**Normas para el desarrollo de los Trabajos:** EL INCUMPLIMIENTO DE LAS NORMAS (\*) SIGNIFICA PERDIDA DE 2 PUNTOS POR CADA INCUMPLIMIENTO.

- En su informe de contestación debe incluir todas las preguntas en el orden y tal y como se les formula en este documento. (\*)
- Las contestaciones irán a continuación de cada pregunta, dejando en blanco las que no conteste. (\*)
- Todas las contestaciones deben ser justificadas con argumentos. Sin argumentos la pregunta se considera no contestada.
- Todas las justificaciones matemáticas deben contener todos y cada uno de los pasos de la misma. En caso de duda la contestación no se considerará válida.
- **Forma de entrega:** Subir el pdf a la web docente de CCIA.

# PREGUNTAS

Todas las preguntas tienen el mismo valor: 1 punto

1. Identificar, para cada una de las siguientes tareas, que tipo de aprendizaje es el adecuado (supervisado, no supervisado, por refuerzo) y los datos de aprendizaje que deberíamos usar en su caso. Si una tarea se ajusta a más de un tipo, explicar como y describir los datos para cada tipo.
  - a) Dada una colección de fotos de caras de personas de distintas razas establecer cuantas razas distintas hay representadas en la colección.
  - b) Clasificación automática de cartas por distrito postal
  - c) Decidir si un determinado índice del mercado de valores subirá o bajará dentro de un periodo de tiempo fijado.
  - d) Aprender un algoritmo que permita a un robot rodear un obstaculo.
2. ¿Cuales de los siguientes problemas son más adecuados para una aproximación por aprendizaje y cuales más adecuados para una aproximación por diseño? Justificar la decisión
  - a) Definir los grupos de de animales vertebrados en pajaros, mamíferos, reptiles, aves y anfibios.
  - b) Determinar si se debe aplicar una campaña de vacunación contra una enfermedad.
  - c) Determinar si un correo electrónico es de propaganda o no.
  - d) Determinar el estado de ánimo de una persona a partir de una foto de su cara.
3. Construir un problema de *aprendizaje desde datos* para un problema de clasificación de fruta en una explotación agraria que produce mangos, papayas y guayabas. Identificar y describir los elementos formales del problema  $\mathcal{X}, \mathcal{Y}, f$  de manera que puedan ser usados por un computador. ¿Considera que en este problema estamos ante un caso de etiquetas con ruido o sin ruido? Justificar las respuestas.
4. La regla de adaptación de los pesos del Perceptron ( $\mathbf{w}_{new} = \mathbf{w}_{old} + y\mathbf{x}$ ) tiene la interesante propiedad de que los mueve en la dirección adecuada para clasificar  $\mathbf{x}$  de forma correcta. Suponga el vector de pesos  $\mathbf{w}$  de un modelo y un dato  $\mathbf{x}(t)$  mal clasificado respecto de dicho modelo. Probar que la regla de adaptación de pesos siempre produce un movimiento en la dirección correcta para clasificar bien  $\mathbf{x}(t)$ .
5. La desigualdad de Hoeffding modificada nos da una forma de caracterizar el error de generalización con una cota probabilística

$$\mathbb{P}[|E_{out}(g) - E_{in}(g)| > \epsilon] \leq 2Me^{-2N\epsilon^2}$$

para cualquier  $\epsilon > 0$ . Si fijamos  $\epsilon = 0,05$  y queremos que la cota probabilística  $2Me^{-2N\epsilon^2}$  sea como máximo 0.03 ¿cual será el valor más pequeño de  $N$  que verifique estas condiciones si  $M = 1$ ?. Repetir para  $M = 10$  y para  $M = 100$

6. El jefe de investigación de una empresa con mucha experiencia en problemas de predicción de datos tras analizar los resultados de los muchos algoritmos de aprendizaje usados sobre todos los problemas en los que la empresa ha trabajado a lo largo de su muy dilatada existencia, decide que para facilitar el mantenimiento del código de la empresa van a seleccionar un único algoritmo y una única clase de funciones con la que aproximar todas las soluciones a sus problemas presentes y futuros. ¿ Considera que dicha decisión es correcta y beneficiará a la empresa? Argumentar la respuesta usando los resultados teóricos estudiados.

7. Para un conjunto  $\mathcal{H}$  con  $d_{VC} = 10$ , ¿que tamaño muestral se necesita (según la cota de generalización) para tener un 95 % de confianza de que el error de generalización sea como mucho 0.05?
8. Identificar de forma precisa las dos condiciones que garantizan que un problema de predicción puede ser aproximado por inducción desde una muestra de datos y una clase de funciones. Justificar la respuesta usando los resultados teóricos estudiados.
9. Considere que le dan una muestra de tamaño  $N$  de datos etiquetados  $\{-1, +1\}$  y le piden que encuentre la función que mejor ajuste dichos datos. Dado que desconoce la verdadera función  $f$ , discuta los pros y contras de aplicar los principios de inducción ERM y SRM para lograr el objetivo. Valore las consecuencias de aplicar cada uno de ellos.

## BONUS

Los BONUS solo serán tenidos en cuenta si en el cuestionario obligatorio se ha conseguido al menos un 75 % de los puntos totales. Hacer como máximo dos de ellos.

1. (2 puntos) Supongamos que tenemos un conjunto de datos  $\mathcal{D}$  de 25 ejemplos extraídos de una función desconocida  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , donde  $\mathcal{X} = \mathbb{R}$  e  $\mathcal{Y} = \{-1, +1\}$ . Para aprender  $f$  usamos un conjunto simple de hipótesis  $\mathcal{H} = \{h_1, h_2\}$  donde  $h_1$  es la función constante igual a  $+1$  y  $h_2$  la función constante igual a  $-1$ .

Consideramos dos algoritmos de aprendizaje, S(smart) y C(crazy). S elige la hipótesis que mejor ajusta los datos y C elige deliberadamente la otra hipótesis. Suponga que hay una distribución de probabilidad sobre  $\mathcal{X}$ , y sea  $P[f(x) = +1] = p$

- a) Si  $p = 0,9$  ¿Cual es la probabilidad de que S produzca una hipótesis mejor que C?
  - b) ¿Existe un valor de  $p$  para el cual es más probable que C produzca una hipótesis mejor que S?
2. (2 puntos) Consideremos el modelo de aprendizaje "M-intervalos" donde la clase de funciones  $H$  está formada por  $h : \mathbb{R} \rightarrow \{-1, +1\}$ , con  $h(x) = +1$  si el punto está dentro de uno de  $m$  intervalos arbitrariamente elegidos y  $-1$  en otro caso. Calcular la dimensión de Vapnik-Chervonenkis para esta clase de funciones.
  3. (1.5 puntos) Considere el enunciado del ejercicio 2 de la sección ERROR Y RUIDO de la relación apoyo.

- a) Si su algoritmo busca la hipótesis  $h$  que minimiza la suma de los valores absolutos de los errores de la muestra,

$$E_{in}(h) = \sum_{n=1}^N |h - y_n|$$

entonces mostrar que la estimación será la mediana de la muestra,  $h_{med}$  (cualquier valor que deje la mitad de la muestra a su derecha y la mitad a su izquierda)

- b) Suponga que  $y_N$  es modificado como  $y_N + \epsilon$ , donde  $\epsilon \rightarrow \infty$ . Obviamente el valor de  $y_N$  se convierte en un punto muy alejado de su valor original. ¿Como afecta esto a los estimadores dados por  $h_{mean}$  y  $h_{med}$ ?