

Aprendizaje automático

Cuestionario 1

Alejandro García Montoro
agarciamontoro@correo.ugr.es

28 de marzo de 2016

1. Ejercicios

Ejercicio 1. *Identificar, para cada una de las siguientes tareas, qué tipo de aprendizaje es el adecuado (supervisado, no supervisado, por refuerzo) y los datos de aprendizaje que deberíamos usar. Si una tarea se ajusta a más de un tipo, explicar cómo y describir los datos para cada tipo.*

- *Categorizar un grupo de animales vertebrados en pajaros, mamíferos, reptiles, aves y anfibios.*
- *Clasificación automática de cartas por distrito postal.*
- *Decidir si un determinado índice del mercado de valores subirá o bajará dentro de un periodo de tiempo fijado.*

Ejercicio 2. *¿Cuáles de los siguientes problemas son más adecuados para una aproximación por aprendizaje y cuáles más adecuados para una aproximación por diseño? Justificar la decisión.*

- *Determinar el ciclo óptimo para las luces de los semáforos en un cruce con mucho tráfico.*
- *Determinar los ingresos medios de una persona a partir de sus datos de nivel de educación, edad, experiencia y estatus social.*
- *Determinar si se debe aplicar una campaña de vacunación contra una enfermedad.*

Ejercicio 3. *Construir un problema de aprendizaje desde datos para un problema de selección de fruta en una explotación agraria (ver transparencias de clase). Identificar y describir cada uno de sus elementos formales. Justificar las decisiones.*

Ejercicio 4. Suponga un modelo PLA y un dato $x(t)$ mal clasificado respecto de dicho modelo. Probar que la regla de adaptación de pesos del PLA es un movimiento en la dirección correcta para clasificar bien $x(t)$.

Ejercicio 5. Considere el enunciado del ejercicio 2 de la sección FACTIBILIDAD DEL APRENDIZAJE de la relación apoyo.

- Si $p = 0,9$, ¿cuál es la probabilidad de que S produzca una hipótesis mejor que C ?
- ¿Existe un valor de p para el cual es más probable que C produzca una hipótesis mejor que S ?

Ejercicio 6. La desigualdad de Hoeffding modificada nos da una forma de caracterizar el error de generalización con una cota probabilística

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

para cualquier $\varepsilon > 0$. Si fijamos $\varepsilon = 0,05$ y queremos que la cota probabilística $2Me^{-2N^2\varepsilon}$ sea como máximo 0,03, ¿cuál será el valor más pequeño de N que verifique estas condiciones si $M = 1$? Repetir para $M = 10$ y para $M = 100$.

Ejercicio 7. Consideremos el modelo de aprendizaje « M -intervalos» donde $h: \mathbb{R} \rightarrow \{1, +1\}$, y $h(x) = +1$ si el punto está dentro de cualquiera de m intervalos arbitrariamente elegidos y 1 en otro caso. ¿Cuál es el más pequeño punto de ruptura para este conjunto de hipótesis?

Ejercicio 8. Suponga un conjunto de k^* puntos x_1, x_2, \dots, x_{k^*} sobre los cuales la clase \mathcal{H} implementa $< 2^{k^*}$ dicotomías. ¿Cuáles de las siguientes afirmaciones son correctas?

- k^* es un punto de ruptura.
- k^* no es un punto de ruptura.
- Todos los puntos de ruptura son estrictamente mayores que k^* .
- Todos los puntos de ruptura son menores o iguales a k^* .
- No conocemos nada acerca del punto de ruptura.

Ejercicio 9. Para todo conjunto de k^* puntos, \mathcal{H} implementa $< 2^{k^*}$ dicotomías. ¿Cuáles de las siguientes afirmaciones son correctas?

- k^* es un punto de ruptura.
- k^* no es un punto de ruptura.
- Todos los $k \geq k^*$ son puntos de ruptura.

- Todos los $k < k^*$ son puntos de ruptura.
- No conocemos nada acerca del punto de ruptura.

Ejercicio 10. Si queremos mostrar que k^* es un punto de ruptura, ¿cuáles de las siguientes afirmaciones nos servirían para ello?:

- Mostrar que existe un conjunto de k^* puntos x_1, x_2, \dots, x_{k^*} que \mathcal{H} puede separar —shatter—.
- Mostrar que \mathcal{H} puede separar cualquier conjunto de k^* puntos.
- Mostrar un conjunto de k^* puntos x_1, x_2, \dots, x_{k^*} que \mathcal{H} no puede separar.
- Mostrar que \mathcal{H} no puede separar ningún conjunto de k^* puntos.
- Mostrar que $m_{\mathcal{H}}(k) = 2^{k^*}$

Ejercicio 11. Para un conjunto \mathcal{H} con $d_{VC} = 10$, ¿qué 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?

Ejercicio 12. Consideremos un escenario de aprendizaje simple. Supongamos que la dimensión de entrada es uno. Supongamos que la variable de entrada x está uniformemente distribuida en el intervalo $[1, 1]$ y el conjunto de datos consiste en 2 puntos $\{x_1, x_2\}$ y que la función objetivo es $f(x) = x^2$. Por tanto el conjunto de datos completo es $\mathcal{D} = \{(x_1, x_1^2), (x_2, x_2^2)\}$. El algoritmo de aprendizaje devuelve la línea que ajusta estos dos puntos como g ; es decir, H consiste en funciones de la forma $h(x) = ax + b$.

- Dar una expresión analítica para la función promedio $\bar{g}(x)$.
- Calcular analíticamente los valores de E_{out} , bias y var.

2. Bonus

Bonus 1. Considere el enunciado del ejercicio 2 de la sección ERROR Y RUIDO de la relación de apoyo.

- 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—.

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

Bonus 2. Considere el ejercicio 12.

- Describir un experimento que podamos ejecutar para determinar — numéricamente— $\bar{g}(x)$, E_{out} , bias y var.
- Ejecutar el experimento y dar los resultados. Comparar E_{out} con bias + var. Dibujar en unos mismos ejes $\bar{g}(x)$, E_{out} y $f(x)$.