Aprendizaje Automático (2015-2016)
Grado en Ingeniería Informática
Universidad de Granada

Cuestionario 1

Laura Tirado López

2 de abril de 2016

Índice

11. Para un conjunto 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 $0{,}05$? 10

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 es conjunto de datos completo es $D = (x_1, x_1^2), (x_2, x_2^2)$. El algoritmo de aprendizaje devuelve la línea que ajusta estos dos puntos como g (i.e. H consiste en funciones de la forma $h(x) = ax + b$). 12

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 deberiamos usar. Si una tarea se ajusta a más de un tipo, explicar como y describir los datos para cada tipo.

1.1. Categorizar un grupo de animales vertebrados en pájaros, mamíferos, reptiles, aves y anfibios.

Para esta tarea, el aprendizaje adecuado sería aprendizaje supervisado dado que se trata de un problema de predicción. Para este caso, utilizaríamos una clasificación multietiqueta. Los datos de aprendizaje a usar debería de ser una muestra formada por los animales y su etiqueta correspondiente. Para reprensar a los animales podríamos usar un vector binario de características como pelo, escamas, branquias, etc.

1.2. Clasificación automática de cartas por distrito postal.

Para esta tarea podría utilizarse tanto aprendizaje supervisado como no supervisado. Al igual que el problema anterior, se trata de un problema de predicción por lo que podría utilizarse aprendizaje supervisado. A su vez, podemos utilizar aprendizaje no supervisado porque al categorizar por distrito postal, categorizamos según se agrupan los distritos postales. Para el aprendizaje supervisado, necesitaríamos los datos y su correspondiente etiqueta. Los datos podrían estar reprensentados por el código postal y la etiqueta sería multietiqueta. En el caso de aprendizaje no supervisado, sólo necesitaríamos los datos sin etiquetas dado que intentamos averiguar el comportamiento de los datos, es decir, cómo se agrupan en distritos postales.

1.3. Decidir si un determinado índice del mercado de valores subirá o bajará dentro de un periodo de tiempo fijado.

Para esta tarea, la mejor opción sería aprendizaje por refuerzo dado que se trata de decidir el comportamiento de una acción (la subida o bajada del índice del mercado de valores).

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.

2.1. Determinar el ciclo óptimo para las luces de los semáforos en un cruce con mucho tráfico.

Este problema es más adecuado para una aproximación por diseño. A partir de las medidas de tiempos, velocidades, número de coches, etc, podría hacerse un modelo que calculase la solución óptima con alguna metaheurística de búsqueda por ejemplo.

2.2. Determinar los ingresos medios de una persona a partir de sus datos de nivel de educación, edad, experiencia y estatus social.

Este problema es más adecuado para una aproximación por aprendizaje dado que no tenemos un modelo claro entre las entradas y las salidas.

2.3. Determinar si se debe aplicar una campaña de vacunación contra una enfermedad.

Este problema es más adecuado para un aproximación por diseño. A partir de situaciones anteriores podemos tener definidos unos umbrales o unos datos probabilísticos a partir de los cuales decidir si la enfermedad necesita una campaña de vacunación.

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.

Los elementos formales de nuestro problema sería:

- Conjunto de características: características de la fruta (color, tamaño, peso, etc).
- Conjunto de etiquetas: utilizaríamos un etiquetado binario (+1 si la fruta es seleccionada y -1 en caso contrario).
- Función de etiquetado: desconocida

- Datos de entrenamiento: conjunto de datos formados por parejas de los conjuntos de características y etiquetas.
- Conjunto de soluciones H: conjunto de funciones para aproximar a la función de etiquetado.

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 $x(t)$.

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

5.1. a) Si $p = 0,9$ ¿Cual es la probabilidad de que S produzca una hipótesis mejor que C?

La probabilidad de que S acierte sería $0,9$ y la de C sería $0,1$. Por tanto, la probabilidad de que S sea mejor que C sería del $80\,\%$.

5.2. b) ¿Existe un valor de p para el cual es más probable que C produzca una hipótesis mejor que S?

No existe un valor para el cual es más probable que C produzca una hipótesis mejor que S porque S siempre coge la mejor hipótesis; por lo que como mucho ambos tendrían la misma probabilidad que seria $p = 0,5$ en cuyo caso la probabilidad de que uno fuese mejor que el otro sería $p = 0$.

6. La desigualdad de Hoeffding modificada nos da una forma de caracterizar el error de generalización con una cota probabilística $P[|E_{out}(g) - E_{in}(g)| > t] \leq 2Me^{-2N\epsilon^2}$ 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 $M = 100$.

$$2Me^{-2N0,05^2} \leq 0,03$$

$$ln(2M) + ln(e^{-0,005N}) \leq ln(0,03)$$

$$ln(2M) + -0,005Nln(e) \leq ln(0,03)$$

$$0,005Nln(e)\ leqln(0,03) - ln(2M)$$

$$N \leq \frac{ln(0,03) - ln(2M)}{-0,005ln(e)}$$

$$N \leq \frac{ln(0,03) - ln(2M)}{-0,005}$$

$$N \leq \frac{-3,5 - ln(2M)}{-0,005}$$

Para $M = 1$:

$$N \leq \frac{-3,5 - ln(2 \cdot 1)}{-0,005}$$

$$N \leq \frac{-3,5 - 0,693}{-0,005}$$

$$N \leq \frac{-4,193}{-0,005}$$

$$N \leq 838,6$$

$$N = 838$$

Para $M = 10$:

$$N \leq \frac{-3,5 - ln(2 \cdot 10)}{-0,005}$$

$$N \leq \frac{-3,5 - 2,99}{-0,005}$$
$$N \leq \frac{-6,49}{-0,005}$$
$$N \leq 1298$$
$$N = 1298$$

Para $M = 100$:

$$N \leq \frac{-3,5 - ln(2 \cdot 100)}{-0,005}$$
$$N \leq \frac{-3,5 - 5,29}{-0,005}$$
$$N \leq \frac{-8,79}{-0,005}$$
$$N \leq 1758$$
$$N = 1758$$

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

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

8.1. k^* es un punto de ruptura

La afirmación es correcta.

8.2. k^* no es un punto de ruptura

La afirmación es incorrecta dado que para $k^* = 3$ se cumple $m_H(k^*) < 2^{k^*}$, por lo que k^* es un punto de ruptura.

8.3. todos los puntos de ruptura son estrictamente mayores que k^*

La afirmación es incorrecta dado que k^* también es un punto de ruptura.

8.4. todos los puntos de ruptura son menores o iguales a k^*

La afirmación es incorrecta porque el punto de ruptura indica a partir de que número de puntos la clase H no puede separar los puntos.

8.5. no conocemos nada acerca del punto de ruptura

La afirmación es incorrecta por la misma razón que la segunda afirmación.

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

9.1. k^* es un punto de ruptura

La afirmación es correcta.

9.2. k^* no es un punto de ruptura

La afirmación es incorrecta dado que para $k^* = 3$ se cumple $m_H(k^*) < 2^{k^*}$, por lo que k^* es un punto de ruptura.

9.3. todos los $k \geq k^*$ son puntos de ruptura

La afirmación es correcta. todos los $k < k^*$ son puntos de ruptura La afirmación es incorrecta porque el punto de ruptura indica a partir de que número de puntos la clase H no puede separar los puntos.

9.4. no conocemos nada acerca del punto de ruptura

La afirmación es incorrecta por la misma razón que la segunda afirmación.

10. Si queremos mostrar que k^* es un punto de ruptura cuales de las siguientes afirmaciones nos servirían para ello:

10.1. Mostrar que existe un conjunto de k^* puntos $x_1, ..., x_{k^*}$ que H puede separar ("shatter").

Esta afirmación no nos serviría porque si H puede separar un conjunto de k^* puntos, $k*$ no sería un punto de ruptura.

10.2. Mostrar que H puede separar cualquier conjunto de k^* puntos.

Esta afirmación no nos serviría por la misma razón que la anterior.

10.3. Mostrar un conjunto de k^* puntos $x_1, ..., x_{k^*}$ que H no puede separar.

Esta afirmación nos serviría porque un punto de ruptura indica el tamaño del conjunto de punto que H no puede separar.

10.4. Mostrar que H no puede separar ningún conjunto de k^* puntos.

Esta afirmación nos serviría por la misma razón que la anterior.

10.5. Mostrar que $m_H(k)2^{k^*}$

. Esta afirmación no nos serviría dado que la ecuación para comprobar si k^* es un punto de ruptura es $m_H(k^*) < 2^{k^*}$.

11. Para un conjunto 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?

$$0,05 = \sqrt{\frac{8}{N} ln \frac{4((2N)^{10} + 1)}{0,05}}$$

$$0,05^2 = \sqrt{\frac{8}{N} ln \frac{4((2N)^{10}+1)}{0,05}}^2$$

$$0,0025 = \frac{8}{N} ln \frac{4((2N)^{10}+1)}{0,05}$$

$$0,0025N = 8ln\frac{2^{12}N^{10}+4}{0,05}$$

$$0,0025N = 8ln(2^{12}N^{10}+4) - ln0,05$$

$$N = \frac{8ln(2^{12}N^{10}+4) - ln0,05}{0,0025}$$

Como podemos ver no podemos despejar N, por lo que aproximamos el valor de N con Wolfram Alpha, un software de cáclulo mateḿático. El resultado aproximado que nos da es $N = 452957$.

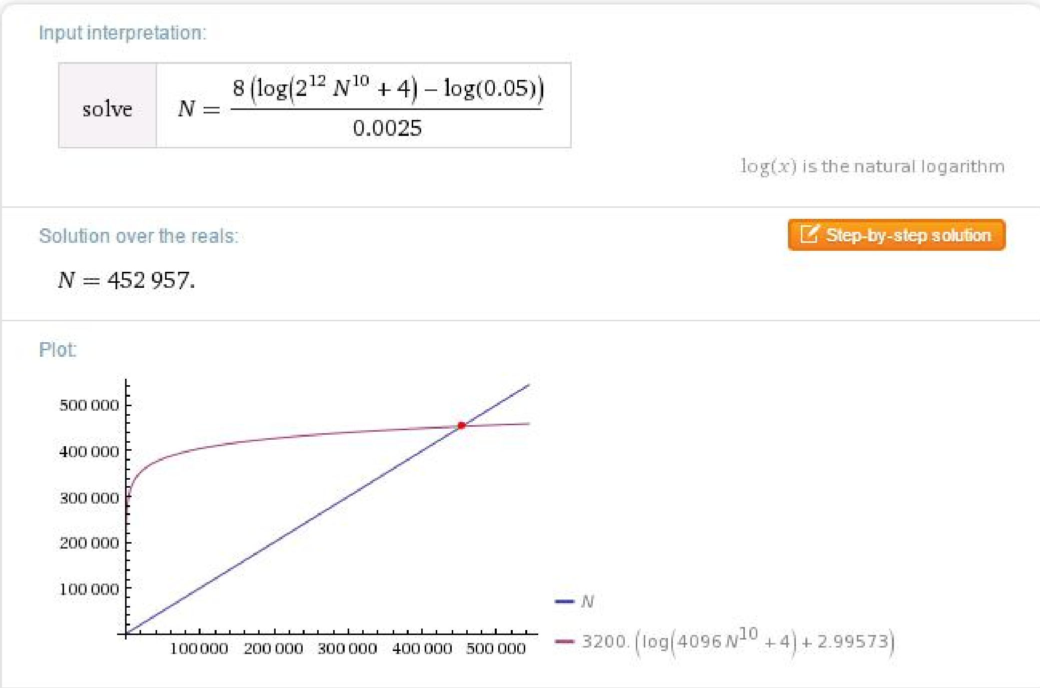

Figura 11.1: Aproximación del valor de N con Wolfram Alpha

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 es conjunto de datos completo es $D = (x_1, x_1^2), (x_2, x_2^2)$. El algoritmo de aprendizaje devuelve la línea que ajusta estos dos puntos como g (i.e. H consiste en funciones de la forma $h(x) = ax + b$).

12.1. Dar una expresión analítica para la función promedio $g(x)$.

La expresión general de la función promedio es:

$$\bar{g}(x) = \frac{1}{K}(g^1(x) + g^2(x) + ... + g^K(x))$$

Como todas las funciones tienen la forma $ax + b$, sustituimos:

$$\bar{g}(x) = \frac{1}{K}(a_1(x) + b_1 + a_2(x) + b_2 + ... + a_K(x) + b_K)$$

La expresión analítica sería:

$$\bar{g}(x) = \frac{x \sum_{i=1}^{K} a_i + \sum_{i=1}^{K} b_i}{K}$$

12.2. Calcular analíticamente los valores de E_{out}, $bias$ y var.