MÁSTER UNIVERSITARIO EN LÓGICA, COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

Aprendizaje Automático
Apellidos: Lorenz Vieta
Nombre: Germán

1. Ejercicio

Sea X un universo y D un conjunto de entrenamiento sobre X. Sean $E^+ = \{x \in X \mid (x,1) \in D\}$ y $E^- = \{x \in X \mid (x,0) \in D\}$. Sea H el conjunto de hipótesis que contiene a todas las hipótesis sobre X. Sea VS el espacio de versiones para (D,H). Sea $x_0 \in X/x_0 \notin E^+ \cup E^-$ y sea $h \in VS/h(x_0) = 0$. Demostrar $h^{`} \in VS/h^{`}(x_0) = 1$.

Considerando la hipótesis $h^{`} \in VS$ definida como: $h^{`}(x) = \begin{cases} 1 & \text{si } x = 1 \\ h(x) & \text{si } x \in X - \{x_0\} \end{cases}$

Por definición $h^{`}(x) = 1$.Veamos que $h^{`} \in VS$.

Como H es el conjunto de todas las hipótesis, $h^{`} \in H$. Además $h^{`}(x) = h(x)\ \forall x \neq x_0$. En particular, $h^{`}(x) = h(x)\ \forall x \in E^+ \cup E^-$. Luego, si $h \in VS \Rightarrow h^{`} \in VS$

Por lo tanto, $\exists h^{`} \in VS/h^{`}(x_0) = 1$

2. Ejercicio

Sea U un universo finito y $C = 2^U$ el conjunto de los objetivos. Sea H un conjunto de hipótesis sobre U y L un algoritmo de aprendizaje tal que su dominio es $\cup_{c \in C} \cup_{m \geq 1} S(m,c)$. Demostrar que si $H \neq 2^U$, entonces L no es consistente.

Sea el algoritmo de aprendizaje $L = \cup_{c \in C} \cup_{m \geq 1} S(m,c) \Rightarrow H \neq 2^U$. Consideremos el vector $s \in Dom(L)$ con $L(s) = h \notin 2^U$ con $s \equiv ((u_1,b_1),\ldots,(u_m,b_m)), u_i \in U, b_i \equiv \{0,1\}$.

L es consistente si $\forall s \in Dom(L), L(s) = h$ es consistente con s. Ademas, $L(s) = h$ es consistente con s si $\forall i \in \{1,\ldots,m\}$ se tiene que $h(u_i) = b_i$. Pero, por definicion, $H \neq 2^U$, i.e., $\exists h \in H \neq 2^U, \exists i \in \{1,\ldots,m\}/h(u_i) \neq b_i$.

Por lo tanto, L no es consistente.

3. Ejercicio

Sea $D = \{\langle x_1, c(x_1)\rangle, \ldots, \langle x_n, c(x_n)\rangle\}$ un conjunto de entrenamiento para un concepto C y sea H un conjunto de hipótesis. Demostrar que el resultado de aplicar el algoritmo de **Eliminación de Candidatos** es el mismo para cualquier ordenación de los elementos de D

Sean dos ordenadores del conjunto D y sean G^1, S^1 y G^2, S^2 las cotas generales obtenidas por eliminación de candidatos para cada una de las ordenaciones.

Supongamos, por reducción al absurdo, que $G^1 \neq G^2$ y $S^1 \neq S^2$. Si $G^1 \neq G^2$, entonces $\exists$ hipótesis $h \in H/h \in G^1$ y $h \notin G^2$

Si $h \in G^1$, h es consistente con $\langle x_1, c(x_1)\rangle\ i = 1,\ldots,n$ cuando $c(x_i) = 1$
Si $h \notin G^1$, h no es consistente con $\langle x_1, c(x_1)\rangle$ para algún $i = 1\ldots,n$ cuando $c(x_i) = 1$

Cuando $c(x_i) = 0$, se eliminaran de las cotas generales G^1, G^2 las hipótesis "menos"generales, y se incluirán las especializaciones minimales que sean consistentes con x_i, siguiendo el razonamiento anterior se tiene:

Si $G^1 \neq G^2 \Rightarrow h$ consistente con $\langle x_1, c(x_1)\rangle \forall i = 1, \ldots, n$ y h inconsistente con $\langle x_1, c(x_1)\rangle$ para algún $i \in \{1, \ldots, n\}$.

Esto es una contradicción, luego $G^1 = G^2$. De la misma forma se puede probar para la cota especifica S.

Si $c(x_i) = 1, S_1$ y S_2 estarán formadas por generalizaciones minimales que sean consistentes con x_i y se eliminaran las menos especificas.

Si $S_1 \neq S_2 \Rightarrow \exists h \in S^1$ consistente con $\langle x_1, c(x_1)\rangle \forall i = 1, \ldots, n$ y $h \notin S^2$ inconsistente con $\langle x_1, c(x_1)\rangle$ para algún $i \in \{1, \ldots, n\}$.

Así, $S^1 = S^2$. Luego las cotas generales y especificas son las mismas para ambas ordenaciones.

Por lo tanto, el resultado obtenido por eliminación de candidatos no depende de la ordenación.

4. Ejercicio

Aplica los algoritmos de *aprendizaje por enumeración* y *Find-S* para los siguientes problemas de aprendizaje:

4.1. Problema

$X = \mathbb{R}^2$
$H = \{h_n : X \Rightarrow \{0, 1\} \mid n \in \mathbb{N} \wedge h_n((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq n^2\}$
$s = \{\langle(1, 1), 1\rangle, \langle(3, 4), 1\rangle, \langle(2, 2), 1\rangle, \langle(4, 7), 0\rangle, \}$

Aprendizaje por enumeración: Recorrer el conjunto de hipótesis buscando $n/h_n(x_i) = c(x_i) \forall i = 1, 2, 3, 4$.

- $h_1 : h_1((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq 1$
  $\langle(1, 1), 1\rangle : h_1((1, 1)) \neq 1 = c((1, 1))$
- $h_2 : h_2((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq 4$
  $\langle(1, 1), 1\rangle : h_2((1, 1)) = 1 = c((1, 1))$
  $\langle(3, 4), 1\rangle : h_2((3, 4)) \neq 1 = c((3, 4))$
- $h_3 : h_3((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq 9$
  $\langle(1, 1), 1\rangle : h_3((1, 1)) = 1 = c((1, 1))$
  $\langle(3, 4), 1\rangle : h_3((3, 4)) \neq 1 = c((3, 4))$
- $h_4 : h_4((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq 16$
  $\langle(1, 1), 1\rangle : h_4((1, 1)) = 1 = c((1, 1))$
  $\langle(3, 4), 1\rangle : h_4((3, 4)) \neq 1 = c((3, 4))$
- $h_5 : h_5((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq 25$
  $\langle(1, 1), 1\rangle : h_5((1, 1)) = 1 = c((1, 1))$
  $\langle(3, 4), 1\rangle : h_5((3, 4)) = 1 = c((3, 4))$
  $\langle(3, 4), 1\rangle : h_5((2, 2)) = 1 = c((2, 2))$
  $\langle(4, 7), 1\rangle : h_5((4, 7)) = 0 = c((4, 7))$
  idem para h_6, h_7, h_8
- $h_9 : h_9((x, y)) = 1 \Leftrightarrow x^2 + y^2 \leq 89$
  $\langle(1, 1), 1\rangle : h_9((1, 1)) = 1 = c((1, 1))$
  $\langle(3, 4), 1\rangle : h_9((3, 4)) = 1 = c((1, 1))$
  $\langle(3, 4), 1\rangle : h_9((2, 2)) = 1 = c((2, 2))$
  $\langle(4, 7), 1\rangle : h_9((4, 7)) = 1 \neq c((4, 7))$

Por lo tanto, $H = \{h_5, h_6, h_7, h_8\}$

Find-S: Partiendo de la hipótesis h_n más específica de H, para cada ejemplo positivo si $h_n(x_i) = 1$, no hacer nada; en caso contrario reemplazar h_n por una menor generalización h_n con $h_n(x_i) = 1$. Los ejemplos negativos se ignoran.

- Paso 0: h_0 es la hipótesis más específica de H
  $H = \{h_0 : X \Rightarrow \{0,1\} \mid h_0((x,y)) = 1 \Leftrightarrow x^2 + y^2 \leq 0\}$

- Paso 1: $\langle(1,1),1\rangle$
  $h_0((1,1)) = 0 \Rightarrow$ se reemplaza h_0 por h_2 ya que h_1 tampoco lo cumple
  $h_2((1,1)) = 1$

- Paso 2: $\langle(3,4),1\rangle$
  $h_2((3,4)) = 0 \Rightarrow$ se reemplaza h_2 por h_5 ya que h_3 y h_4 tampoco lo cumplen

- Paso 3: $\langle(2,2),1\rangle$
  $h_5((2,2)) = 1 \Rightarrow$ se mantiene h_5

- Paso 4: $\langle(4,7),0\rangle$
  Al ser un ejemplo negativo se ignora.

En este caso la hipótesis de máxima especificidad consistente con todos los ejemplos positivos es h_5:

$$\text{Por lo tanto, } H = \{h_5 : X \Rightarrow \{0,1\} \mid h_5((x,y)) = 1 \Leftrightarrow x^2 + y^2 \leq 25\}$$

4.2. Problema

$X = \mathbb{R}^2$
$H = \{h_n \mid n \in \mathbb{N}\}$ con $h_0 = \emptyset$ y si $n \geq 1$, entonces $h_n = \{(x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, n = \frac{b(b-1)}{2} + a + 1\}$
$s = \{\langle(0,0),0\rangle, \langle(3,4),1\rangle, \langle(2,2),1\rangle\}$

Aprendizaje por enumeración:

- $h_0 : h_0 = \emptyset$

- $h_1 : h_1 = \{(x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, 1 = \frac{b(b-1)}{2} + a + 1\}$
  $\langle(0,0),0\rangle : a \leq 0 < b$. Como $a \in \mathbb{N}, h_1((0,0)) = 0 = c(0,0)$
  $\langle(3,4),1\rangle : a \leq 3 < b, 1 = \frac{b(b-1)}{2} + 1 + 1 \Rightarrow \frac{b(b-1)}{2} + a = 0$

  - Para $a = 1 : \frac{b(b-1)}{2} + 1 = 0 \Rightarrow b^2 - b + 2 = 0 \Rightarrow \nexists b \in \mathbb{N}/b^2 - b + 2 = 0$
  - Idem para $a = 2$ y $a = 3$
  - $h_1((3,4)) = 0 \neq c(3,4)$

- $h_2 : h_2 = \{(x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, 2 = \frac{b(b-1)}{2} + a + 1\}$
  $\langle(0,0),0\rangle : a \leq 0 < b$. Como $a \in \mathbb{N}, h_1((0,0)) = 0 = c(0,0)$
  $\langle(3,4),1\rangle : a \leq 3 < b, 2 = \frac{b(b-1)}{2} + 1 + 1 \Rightarrow \frac{b(b-1)}{2} + a = 0$

  - Para $a = 1 : \frac{b(b-1)}{2} = 0 \Rightarrow b = 0 \notin \mathbb{N}, b = 1 \ngtr 3$
  - Para $a = 2$ y $a = 3, \nexists b \in \mathbb{N}/\frac{b(b-1)}{2} + a - 1 = 0$
  - $h_2((3,4)) = 0 \neq c(3,4)$

- Buscamos una generalización: a $\leq x < b$ con x $\in \{2,3\}$, ya que siempre tendremos que se cumple para $\langle(0,0),0\rangle$ porque $0 \notin \mathbb{N}$. Tenemos: $a \leq 3 < b \Rightarrow a \in \{1,2,3\}$
  $n = \frac{b(b-1)}{2} + a + 1 \Rightarrow b^2 - b + 2a + 2 - 2n = 0 \Rightarrow b = \frac{1 \pm \sqrt{1-4(2a+2-2n)}}{2}$
  Para que b tenga solución, $1 - 4(2a + 2 - 2n) \geq 0 \Leftrightarrow n \geq \frac{7+8a}{8}$.

  - Para a $= 1$: $b = \frac{1\pm\sqrt{1-4(2+2-2n)}}{2} = \frac{1\pm\sqrt{-15+8n)}}{2}$. Como $b \in \mathbb{N}$ y $b > 3 : \frac{1\pm\sqrt{-15+8n)}}{2} > n > 5$. Por lo tanto, para existir solución $n > 5$ y $\frac{1\pm\sqrt{-15+8n)}}{2} \in \mathbb{N}$.
  - Para a $= 2$: $b = \frac{1\pm\sqrt{1-4(4+2-2n)}}{2} = \frac{1\pm\sqrt{-23+8n)}}{2}$. Como $b \in \mathbb{N}$ y $b > 3 : \frac{1\pm\sqrt{-23+8n)}}{2} > n > 6$. Por lo tanto, para existir solución $n > 6$ y $\frac{1\pm\sqrt{-23+8n)}}{2} \in \mathbb{N}$.
  - Para a $= 3$: $b = \frac{1\pm\sqrt{1-4(6+2-2n)}}{2} = \frac{1\pm\sqrt{-31+8n)}}{2}$. Como $b \in \mathbb{N}$ y $b > 3 : \frac{1\pm\sqrt{-31+8n)}}{2} > n > 7$. Por lo tanto, para existir solución $n > 7$ y $\frac{1\pm\sqrt{-31+8n)}}{2} \in \mathbb{N}$.

- $h_n((x,y)) = 1$ si $\exists(a,b) \in \mathbb{N}$ con $a \leq x < b, n = \frac{b(b-1)}{2} + a + 1$, luego:

Por lo tanto, $H = \{h_n \mid n \in \mathbb{N}, n > 5, \frac{1+\sqrt{-15+8n}}{2} \in \mathbb{N}\} \cup \{h_n \mid n \in \mathbb{N}, n > 6, \frac{1+\sqrt{-23+8n}}{2} \in \mathbb{N}\} \cup \{h_n \mid n \in \mathbb{N}, n > 7, \frac{1+\sqrt{-31+8n}}{2} \in \mathbb{N}\}$

Find-S:

- Paso 0: h_0 es la hipótesis más especifica de H.
- Paso 1: $\langle(0,0),0\rangle$. Al ser un ejemplo negativo se ignora.
- Paso 2: $\langle(3,4),1\rangle$.Usare los resultados obtenidos en el algoritmo de aprendizaje por enumeración para facilitar el cálculo. Para que haya solución, n debe ser, al menos, mayor que 5.
  - $h_6 = \{(x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, 6 = \frac{b(b-1)}{2} + a + 1\}$
    - Para $a = 1 : a \leq 3 < b, 6 = \frac{b(b-1)}{2} + 1 + 1 \Rightarrow b^2 + b - 4 = 0 \Rightarrow b = \frac{1 \pm \sqrt{33}}{2} \notin \mathbb{N}$.
    - Igual para $a = 2$ y $a = 3$.
  - $h_7 = \{(x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, 7 = \frac{b(b-1)}{2} + a + 1\}$
    - Para $a = 1 : a \leq 3 < b, 7 = \frac{b(b-1)}{2} + 1 + 1 \Rightarrow b^2 + b - 10 = 0 \Rightarrow b = \frac{1 \pm \sqrt{41}}{2} \notin \mathbb{N}$.
    - Igual para $a = 2$ y $a = 3$.
  - $h_8 = \{(x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, 8 = \frac{b(b-1)}{2} + a + 1\}$
    - Para $a = 1 : a \leq 3 < b, 8 = \frac{b(b-1)}{2} + 1 + 1 \Rightarrow b^2 + b - 12 = 0 \Rightarrow b = \frac{1 \pm \sqrt{49}}{2} \Rightarrow = 4 \in \mathbb{N}$ y $4 > 3$. Luego, $\exists (a,b) = (1,4) \in \mathbb{N} / a \leq 3 < b, n = \frac{b(b-1)}{2} + a + 1$.
  - $h_n((3,4)) = 0, n = 1, \ldots, 7 \Rightarrow$ se reemplaza la hipótesis por h_8 cumpliendo que $h_8((3,4)) = 1 = c(3,4)$.
- Paso 3: $\langle(2,2),1\rangle$
  $h_8((2,2)) = 1$ para $a = 1, b = 4 \Rightarrow$ se mantiene h_8. Find-S encuentra una hipótesis de máxima especificidad consistente con todos los ejemplos positivos que es h_8:

$$\text{Por lo tanto, } H = \{h_8 : (x,y) \in X \mid a,b \in \mathbb{N}, a \leq x < b, 8 = \frac{b(b-1)}{2} + a + 1\}$$

5. Ejercicio

En este ejercicio consideraremos $X = \{0,1\}^n$, i.e., X es el conjunto de todas las cadenas de longitud n formadas por ceros y unos.

1. ¿Cuantos ejemplos positivos del concepto palíndromo hay en X?

   Palíndromo es una palabra o frase cuyas letras están dispuestas de tal manera que resulta la misma leída de izquierda a derecha que de derecha a izquierda. Analizaré las posibles combinaciones para los distintos valores de n.

   - $n = 1$
     - $X = \{\{0\},\{1\}\} \Rightarrow p_1 = 2$
   - n = 2
     - $X = \{\{0,0\},\{1,1\}\} \Rightarrow p_2 = 2$
   - n = 3
     - $X = \{\{0,0,0\},\{0,1,0\},\{1,1,1\},\{1,0,1\}\} \Rightarrow p_3 = 4$
   - n = 4
     - $X = \{\{0,0,0,0\},\{\{1,0,0,1\},\{\{0,1,1,0\},\{\{1,1,1,1\}\} \Rightarrow p_4 = 4$
   - n = 5
     - $X = \{\{0,0,0,0,0\},\{0,0,1,0,0\},\{1,1,1,1,1\},\{1,0,0,0,1\},\{0,1,0,1,0\},\{1,0,1,0,1\},\{0,1,1,1,0\},\{1,1,0,1,1$
       $p_5 = 8$
   - n = 6
     - $X = \{\{0,0,0,0,0,0\},\{0,0,1,1,0,0\},\{1,1,1,1,1,1\},\{1,0,0,0,0,1\},\{1,1,0,0,1,1\},\{0,1,1,1,1,0\},\{0,1,0,0,1$
       $p_6 = 8$

Por definición de palíndromo y su simetría podemos analizar únicamente la primer mitad de la cadena. Veamos cuando n es par:

- Si $n = 2$: Las opciones son:
  - Cero 1's y un 0
  - Un 1 y cero 0's
  Ambas son permutaciones de elementos que se repiten. Supongamos que tenemos a veces el número 1 y b veces el número 0, tendremos una permutación de $n = a + b$ elementos en las que el número 1 se repite a veces y el número 0 b veces. Una forma de expresarlo es:
  $P_n^{a,b} = \frac{n!}{a!b!}, n = a + b$
  Por tanto, $p_2 = P_1^{0,1} + P_1^{1,0} = \frac{1!}{0!1!} + \frac{1!}{1!0!} = 2$
- Si $n = 4$:
  - Cero 1's y dos 0
  - Un 1 y un 0
  - Dos 1's y cero 0's
  Por tanto, $p_4 = P_2^{0,2} + P_2^{1,1} + P_2^{2,0} = \frac{2!}{0!2!} + \frac{2!}{1!1!} + \frac{2!}{2!0!} = 4$
- Si $n = 6$:
  - Cero 1's y tres 0's
  - Dos 1's y un 0
  - Un 1 y dos 0's
  - Tres 1's y cero 0's
  Por tanto, $p_6 = P_3^{0,3} + P_3^{2,1} + P_3^{1,2} + P_3^{3,0} = \frac{3!}{0!3!} + \frac{3!}{2!1!} + \frac{3!}{1!2!} + \frac{3!}{3!0!} = 8$
- Si n es par por lo expuesto:
  - Cero 1's y $(n/2)$ 0's
  - Un 1 y $(n/2 - 1)$ 0's
  - Dos 1's y $(n/2 - 2)$ 0's
  - ...
  - $(n/2 - 2)$ 1's y dos 0's
  - $(n/2 - 1)$ 1's y un 0
  - $(n/2)$ 1's y cero 0's
  Por tanto, $p_{n-par} = P_{n/2}^{0,n/2} + P_{n/2}^{2,n/2-2} + \ldots + P_{n/2}^{n/2-2,2} + +P_{n/2}^{n/2-1,1} + +P_{n/2}^{n/2,0}$

Veamos cuando n es impar:

- Si $n = 3$:
  - Cero 1's: 1 posibilidad
  - Un 1: 1 posibilidad
  - Dos 1's: 1 posibilidad
  - Tres 1's: 1 posibilidad
- Si $n = 5$:
  - Cero 1's: 1
  - Un 1: Equivalente a $n = 4$ con cero 1's
  - Dos 1's: Equivalente a $n = 4$ con dos 1's
  - Tres 1's: Equivalente a $n = 4$ con dos 1's
  - Cuatro 1's: Equivalente a $n = 4$ con cuatro 1's
  - Cinco 1's: 1
- Si $n = 7$:
  - Cero 10s: 1
  - Un 1: Equivalente a $n = 6$ con cero 1's
  - Dos 1's: Equivalente a $n = 6$ con dos 1's
  - Tres 1's: Equivalente a $n = 6$ con dos 1's
  - Cuatro 1's: Equivalente a $n = 6$ con cuatro 1's
  - Cinco 1's: Equivalente a $n = 6$ con cuatro 1's
  - Seis 1's: Equivalente a $n = 6$ con seis 1's
  - Siete 1's: 1

Por tanto, podemos verificar que al ser un numero impar, el numero del centro de la cadena no tiene ninguna restricción, Entonces para el n impar vamos a tener siempre el doble de opciones que para $n - 1$. Osea $p_{n-impar} = 2P_{n-1}$

En suma, el numero de ejemplos positivos de palíndromo en X sigue la siguiente regla:

$$P_n = \begin{cases} P_{n/2}^{0,n/2} + P_{n/2}^{2,n/2-2} + \ldots + P_{n/2}^{n/2-2,2} + +P_{n/2}^{n/2-1,1} + +P_{n/2}^{n/2,0} & \text{n } \textit{par} \\ 2P_{n-1} & \text{n } \textit{impar} \end{cases}$$

2. Sea ω el concepto definido en X de la siguiente manera: $\omega(y) = 1$ si y sólo si y contiene como máximo dos 1's. Prueba que el n´umero de ejemplos positivos de ω es una función cuadrática de n.

   Por definición y por ejercicio anterior encontramos la posibilidad de:

   - Cero 1's, hay 1 posible combinación.
   - Un 1, hay n combinaciones.
   - Dos 1's, hay $P_n^{2,n-2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$.

   Por tanto, el numero de positivos seguirá: $1 + n + \frac{n(n-1)}{2} = \frac{1}{2}n^2 + \frac{1}{2}n + 1$ que efectivamente es cuadrática

3. Supongamos que en un problemas de aprendizaje sobre$X$ aplicamos el algoritmo de *aprendizaje por enumeración* sobre el con junto de todas las hipótesis y las hipótesis están enumeradas de manera que la que buscamos está en la primera mitad. Si podemos probar un millón de hipótesis por segundo y $X = \{0,1\}^9$, ¿cuánto tiempo llevará encontrar la hipótesis buscada en el peor de los casos?

   Utilizando la formular de permutaciones tenemos que el conjunto de X esta formado por 512 elementos $P_9^{0,9} + P_9^{1,8} + \ldots + P_9^{9,1} + P_9^{9,0}$.
   Si el conjunto de hipótesis contiene todas las hipótesis, tenemos que $H = 2^X$, lo que quiere decir que en H hay $2^{512} \simeq 10^{153}$ hipótesis.
   Suponiendo que esta en la primera mitad, tenemos $10^{153}/2 \simeq 10^{153}$.

   Si podemos probar un millón de hipótesis por segundo, necesitamos $10^{153}/(60 * 10^6) \simeq 10^{145}$ segundos para probar nuestra hipótesis.