Máster Universitario en Lógica, Computación e Inteligencia Artificial.

Aprendizaje Automático

Diego Enrique Morales Polanco

Cuestionario 2

Ejercicio 1. Demostrar que la ganancia de información es siempre mayor o igual a cero. (Sugerencia: Desigualdad de Jensen).

$$Ganancia(D, A) = Ent(D) \ - \sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V)$$

Se quiere demostrar que:

$$Ent(D) \ - \sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V) \geq 0$$

Si el enunciado anterior es cierto, pues también es cierto que la entropía de un conjunto de ejemplos es mayor o igual que la entropía esperada de un subconjunto del conjunto D con valor del atributo A igual a v. Y si se demuestra cualquiera de los dos enunciados, el otro queda demostrado.

$$Ent(D) \geq \sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V)$$

Prueba:

$$Lado\ izquierdo = Ent(D)$$

$$Lado\ izquierdo = -\frac{|P|}{|D|}\log_2 \frac{|P|}{|D|} - \frac{|N|}{|D|}\log_2 \frac{|N|}{|D|}$$

Esta expresión se puede dividir en dos partes:

$$-\frac{|P|}{|D|}\log_2 \frac{|P|}{|D|} = a$$

$$-\frac{|N|}{|D|}\log_2 \frac{|N|}{|D|} = b$$

$$Ent(D) = a + b$$

Donde:

$$a, b \in \mathbb{R}^+$$

y

$$0 \leq a, b \leq 1$$

Consideremos:

$$|Valores(A)| = 2$$

$$Valores(A) = \{a, b\}$$

$$Lado\ derecho = \sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V)$$

$$Lado\ derecho = \frac{|P_a| + |N_a|}{|D|} * \left(-\frac{|P_a|}{|P_a| + |N_a|}\log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) - \frac{|N_a|}{|P_a| + |N_a|}\log_2\left(\frac{|N_a|}{|P_a| + |N_a|}\right)\right) + \frac{|P_b| + |N_b|}{|D|} * \left(-\frac{|P_b|}{|P_b| + |N_b|}\log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right) - \frac{|N_b|}{|P_b| + |N_b|}\log_2\left(\frac{|N_b|}{|P_b| + |N_b|}\right)\right)$$

Al igual que el lado izquierdo, el lado derecho se puede organizar en base a los casos positivos y negativos de cada valor del atributo A, de la siguiente manera:

Partiendo de:

$$Lado\ derecho = \frac{|P_a| + |N_a|}{|D|} * \left(-\frac{|P_a|}{|P_a| + |N_a|}\log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) - \frac{|N_a|}{|P_a| + |N_a|}\log_2\left(\frac{|N_a|}{|P_a| + |N_a|}\right)\right) + \frac{|P_b| + |N_b|}{|D|} * \left(-\frac{|P_b|}{|P_b| + |N_b|}\log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right) - \frac{|N_b|}{|P_b| + |N_b|}\log_2\left(\frac{|N_b|}{|P_b| + |N_b|}\right)\right)$$

Expandiendo:

$$Lado\ derecho = \left(-\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a| + |N_a|}\log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) - \frac{|P_a| + |N_a|}{|D|}\frac{|N_a|}{|P_a| + |N_a|}\log_2\left(\frac{|N_a|}{|P_a| + |N_a|}\right)\right) + \left(-\frac{|P_b| + |N_b|}{|D|}\frac{|P_b|}{|P_b| + |N_b|}\log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right) - \frac{|P_b| + |N_b|}{|D|}\frac{|N_b|}{|P_b| + |N_b|}\log_2\left(\frac{|N_b|}{|P_b| + |N_b|}\right)\right)$$

$$Lado\ derecho = \left(-\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a| + |N_a|}\log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) - \frac{|P_b| + |N_b|}{|D|}\frac{|P_b|}{|P_b| + |N_b|}\log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right)\right) + \left(-\frac{|P_a| + |N_a|}{|D|}\frac{|N_a|}{|P_a| + |N_a|}\log_2\left(\frac{|N_a|}{|P_a| + |N_a|}\right) - \frac{|P_b| + |N_b|}{|D|}\frac{|N_b|}{|P_b| + |N_b|}\log_2\left(\frac{|N_b|}{|P_b| + |N_b|}\right)\right)$$

Luego de agrupar de esta manera, podemos dividir el lado derecho en las siguientes expresiones:

$$-\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a| + |N_a|}\log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) - \frac{|P_b| + |N_b|}{|D|}\frac{|P_b|}{|P_b| + |N_b|}\log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right) = c$$

$$-\frac{|P_a| + |N_a|}{|D|}\frac{|N_a|}{|P_a| + |N_a|}\log_2\left(\frac{|N_a|}{|P_a| + |N_a|}\right) - \frac{|P_b| + |N_b|}{|D|}\frac{|N_b|}{|P_b| + |N_b|}\log_2\left(\frac{|N_b|}{|P_b| + |N_b|}\right) = d$$

$$\sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V) = c + d$$

Donde:

$$c, d \,\epsilon\, \mathbb{R}^{+}$$

y

$$0 \leq c, d \leq 1$$

Desigualdad de Jensen:

Está definida tanto para funciones cóncavas como para convexas. Indica que para una variable aleatoria integrable X a la cual se le aplica una función con dominio real y que también es integrable, la siguiente desigualdad se cumple:

Sea g una función convexa:

$$\mathrm{E}[g(X)] \geq g(\mathrm{E}[X])$$

Sea g una función cóncava:

$$\mathrm{E}[g(X)] \leq g(\mathrm{E}[X])$$

Siendo E la función esperanza matemática de la siguiente manera:

$$E(X) = \sum_{i=1}^{n} P(x_i) * x_i$$

Evaluando (c):

$$-\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a| + |N_a|} \log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) - \frac{|P_b| + |N_b|}{|D|} \frac{|P_b|}{|P_b| + |N_b|} \log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right)$$

$$-\left(\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a| + |N_a|} \log_2\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} \frac{|P_b|}{|P_b| + |N_b|} \log_2\left(\frac{|P_b|}{|P_b| + |N_b|}\right)\right)$$

Esto se puede escribir como:

$$\frac{|P_a| + |N_a|}{|D|} * g\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} g\left(\frac{|P_b|}{|P_b| + |N_b|}\right)$$

Siendo:

$$g(X) = -\mathrm{X} * \log_2 X$$

Se puede apreciar que tenemos la esperanza matemática del resultado de la evaluación de una función g, el lado izquierdo de la desigualdad de Jensen vista anteriormente, con una función g(X) la cual en nuestro caso es -X*log2(X), una función cóncava, por lo que podemos decir lo siguiente, aplicando la desigualdad de Jensen para funciones cóncavas:

$$\frac{|P_a| + |N_a|}{|D|} * g\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} g\left(\frac{|P_b|}{|P_b| + |N_b|}\right) \leq g(\mathrm{E}[X])$$

Siendo:

$$g(X) = -\mathrm{X} * \log_2 X$$

$$X = \frac{|P_a|}{|P_a| + |N_a|}, \frac{|P_b|}{|P_b| + |N_b|}$$

$$P(X) = \frac{|P_a| + |N_a|}{|D|}, \frac{|P_b| + |N_b|}{|D|}$$

Desarrollando:

$$\frac{|P_a| + |N_a|}{|D|} * g\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} g\left(\frac{|P_b|}{|P_b| + |N_b|}\right) \leq g(\mathrm{E}[X])$$

$$\frac{|P_a| + |N_a|}{|D|} * g\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} g\left(\frac{|P_b|}{|P_b| + |N_b|}\right)$$
$$\leq g\left(\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a + N_a|} + \frac{|P_b| + |N_b|}{|D|} * \frac{|P_b|}{|P_b| + |N_b|}\right)$$

Desarrollando el lado derecho de esta desigualdad:

$$g\left(\frac{|P_a| + |N_a|}{|D|} * \frac{|P_a|}{|P_a + N_a|} + \frac{|P_b| + |N_b|}{|D|} * \frac{|P_b|}{|P_b| + |N_b|}\right)$$

$$= g\left(\frac{|P_a|}{|D|} + \frac{|P_b|}{|D|}\right)$$

$$= g\left(\frac{|P_a| + |P_b|}{|D|}\right) = g\left(\frac{|P|}{|D|}\right)$$

Recordando que g(X) es

$$g(X) = -\mathrm{X} * \log_2 X$$

$$g\left(\frac{|P|}{|D|}\right) = -\frac{|P|}{|D|} * \log_2 \frac{|P|}{|D|}$$

Por lo tanto:

$$\frac{|P_a| + |N_a|}{|D|} * g\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} g\left(\frac{|P_b|}{|P_b| + |N_b|}\right) \leq g(\mathrm{E}[X])$$

$$\frac{|P_a| + |N_a|}{|D|} * g\left(\frac{|P_a|}{|P_a| + |N_a|}\right) + \frac{|P_b| + |N_b|}{|D|} g\left(\frac{|P_b|}{|P_b| + |N_b|}\right) \leq -\frac{|P|}{|D|} * \log_2 \frac{|P|}{|D|}$$

Podemos ver que el lado derecho corresponde con la parte de los valores de la clasificación P de la expresión de la entropía de D, la expresión (a) obtenida anteriormente, el lado izquierdo corresponde con la parte de los valores de la clasificación P de la expresión de la entropía esperada, ósea (c). También se puede apreciar que el número de valores posibles no hubiera afectado el resultado, ya que la función g(X) obtenida hubiera sido la misma, solo se tuvieran términos extra |P_c|, |P_d|… cuya sumatoria sería |P|, llegando a la misma conclusión aplicando la desigualdad de Jensen.

$$-\left(\frac{|P_a|+|N_a|}{|D|}*\frac{|P_a|}{|P_a|+|N_a|}\log_2\left(\frac{|P_a|}{|P_a|+|N_a|}\right)+\frac{|P_b|+|N_b|}{|D|}\frac{|P_b|}{|P_b|+|N_b|}\log_2\left(\frac{|P_b|}{|P_b|+|N_b|}\right)\right) \leq -\frac{|P|}{|D|}*\log_2\frac{|P|}{|D|}$$

Lo que significa que:

$$c \leq a$$

Haciendo el mismo desarrollo con la desigualdad de Jensen en la expresión (d), se puede llegar a la siguiente desigualdad:

$$-\left(\frac{|P_a|+|N_a|}{|D|}\frac{|N_a|}{|P_a|+|N_a|}\log_2\left(\frac{|N_a|}{|P_a|+|N_a|}\right)+\frac{|P_b|+|N_b|}{|D|}\frac{|N_b|}{|P_b|+|N_b|}\log_2\left(\frac{|N_b|}{|P_b|+|N_b|}\right)\right) \leq -\frac{|N|}{|D|}*\log_2\frac{|N|}{|D|}$$

Y:

$$d \leq b$$

Finalmente, y recordando que:

$$Ent(D) = a + b$$

Donde:

$$a, b \, \epsilon \, \mathbb{R}^+$$

y

$$0 \leq a, b \leq 1$$

Y:

$$\sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V) \ = c + d$$

Donde:

$$c, d \, \epsilon \, \mathbb{R}^+$$

y

$$0 \leq c, d \leq 1$$

$$a \geq c \, , b \geq d$$

Por tanto

$$a + b \geq c + d$$

$$Ent(D) \geq \sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V)$$

Y:

$$Ent(D) \; - \sum_{v \epsilon Valores(A)} \frac{|D_V|}{|D|} * Ent(D_V) \geq 0$$

Ejercicio 2. Dado un conjunto de entrenamiento D y dos atributos discretos X e Y, denotaremos por GanY (D, X) a la ganancia de información del atributo X cuando los valores del atributo Y se consideran las etiquetas de clasificación. Demostrar que GanY (D, X) = GanX(D, Y). (Sugerencia: Entropía Condicional).

Se quiere demostrar que:

$$GanY(D,X) = GanX(D,Y)$$

Expandiendo la expresión con la fórmula de la ganancia de información esperada:

$$Ent(D_Y) - \sum_{v \epsilon valores(X)} \frac{|D_v|}{|D|} * Ent(D_v) = Ent(D_X) - \sum_{v \epsilon valores(Y)} \frac{|D_v|}{|D|} * Ent(D_v)$$

Donde:

D_Y es el conjunto de entrenamiento con los valores del atributo Y como las etiquetas de clasificación.

D_X es el conjunto de entrenamiento con los valores del atributo X como las etiquetas de clasificación.

Es importante hacer esta distinción porque la evaluación de la función entropía cambia si se toma un atributo u otro como las etiquetas de clasificación.

Consideremos, por simplicidad, que los atributos tanto X como Y tienen solo 2 valores (positivo y negativo):

$$|Valores(X)| = 2$$

$$|Valores(Y)| = 2$$

Analizando la expresión de la entropía esperada, se puede escribir como:

$$\begin{aligned}\sum_{v \epsilon valores(X)} \frac{|D_v|}{|D|} * Ent(D_v) \\ = \frac{|P_{Y,(P_X)}| + |N_{Y,(P_X)}|}{|D|} * \left(- \frac{|P_{Y,(P_X)}|}{|P_{Y,(P_X)}| + |N_{Y,(P_X)}|} \log_2 \left(\frac{|P_{Y,(P_X)}|}{|P_{Y,(P_X)}| + |N_{Y,(P_X)}|} \right) - \frac{|N_{Y,(P_X)}|}{|P_{Y,(P_X)}| + |N_{Y,(P_X)}|} \log_2 \left(\frac{|N_{Y,(P_X)}|}{|P_{Y,(P_X)}| + |N_{Y,(P_X)}|} \right) \right) \\ + \frac{|P_{Y,(N_X)}| + |N_{Y,(N_X)}|}{|D|} * \left(- \frac{|P_{Y,(N_X)}|}{|P_{Y,(N_X)}| + |N_{Y,(N_X)}|} \log_2 \left(\frac{|P_{Y,(N_X)}|}{|P_{Y,(N_X)}| + |N_{Y,(N_X)}|} \right) - \frac{|N_{Y,(N_X)}|}{|P_{Y,(N_X)}| + |N_{Y,(N_X)}|} \log_2 \left(\frac{|N_{Y,(N_X)}|}{|P_{Y,(N_X)}| + |N_{Y,(N_X)}|} \right) \right) = a\end{aligned}$$

Esta expansión la llamaremos a.

Donde,

$P_{Y,(P_X)}$ son los elementos con valor positivo de Y cuyo atributo X también es positivo,

$P_{Y,(N_X)}$ son los elementos con valor positivo de Y cuyo atributo X es negativo,

$N_{Y,(P_X)}$ son los elementos con valor negativo de Y cuyo atributo X es positivo, y

$N_{Y,(N_X)}$ son los elementos con valor negativo de Y cuyo atributo X es negativo, Podemos reescribir tanto la expresión la expresión anterior en términos de probabilidad, considerando que X y Y son variables aleatorias con dos posibles valores.

Tomando en cuenta que:

$p(Y)$ es la probabilidad que la variable Y tome el valor positivo de clasificación,

$\left(1-p(Y)\right)$ es la probabilidad que la variable Y tome el valor negativo de clasificación

Reescribiendo:

$\frac{\left|P_{Y,(P_X)}\right|+\left|N_{Y,(P_X)}\right|}{|D|}$, puede considerarse como la probabilidad de que x tenga valor positivo $p(x)$

$\frac{|P_{Y,(N_X)}|+|N_{Y,(N_X)}|}{|D|}$, puede considerarse como la probabilidad de que x tenga valor negativo $\left(1-p(x)\right)$

$\frac{\left|P_{Y,(P_X)}\right|}{\left|P_{Y,(P_X)}\right|+\left|N_{Y,(P_X)}\right|}$ puede considerarse como la probabilidad de que la variable Y tome el valor positivo de la clasificación dado que sepamos que el valor de la variable X es positivo, por lo que podemos reescribir este término como una probabilidad condicional $p\left(y\middle|(x=P)\right)$

$\frac{\left|N_{Y,(P_X)}\right|}{\left|P_{Y,(P_X)}\right|+\left|N_{Y,(P_X)}\right|}$ puede considerarse como la probabilidad de que la variable Y tome el valor negativo de la clasificación dado que sepamos que el valor de la variable X es positivo, por lo que podemos reescribir este término como una probabilidad condicional $\left(1-p\left(y\middle|(x=P)\right)\right)$

$\frac{\left|P_{Y,(P_X)}\right|}{\left|P_{Y,(P_X)}\right|+\left|N_{Y,(P_X)}\right|}$ puede considerarse como la probabilidad de que la variable Y tome el valor positivo de la clasificación dado que sepamos que el valor de la variable N es positivo, por lo que podemos reescribir este término como una probabilidad condicional $p\left(y\middle|(x=N)\right)$

$\frac{\left|P_{Y,(P_X)}\right|}{\left|P_{Y,(P_X)}\right|+\left|N_{Y,(P_X)}\right|}$ puede considerarse como la probabilidad de que la variable Y tome el valor positivo de la clasificación dado que sepamos que el valor de la variable X es positivo, por lo que podemos reescribir este término como una probabilidad condicional $\left(1-p\left(y\middle|(x=N)\right)\right)$

Estas probabilidades condicionales se pueden reescribir de la siguiente manera a partir de la probabilidad conjunta:

$$p\left(y\middle|(x=P)\right)=\frac{p(y,x=P)}{p(y)}$$

$$\left(1-p\left(y\middle|(x=P)\right)\right)=\left(1-\frac{p(y,x=P)}{p(y)}\right)$$

$$p\left(y\middle|(x=N)\right)=\frac{p(y,x=N)}{p(y)}$$

$$\left(1-p\left(y\middle|(x=N)\right)\right)=\left(1-\frac{p(y,x=N)}{p(y)}\right)$$

De una forma más reducida, la expresión “a” puede ser escrita con las probabilidades condicionales y conjuntas como:

$$\sum_{v \epsilon valores(X)} \frac{|D_v|}{|D|} * Ent(D_v) = - \sum_{x \in X, y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(y)}\right)$$

Esto es equivalente a un concepto llamado entropía condicional, la cual se denota como:

$$H(X|Y) = - \sum_{x \in X, y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)}\right)$$

La ganancia de información al fijar un atributo en términos generales se puede escribir a partir de la entropía condicional, la cuál es equivalente a la entropía esperada al fijar un atributo del árbol:

$$GanX(D,Y) = H(X) - H(X|Y)$$

$$GanY(D,X) = H(Y) - H(Y|X)$$

Siendo:

$H(X)$ la entropía del conjunto de entrenamiento D, cuando se fija X como variable target, y

$H(X|Y)$ es la entropía condicional o entropía esperada cuando se conoce el valor de Y.

Partiendo de la regla de la cadena de las probabilidades:

$$p(x,y) = p(x)p(y|x)$$

Esta regla se expande a la entropía, pero sumando en lugar de multiplicando debido a los logaritmos presentes en la función de entropía:

$$H(X,Y) = H(X) + H(Y|X)$$

pero también:

$$H(X,Y) = H(Y) + H(X|Y)$$

Por lo que:

$$GanX(D,Y) = H(X) - H(X|Y)$$

$$GanX(D,Y) = H(X) - \big(H(X,Y) - H(Y)\big)$$

$$GanX(D,Y) = H(X) - H(X,Y) + H(Y)$$

Recordando que:

$$H(X,Y) = H(X) + H(Y|X)$$

$$H(X,Y) - H(X) = H(Y|X)$$

$$H(X) - H(X,Y) = -H(Y|X)$$

Por lo que:

$$GanX(D,Y) = H(X) - H(X,Y) + H(Y)$$

$$GanX(D,Y) = H(Y) - H(Y|X)$$

Recordando que:

$$GanY(D,X) = H(Y) - H(Y|X)$$

Queda demostrado que:

$$GanY(D,X) = GanX(D,Y)$$

Ejercicio 3. 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 $\bigcup_{c \in C} \bigcup_{m \geq 1} S(m,c)$. Demostrar que si $H \neq 2^U$, entonces L no es consistente.

Por definición:

Un algoritmo de aprendizaje L es consistente si para todo $s \in Dom(L)$, $L(s) = h$ es consistente con s.

$$Dom(L) = \bigcup_{c \in C} \bigcup_{m \geq 1} S(m,c)$$

El dominio de L es la unión de todos los conjuntos de vectores de entrenamiento S, formados por todos los conceptos pertenecientes a los conceptos objetivos c, de tamaño $m \geq 1$.

Si:

$$H \neq 2^U$$

Significa que existirá por lo menos un $c_j \in C$, y una $h_j \in H$, que para un elemento $x_j \in U$

$$c_j(x_j) \neq h_j(x_j)$$

Esto a su vez significa que los vectores de entrenamiento $s(c_j, m)$ que contengan el concepto c_j y el elemento x_j no serán consistentes con h_j. Por lo que L no es consistente.

Ejercicio 4. Aplica los algoritmos de aprendizaje por enumeración y Find-S para el problema de aprendizaje definido a continuación. Consideraremos que el universo de instancias X es $\mathbb{R}^2$. Consideraremos un conjunto infinito de hipótesis $H = \{h_n | n \in \mathbb{N}\}$ e identificaremos cada hipótesis h_n como un conjunto de puntos del plano. De este modo consideraremos que $h_n(x) = 1$ si $x \in h_n$ y $h_n(x) = 0$ si $x \notin h_n$. Por ejemplo, h_0 lo describimos como el conjunto vació $h_0 = \emptyset$. Esto significa que para todo $x \in \mathbb{R}^2$, se tiene que $h_0(x) = 0$, ya que para todo $x \in \mathbb{R}^2$, se tiene que $x \notin \emptyset$. Para aplicar los algoritmos de aprendizaje consideraremos también un vector de entrenamiento s. Los datos son los siguientes:

$X = \mathbb{R}^2$

$H = \{h_n | n \in \mathbb{N}\}$ con $h_0 = \emptyset$ y si $n > 1$, entonces

$$h_n = \{(x,y) \in X | a, b \in \mathbb{N}, a \leq x < b, n = \frac{b(b-1)}{2} + a + 1\}$$

$$s = \{\langle (0,0), 0 \rangle, \langle (3,4), 1 \rangle, \langle (2,2), 1 \rangle\}$$

-Aprendizaje por enumeración:

1. Enumeración del conjunto de entrenamiento:

$D = \{\langle(0,0),0\rangle, \langle(3,4),1\rangle, \langle(2,2),1\rangle\}$

2. Enumeración de las hipótesis expresables:

$h_0 = \emptyset$

Para n>1

$h_n = \{(x,y) \in X | a,b \in \mathbb{N}, a \leq x < b, n = \frac{b(b-1)}{2} + a + 1\}$

3. Recorrer el conjunto de hipótesis buscando una hipótesis que cumpla con la clasificación correcta de todo el conjunto de entrenamiento:

Se empieza desde h_2 porque la expresión de las hipótesis no está definida para $n = 1$.

$h_2 = \{(x,y) \in X | a,b \in \mathbb{N}, a \leq x < b, 2 = \frac{b(b-1)}{2} + a + 1\}$

$a, b \in \mathbb{N}$

Para x_1:

$$x = 0$$

$$a \leq 0 < b$$

Intentar con:

$$a = 0, b = 1$$

$$\frac{1(1-1)}{2} + 0 + 1 = 1$$

$$2 \neq 1$$

x_1 no se encuentra en el conjunto de puntos definido por los valores de a y b elegidos, por lo que la hipótesis con estos valores clasifica correctamente el primer valor del conjunto de entrenamiento. Continuando con x_2

$$x = 3$$

$$0 \leq 3 < 1?$$

Estos valores de a y b no clasifican correctamente x_2.

Intentando con:

$$a = 0, b = 4$$

$$\frac{4(4-1)}{2} + 0 + 1 = 7$$

$$2 \neq 7$$

a, no puede tener un valor más pequeño que 0 y b no puede tener un valor menor que 4 y que se siga cumpliendo la condición $a \leq x < b$, pero a la vez un aumento en el valor de a o b hará que el valor de la expresión $\frac{b(b-1)}{2} + a + 1$

aumente, por lo que no se cumplirá que esta expresión será igual a n, en este caso 2. Por lo que se continuará evaluando la siguiente hipótesis.

$h_3 = \{(x, y) \in X | a, b \in \mathbb{N}, a \leq x < b, 3 = \dfrac{b(b-1)}{2} + a + 1\}$

$a, b \in \mathbb{N}$

Para x_1:

$$x = 0$$

$$a \leq 0 < b$$

Intentar con:

$$a = 0, b = 1$$

$$\frac{1(1-1)}{2} + 0 + 1 = 1$$

$$3 \neq 1$$

x_1 no se encuentra en el conjunto de puntos definido por los valores de a y b elegidos, por lo que la hipótesis con estos valores clasifica correctamente el primer valor del conjunto de entrenamiento. Continuando con x_2

$$x = 3$$

$$0 \leq 3 < 1?$$

Estos valores de a y b no clasifican correctamente x_2.

Intentando con:

$$a = 0, b = 4$$

$$\frac{4(4-1)}{2} + 0 + 1 = 7$$

$$3 \neq 7$$

Se puede apreciar que ocurre lo mismo que en la hipótesis anterior. 7 es el valor más pequeño que se puede obtener de la expresión $\frac{b(b-1)}{2} + a + 1$ con valores de a y b que cumplan $a \leq x < b$ para el segundo elemento del conjunto de entrenamiento ($x = 3$). Por esto, se continuará evaluando la hipótesis 7.

$h_7 = \{(x, y) \in X | a, b \in \mathbb{N}, a \leq x < b, 7 = \dfrac{b(b-1)}{2} + a + 1\}$

$a, b \in \mathbb{N}$

Para x_1:

$$x = 0$$

$$a \leq 0 < b$$

Intentar con:

$$a = 0, b = 1$$

$$\frac{1(1-1)}{2}+0+1=1$$

$$7 \neq 1$$

x_1 no se encuentra en el conjunto de puntos definido por los valores de a y b elegidos, por lo que la hipótesis con estos valores clasifica correctamente el primer valor del conjunto de entrenamiento. Continuando con x_2

$$x=3$$

$$0 \leq 3 < 1?$$

Estos valores de a y b no clasifican correctamente x_2.

Intentando con:

$$a=0, b=4$$

$$0 \leq 3 < 4$$

$$\frac{4(4-1)}{2}+0+1=7$$

$$7=7$$

Con estos valores de a y b, el segundo punto del conjunto de entrenamiento se encuentra dentro del conjunto de puntos de la hipótesis 7, por lo que lo clasifica correctamente. Pero clasifica incorrectamente x_1

$$x=0$$

$$0 \leq 0 < 4$$

$$\frac{4(4-1)}{2}+0+1=7$$

$$7=7$$

Se puede apreciar que para estos valores de a y b, x_1 se encuentra en el conjunto de puntos descrito por h_7. Si se aumenta el valor de a o b, x_2 dejaría de cumplir con la expresión $\frac{b(b-1)}{2}+a+1$, por lo que dejaría de ser parte del conjunto de puntos descrito. Debido a esto se descarta h_7 y se continúa evaluando h_8.

$h_8 = \{(x,y) \in X | a,b \in \mathbb{N}, a \leq x < b, 8 = \dfrac{b(b-1)}{2}+a+1\}$

$a,b \in \mathbb{N}$

Para x_1:

$$x=0$$

$$a \leq 0 < b$$

Intentar con:

$$a=0, b=1$$

$$\frac{1(1-1)}{2}+0+1=1$$

$$8 \neq 1$$

x_1 no se encuentra en el conjunto de puntos definido por los valores de a y b elegidos, por lo que la hipótesis con estos valores clasifica correctamente el primer valor del conjunto de entrenamiento. Continuando con x_2

$$x = 3$$

$$0 \leq 3 < 1?$$

Estos valores de a y b no clasifican correctamente x_2.

Intentando con:

$$a = 0, b = 4$$

$$0 \leq 3 < 4$$

$$\frac{4(4-1)}{2} + 0 + 1 = 7$$

$$8 \neq 7$$

En este caso, el valor de la expresión es menor que n, por lo que se puede aumentar el valor de a o b en lugar de descartar la hipótesis.

Intentando con:

$$a = 1, b = 4$$

$$1 \leq 3 < 4$$

$$\frac{4(4-1)}{2} + 1 + 1 = 8$$

$$8 = 8$$

Se puede ver que h_8 con $a = 1$ y $b = 4$, clasifica correctamente x_2, también clasifica correctamente x_1 ya que la expresión $a \leq x < b$ no se cumple para $x = 0$, $(1 \leq 0 < 4)$. Continuando con x_3:

$$x = 2$$

Intentando con:

$$a = 1, b = 4$$

$$1 \leq 2 < 4$$

$$\frac{4(4-1)}{2} + 1 + 1 = 8$$

$$8 = 8$$

Con $a = 1\ y\ b = 4$, h_8 clasifica correctamente todos los elementos del conjunto de entrenamiento.

-Find-S:

$s = \{\langle(0,0),0\rangle, \langle(3,4),1\rangle, \langle(2,2),1\rangle\}$

$h_n = \{(x,y) \in X | a,b \in \mathbb{N}, a \leq x < b, n = \dfrac{b(b-1)}{2} + a + 1\}$

Partiendo de la hipótesis más específica:

$$h = \emptyset$$

Ejemplo positivo:

$$x_2 = \langle(3,4),1\rangle$$

$$h(x_2) = 0$$

Reemplazar h por una menor generalización h' de h, tal que $h'(x_2) = 1$. Del ejercicio anterior, sabemos que la primera hipótesis que puede clasificar correctamente x_2 es h con n=7.

$$x = 3$$

$$n = 7$$

$$a = 0, b = 4$$

$$0 \leq x \leq 4$$

$$\frac{b(b-1)}{2} + a + 1 = \frac{4(4-1)}{2} + 0 + 1 = 7$$

$$7 = 7$$

$$h' = \{(x,y) \in X|, 0 \leq x < 4, 7 = \frac{4(4-1)}{2} + 0 + 1\}$$

$$h'(x_2) = 1$$

Siguiente ejemplo positivo:

$$x_3 = \langle(2,2),1\rangle$$

$$h' = \{(x,y) \in X|, 0 \leq x < 4, 7 = \frac{4(4-1)}{2} + 0 + 1\}$$

$$h'(x_3) = 1$$

La hipótesis encontrada clasifica correctamente todos los ejemplos positivos. La hipótesis final encontrada por find-S es:

$$h_7 = \{(x,y) \in X|, 0 \leq x < 4, 7 = \frac{4(4-1)}{2} + 0 + 1\}$$

Se pudo encontrar más rápido que con el algoritmo de enumeración, pero no clasifica correctamente el ejemplo negativo del conjunto de entrenamiento.

Ejercicio 5. Responder razonadamente si las siguientes afirmaciones son verdaderas o falsas. Si la respuesta es **Verdadera**, debe dar razones que apoyen tu decisión. Si la respuesta es **Falsa** debes dar un ejemplo en el que no se verifique la afirmación.

(a) Sean D_1 y D_2 dos conjuntos de entrenamiento para el mismo concepto. Entonces:

$$\frac{Ent(D_1) + Ent(D_2)}{2} \leq Ent(D_1 \cup D_2)$$

Verdadero.

Si en lugar de D_1 y D_2 utilizamos las variables aleatorias "X" y "Y", las cuales se refieren a la distribución de probabilidad de la clasificación del concepto en D_1 y D_2 respectivamente, podemos escribir la expresión anterior como:

$$\frac{H(X) + H(Y)}{2} \leq H(X, Y)$$

Donde $H(X, Y)$ es la entropía conjunta de ambas variables aleatorias. La entropía conjunta de dos o más variables aleatorias tiene la propiedad de ser mayor o igual que la entropía de cada variable aleatoria individualmente:

$$H(X, Y) \geq \max\big(H(X), H(Y)\big)$$

Esto intuitivamente hace sentido, ya que la unión de dos conjuntos tendrá, en el peor de los casos (que uno de los conjuntos esté vacío o que ambos conjuntos contengan exactamente los mismos elementos), la misma medida de dispersión que el conjunto más disperso individualmente. Intuitivamente, si tenemos un conjunto con una medida de dispersión y unimos ese conjunto con otro, la dispersión del resultado no puede ser menor que la del conjunto original.

$$\frac{H(X) + H(Y)}{2}$$

En este promedio de las entropías, sabiendo que:

$$0 \leq H(X) \leq 1$$

$$0 \leq H(Y) \leq 1$$

Pues, podemos afirmar que se cumple la siguiente propiedad de los promedios aritméticos:

$$\frac{H(X) + H(Y)}{2} \leq \max\big(H(X), H(Y)\big)$$

Finalmente:

$$\frac{H(X) + H(Y)}{2} \leq \max\big(H(X), H(Y)\big) \leq H(X, Y)$$

Por lo que:

$$\frac{H(X) + H(Y)}{2} \leq H(X, Y)$$

$$\frac{Ent(D_1) + Ent(D_2)}{2} \leq Ent(D_1 \cup D_2)$$

(b) Sean D_1 y D_2 dos conjuntos de entrenamiento para el mismo concepto tales que $D_1 \cap D_2 \neq \emptyset$. Entonces $Ent(D_1 \cup D_2) < Ent(D_1) + Ent(D_2)$.

Verdadero.

Si en lugar de D_1 y D_2 utilizamos las variables aleatorias "X" y "Y", las cuales se refieren a la distribución de probabilidad de la clasificación del concepto en D_1 y D_2 respectivamente, podemos escribir la expresión anterior como:

$$H(X,Y) < H(X) + H(Y)$$

Donde $H(X,Y)$ es la entropía conjunta de ambas variables aleatorias. La función de entropía, por definición, es una función subaditiva. La subaditividad es una propiedad de una función que establece, resumidamente, que la evaluación de la función para la suma de dos elementos del dominio siempre devuelve algo menor o igual que la suma de los valores de la función en cada elemento.

$$H(X,Y) \leq H(X) + H(Y)$$

La igualdad solo se cumple si uno de los conjuntos está vacío o si las variables son estadísticamente independientes ($D_1 \cap D_2 = \emptyset$). Por lo que, para el caso en que haya elementos repetidos en ambos conjuntos ($D_1 \cap D_2 \neq \emptyset$):

$$H(X,Y) < H(X) + H(Y)$$

$$Ent(D_1 \cup D_2) < Ent(D_1) + Ent(D_2)$$