

Aprendizaje automático
Departamento de Ingeniería en Informática
ITBA

Trabajo Práctico 1

Objetivo: Aprender a diseñar un sistema de aprendizaje y a aplicar algoritmos para el aprendizaje de conceptos

I. Diseño de sistemas de aprendizaje

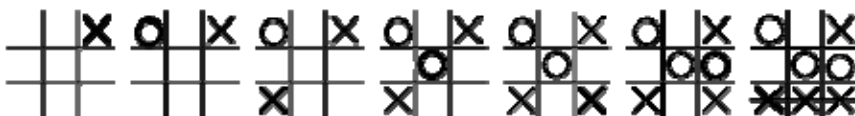
1. Dar un ejemplo de una tarea de aprendizaje.
 - (a) Describir la tarea, la medida de rendimiento y la experiencia de aprendizaje.
 - (b) Proponer una función objetivo que deba ser aprendida y su representación.
2. Considerar estrategias alternativas para el generador de experimentos visto en clase para el juego de damas.

En particular, considerar estrategias en las cuales el generador de experimentos sugiere nuevas posiciones del tablero

 - generando al azar posiciones del tablero válidas
 - generando una posición escogiendo un estado del tablero de un juego previo, y luego aplicando un movimiento que no haya sido ejecutado
 - una nueva propuesta de estrategia
3. Implementar un algoritmo similar al de juego de damas pero para el juego del tatetí.

El tatetí es un juego entre dos jugadores: O y X, que marcan los espacios de un tablero de 3×3 alternadamente. Un jugador gana si consigue tener una línea de tres de sus símbolos: la línea puede ser horizontal, vertical o diagonal.

Por ejemplo, una partida ganada por el primer jugador, X:



Una partida que termina en empate:



- (a) Representar la función aprendida como una combinación lineal de características del tablero.
- (b) Para entrenar al programa, jugar repetidamente contra una segunda copia del programa que use una función de evaluación fija.

Hacer un gráfico del porcentaje de juegos ganados por su sistema versus el número de juegos de entrenamiento jugados.

II. Aprendizaje de conceptos

1. Explicar porqué el tamaño del espacio de hipótesis en la tarea de aprendizaje **disfruta-deporte** es 973.
 - (a) ¿Cómo aumentaría el número de instancias y de hipótesis posibles si se agrega el atributo **corriente** (del agua) el cual puede tomar los valores *débil*, *moderada* o *fuerte*?
 - (b) En general, ¿cómo crece el número de posibles instancias e hipótesis si se agrega un nuevo atributo A que puede tomar k valores posibles?
2. Dar la secuencia de conjunto frontera S y G obtenida mediante el algoritmo de ELIMINACIÓN-DE-CANDIDATOS si la secuencia de ejemplos de entrenamiento se da en orden *inverso* (4-3-2-1) a los de la siguiente tabla:

ejemplo	cielo	temperatura del aire	humedad	viento	temperatura del agua	pronóstico del tiempo	disfruta deporte
1	soleado	cálida	normal	fuerte	cálida	igual	si
2	soleado	cálida	alta	fuerte	cálida	igual	si
3	nublado	fría	alta	fuerte	cálida	cambiante	no
4	soleado	cálida	alta	fuerte	fría	cambiante	si

3. Considerar la tarea de aprendizaje **disfruta-deporte** y el espacio de hipótesis H descrito como una conjunción de restricciones sobre los atributos **Cielo, Aire, Humedad, Viento, Agua y Pronóstico**.
 Definamos un nuevo espacio de hipótesis H' que contenga todas las disyunciones de pares de las hipótesis en H . Por ejemplo, una hipótesis en H' es
 $\langle ?, \text{cálida}, \text{alta}, ?, ?, ? \rangle \vee \langle \text{soleado}, ?, \text{alta}, ?, ?, \text{igual} \rangle$
 Aplicar el algoritmo de ELIMINACIÓN-DE-CANDIDATOS para el espacio de hipótesis H' para la secuencia de entrenamiento de la tabla del ejercicio 2.
4. Considerar la siguiente secuencia de ejemplos de entrenamiento positivos y negativos que describen el concepto “pares de personas que viven en la misma casa”.
 Cada ejemplo de entrenamiento se describe como un par *ordenado* de personas, en el cual cada persona es descrita por su sexo, color de cabello (morocha, castaño o rubio), altura (alto, mediano o bajo) y nacionalidad (argentino, italiano, español, francés, alemán, inglés o japonés)
 - + <<mujer morocha baja argentina> <varón castaño alto argentino>>
 - + <<mujer morocha baja argentina> <varón castaño bajo francés>>
 - <<mujer morocha baja inglesa> <mujer castaña alta alemana>>
 - + <<mujer castaña baja española> <varón castaño alto español>>

Considerar un espacio de hipótesis basado en estas instancias, en el cual cada hipótesis se represente con una 4-upla y donde las restricciones de cada atributo pueden ser un valor específico, “?” o “∅”. Por ejemplo, la hipótesis

<<mujer ? ? japonesa> <varón ? alto ?>>

representa a todos los pares de personas en los que la primera persona es una mujer japonesa y la segunda un varón alto.

(a) Aplicar el algoritmo de ELIMINACIÓN-DE-CANDIDATOS para los ejemplos de entrenamiento previos.

(b) ¿Cuántas hipótesis distintas del espacio hipótesis dado son consistentes con el siguiente ejemplo entrenamiento positivo?

+ << mujer rubia alta inglesa> <varón morocho bajo italiano>>

(c) Supongamos que el aprendiz ha encontrado sólo un ejemplo positivo en la parte (b), y que se le permite consultar al entrenador, después de haber generado cualquier instancia, para que la clasifique.

- i. Dar una secuencia específica de consultas que asegure al aprendiz que convergerá a una única hipótesis correcta.
- ii. Dar la secuencia más corta de consultas que puede encontrar. ¿Cómo se relaciona el largo de esta secuencia con la respuesta a la pregunta (b)?

5. Implementar el algoritmo ENCONTRAR-S.

(a) Verificar que produce la traza vista en clase para el ejemplo **disfruta-deporte**.

(b) Usar el programa para estudiar el número de ejemplos de entrenamiento al azar requeridos para aprender exactamente el concepto objetivo, y luego clasificarlos de acuerdo al siguiente concepto objetivo <soleado,cálida, ?,?,?,?>

(c) Aplicar su programa de entrenamiento sobre ejemplos generados al azar y medir el número de ejemplos necesarios hasta que la hipótesis del programa sea idéntica a la del concepto objetivo.

- i. ¿Podría predecir el número promedio de ejemplos necesarios?. Realizar el experimento 20 veces y reportar el número medio de experimentos necesarios.
- ii. ¿Cómo espera que varíe este número con respecto al número de “?” en el concepto objetivo?
- iii. ¿Cómo variaría este número en función del número de atributos usados para describir las instancias y las hipótesis?