



4 de diciembre, 2018.

Examen

Nombre:

- a. Tiempo: 9:00 a 12:00.
- b. Está permitido usar apuntes personales.
- c. No está permitido usar impresiones o fotocopias.
- d. A excepción de cuando el profesor o los ayudantes indiquen, cualquier tipo de comunicación con otra persona que habite este planeta será considerada como caso de copia. Conexiones espirituales o con otros mundos son bienvenidas.
- e. No olvide silenciar su teléfono celular.

Pregunta	Puntaje Máximo	Puntaje Obtenido
P1	14	
P2	12	
P3	12	
P4	12	
P5	12	
P6	12	
P7	12	
	TOTAL	
	NOTA FINAL	

1. (14 puntos) Responda indicando si la expresión es verdadera o falsa, o respondiendo directamente la consulta respectiva. En todo los casos fundamente brevemente su respuesta.

- 1.1. El módulo más importante para construir sistemas que contestan exitosamente esquemas de Winograd, es el que hace análisis gramatical de las oraciones.

Solución: Falso. Los esquemas de Winograd exigen mucho más que eso. Específicamente, un esquema de Winograd en general no se puede contestar sin usar conocimiento de sentido común.

- 1.2. Si el conjunto de fórmulas de lógica de primer orden, Σ , no menciona P , entonces $\Sigma \not\models \exists x P(x)$.

Solución: Verdadero. Dado que Σ no menciona a P es posible construir interpretaciones para Σ que interpreten a P de forma arbitraria. En particular, con extensión vacía.

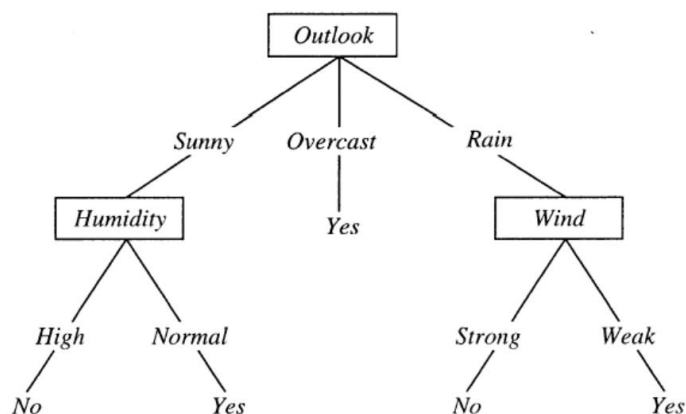
- 1.3. A^* retorna cuando un nodo objetivo es extraído de la lista Open. Imagine que una implementación de A^* se modifica para que retorne cuando un nodo objetivo es agregado a la Open. Entonces, la implementación resultante no retorna siempre soluciones óptimas, incluso cuando la heurística es admisible.

Solución: Verdadero. Para justificar basta con mostrar un ejemplo en el que el objetivo es alcanzable por dos caminos: uno de una sola arista, pero cara y otro de dos aristas, pero mucho más barato. La modificación de A^* encontrará el camino de una arista; es decir, el más caro.

- 1.4. Sea h una heurística admisible y h' una heurística inadmisble. Entonces A^* tiene un tiempo de ejecución menor o igual cuando es ejecutado con h que cuando es ejecutado con h' .

Solución: Falso. Con una heurística como $h' = 2h$ usualmente se observará que A^* ejecuta más rápido que con h .

- 1.5. Para las siguientes 5 preguntas considere el árbol de decisión de la figura. El árbol es similar al visto en clases y permite clasificar si una persona irá o no a jugar tenis.



Para la instancia [outlook=sunny, temperature=hot, wind=strong, humidity=normal] la predicción del árbol es : Jugar Tenis.

Solución: Verdadero. Verificar siguiendo una de las ramas del lado izquierdo del árbol.

- 1.6. Para la instancia [outlook=rain, temperature=cool, wind=strong, humidity=normal] la predicción del árbol es : No Jugar Tenis.

Solución: Verdadero. Verificar siguiendo una de las ramas del lado derecho del árbol.

- 1.7. Asumiendo que este árbol es obtenido con el algoritmo de construcción de árboles de decisión visto en clases (ID3), indique si es posible garantizar que para los datos usados en el entrenamiento, este árbol es el que entrega mejor rendimiento entre todos los árboles de decisión que componen el espacio de hipótesis.

Solución: *No es posible garantizar lo indicado. El algoritmo ID3 es de tipo codicioso (greedy) por tanto no garantiza encontrar un óptimo global.*

- 1.8. Asumiendo que este árbol es obtenido con el algoritmo de construcción de árboles de decisión visto en clases (ID3), indique si es posible garantizar que la ganancia de información del atributo outlook es mayor que la del atributo rain.

Solución: *Falso. Rain no es un atributo, sino un valor del atributo outlook.*

- 1.9. Asumiendo que este árbol es obtenido con el algoritmo de construcción de árboles de decisión visto en clases (ID3) y que todos los registros de entrenamiento son distintos, indique si es posible garantizar que todos los ejemplos de entrenamiento de la clase "Jugar Tenis" fueron obtenidos en días nublados (overcast).

Solución: *Verdadero. La rama del árbol correspondiente a "outlook=overcast" es terminal, por lo cual se cumple alguna de las 2 condiciones de término de la recursión del algoritmo ID3. Dado que se indica que todos los registros de entrenamiento son distintos, la única condición restante es que todos sean de la misma clase.*

- 1.10. El siguiente enunciado es común para las siguientes 5 preguntas. Considere un problema de clasificación donde se tiene un set de n datos de entrenamiento: $TS = \{(x_i, y_i), i = 1, \dots, n\}$, $x_i \in \mathbb{R}^2$ y $y_i \in [-1, 1]$. Sea $w = w^*$, $c = c^*$ los parámetros de un clasificador lineal de máximo margen que se obtiene al resolver el siguiente problema de optimización:

$$\underset{w, c}{\operatorname{argmin}} \frac{1}{2} \|w\|^2$$

sujeto a: $y_i(w \cdot x_i + c) \geq 1, \quad \forall (x_i, y_i) \in TS, \quad \|\cdot\| = \text{norma } \ell_2$

Bajo las condiciones descritas, indique si es posible garantizar que los datos de entrenamiento TS son linealmente separables.

Solución: *Si es posible garantizar lo indicado. Esto debido a la existencia de un plano separador y el no uso de variables de slack.*

- 1.11. Indique si es posible garantizar que los vectores de soporte tienen un margen de clasificación igual a 1.

Solución: *Si es posible garantizar lo indicado. Esto pues en el problema de optimización planteado el mínimo margen requerido es 1.*

- 1.12. Se tiene además un clasificador lineal con parámetros $w = \hat{w}$, $c = \hat{c}$ que satisfacen $y_i(\hat{w} \cdot x_i + \hat{c}) \geq 0 \quad \forall (x_i, y_i) \in TS$. Es posible garantizar que $\|(w^*, c^*)\| > \|(\hat{w}, \hat{c})\|$

Solución: *No es posible garantizar lo indicado. Esto pues es posible escalar la ecuación del hiperplano (\hat{w}, \hat{c}) y obtener un separador equivalente.*

- 1.13. Si x_k es un vector de soporte del clasificador (w^*, c^*) , indique si es posible garantizar que $(w^* \cdot x_k + c^*) = 1$

Solución: *No es posible garantizar lo indicado. Si bien x_k es un vector de soporte, la expresión $(w^* \cdot x_k + c^*)$ puede tomar el valor $+1$ o -1 .*

- 1.14. Si el registro (x_k, y_k) está bien clasificado por el clasificador (\hat{w}, \hat{c}) , indique si es posible garantizar que $y_k(\hat{w} \cdot x_k + \hat{c}) \geq 0$

Solución: *Si es posible garantizar lo indicado. Esto pues según lo indicado el registro (x_k, y_k) debe estar en el lado correcto del plano separador, es decir, respetando el rótulo y_k . Por ende, al multiplicar y_k por la ecuación del plano se debe cumplir que $y_k(\hat{w} \cdot x_k + \hat{c}) \geq 0$.*

2. (12 puntos)

a. (6 pts) (Resolución) Demuestre, usando resolución, que

$$\{\neg\forall x \neg P(x), \forall x (P(x) \rightarrow P(f(x)))\} \models \exists x P(f(f(x))).$$

Solución: La demostración es así:

- a.** $P(C)$ (desde Σ y C constante de skolem)
- b.** $\neg P(x) \vee P(f(x))$ (desde Σ)
- c.** $\neg P(f(f(x)))$ (negación de $\exists x P(f(f(x)))$)
- d.** $P(f(C))$ (1 con 2 y x/C)
- e.** $P(f(f(C)))$ (4 con 2 y $x/f(C)$)
- f.** (5 y 3 con x/C)

b. (6 pts) (Diagnóstico) Suponga que se ejecuta el algoritmo de resolución para verificar $\Sigma \models \varphi$, y que el algoritmo termina sin haber generado la cláusula vacía (es decir, se determinó que $\Sigma \not\models \varphi$). Suponga, sin embargo, que durante la ejecución el algoritmo generó la cláusula $\neg Q(C)$, donde C es una constante de skolem. Demuestre que $\Sigma \cup \{\psi\} \models \varphi$, donde ψ es una fórmula no contradictoria que menciona solo al predicado Q . Diga cuál es ese ψ .

Solución: Basta con hacer $\psi = \forall x Q(x)$, así al ingresar esta fórmula en la resolución, quedaría como $Q(x)$ que luego se puede resolver siempre con $\neg Q(C)$. Importante: una respuesta correcta no puede mencionar a C puesto que C es una constante de skolem.

3. (12 puntos)

a. (6 pts) Imagine nuevamente el mundo de grúas que fue presentado en la tarea 2. El siguiente es el código Prolog usado para representar la precondition y efectos de la acción `soltar(G, C, Sup, L)` que es la acción en donde la grúa `G` suelta un container `C` sobre una superficie `Sup`, que está en un lugar `L`.

```
%%% Accion soltar: precondition

poss(soltar(G,C,Sup,L),S) :-
    holds(en(G,L),S),
    holds(en(Sup,L),S),
    holds(levantando(G,C),S),
    holds(despejada(Sup),S),
    grua(G),superficie(Sup),lugar(L).

%%% Accion soltar: efectos positivos

is_positive_effect(soltar(G,C,Sup,L),sobre(C,Sup)).
is_positive_effect(soltar(G,C,Sup,L),despejada(C)).
is_positive_effect(soltar(G,C,Sup,L),disponible(G)).
is_positive_effect(soltar(G,C,Sup,L),en(C,L)).

%%% Accion soltar: efectos negativos

is_negative_effect(soltar(G,C,Sup,L),levantando(G,C)).
is_negative_effect(soltar(G,C,Sup,L),despejada(Sup)).
```

Suponga que una nueva restricción portuaria ha impuesto un límite de 4 al número de containers que pueden estar apilados uno sobre otro.

Para ello, un experto en modelación de problemas sugiere dos modificaciones:

- Agregar el predicado dinámico `containers_bajo`, que especifica el número de containers que hay bajo un cierto container.
- Agregar un parámetro adicional a la acción `soltar`. Este es un parámetro numérico de debe coincidir con el número de bloques bajo `Sup`.

Siguiendo estas recomendaciones, especifique la nueva acción `soltar`, con sus nuevos efectos y precondiciones. ¿Es necesario decir algo adicional en el estado inicial?

Solución:

A la precondition es necesario agregar

```
holds(containers_bajo(Sup,N),S), N=<=3
```

A los efectos es necesario agregar

```
is_positive_effect(soltar(G,C,Sup,L,N),containers_bajo(C,M)) :- M is N+1
```

Es necesario agregar al estado inicial `containers_bajo`, bien definido, para cada container/superficie.

b. (6 pts) Considere el siguiente código Prolog:

```
quehago(L1,_,R) :- member(R,L1).
quehago(L1,L2,do(X,R)) :-
    quehago(L1,L2,R),
    member(X,L2).
```

Describe la lista (infinita) de respuestas que entrega el intérprete frente a la consulta

```
quehago([s0,s1],[a,b,c],X)
```

Solución:

El intérprete responde

```
X = s0 ;  
X = s1 ;  
X = do(a, s0) ;  
X = do(b, s0) ;  
X = do(c, s0) ;  
X = do(a, s1) ;  
X = do(b, s1) ;  
X = do(c, s1) ;  
X = do(a, do(a, s0)) ;  
X = do(b, do(a, s0)) ...
```

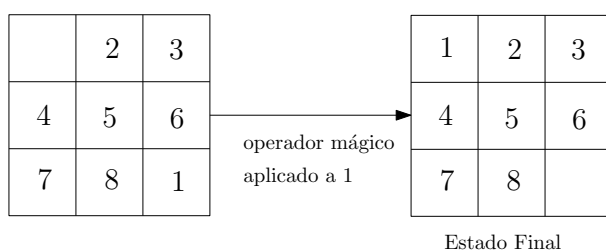
Es decir muestra, en forma alternada, el recorrido de dos árboles de situaciones cuyas raíces son s_1 y s_2 y cuyas acciones son a , b y c .

4. (12 puntos)

“¿Qué es la magia? La magia es una causalidad distinta. Es suponer que, además de las relaciones causales que conocemos, hay otra relación causal.” – Jorge Luis Borges

El *puzle de 8*, tal como fue visto en clases, es un clásico juego en el que se tiene un tablero de 4×4 con 8 fichas numeradas y un espacio vacío. Una movida desplaza una ficha hacia el espacio vacío adyacente. El objetivo es llevar el puzle desde una configuración inicial, hasta el estado final en donde las fichas están en orden (el estado final aparece en la figura).

En esta pregunta consideraremos una variante del puzle de 8: el *puzle mágico de 8*. Este juego funciona tal como el original, pero contiene un operador mágico, que puede ser aplicado N veces y que permite mover cualquier pieza del tablero hacia el espacio vacío. (N es un parámetro del problema.) Observe en la figura al aplicación del operador mágico sobre la ficha 1 hacia el cuadro vacío y así llegar al estado final. El operador mágico es “barato”: su costo es igual al de las otras movidas.



a. (4 pts) Sean

$h_1 = \text{“número de piezas fuera de lugar”}$

$h_2 = \text{“suma de distancias Manhattan de cada ficha hasta su posición final”}.$

Muestre que, para el puzle mágico, h_1 es admisible, mientras que h_2 no lo es.

Solución: h_1 es admisible porque al menos un operador es necesario para mover cada ficha a su posición final en el problema original, por lo que siempre subestimaré el costo de una solución real. h_2 en cambio no es admisible para verlo podemos usar la misma situación mostrada en la figura, en donde h_2 retorna 4 para el estado a la izquierda cuando la solución óptima tiene costo 1.

b. (4 pts) Una alumna intenta modificar una implementación de A* para el puzle de 8 a este nuevo problema. Luego de pensar por unos minutos, alza la voz diciendo: “como es posible que uno llegue al mismo estado de distintas formas, eso hace a este problema bastante distinto”. La mitad del curso no está de acuerdo (o no comprende) la afirmación. ¿Está usted de acuerdo? Discuta.

Solución: La alumna tiene razón en el sentido que es necesario llevar una cuenta de cuántos operadores mágicos se han aplicado dentro de la representación del estado. Esto es necesario porque ese contador debe ser usado para determinar si es posible o no ejecutar más operaciones mágicas.

c. (4 pts) Defina formalmente una heurística admisible para A* que se comporte mucho mejor que h_1 y que dependa de N .

Solución: Dado un estado s , sea n el número de operadores mágicos que aún se pueden aplicar. Calculamos la distancia Manhattan de cada una de las fichas y las ordenamos en una lista, de menor a mayor. Sumamos los $8-n$ primeros elementos. Ahora, por cada una de las n fichas restantes, sumamos 1 por cada elemento que tiene distancia Manhattan mayor que 0.

5. (12 puntos) La figura muestra el resultado del análisis del rendimiento de clasificación de un clasificador SVM (figura a) y una red neuronal (figura b).

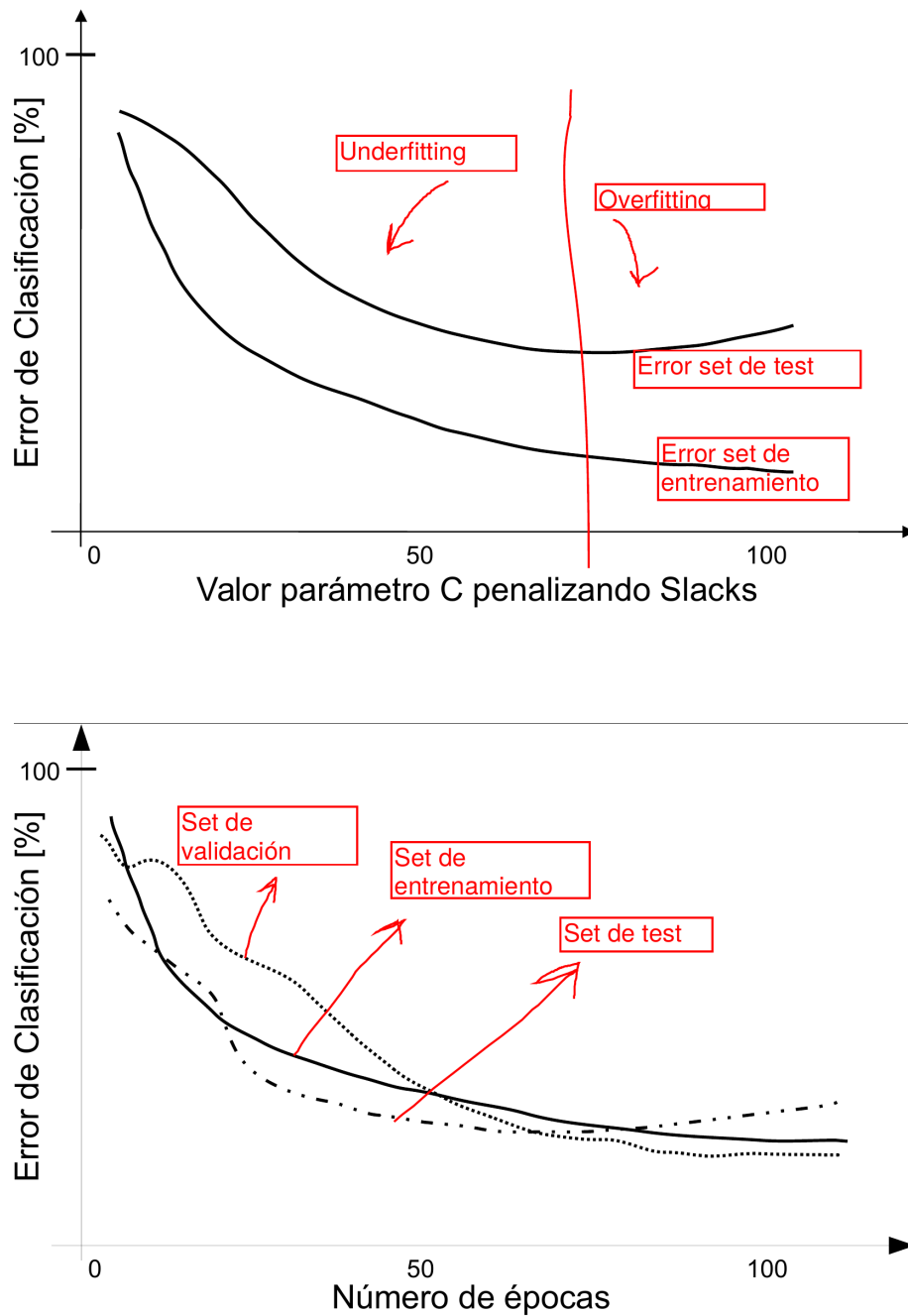


Figure 1: a) Clasificador SVM.

b) Red Neuronal.

a. (3 pts) En la figura a), indique ¿Cuál de las curvas corresponde al rendimiento esperado en el set de entrenamiento y cuál al set de test?. Adicionalmente, para cada curva indique las zonas en que el clasificador no alcanza a ajustar los datos (underfitting) o sobreajusta los datos (overfitting).

Solución: Ver figura a).

b. (3 pts) En la figura b), indique ¿Cuál de las curvas corresponde al rendimiento esperado en el set de entrenamiento, cuál al set de validación, y cuál al set test?.

Solución: Ver figura b). Note que en general es esperable que el error en el set de entrenamiento sea

estrictamente decreciente, no así en set de validación.

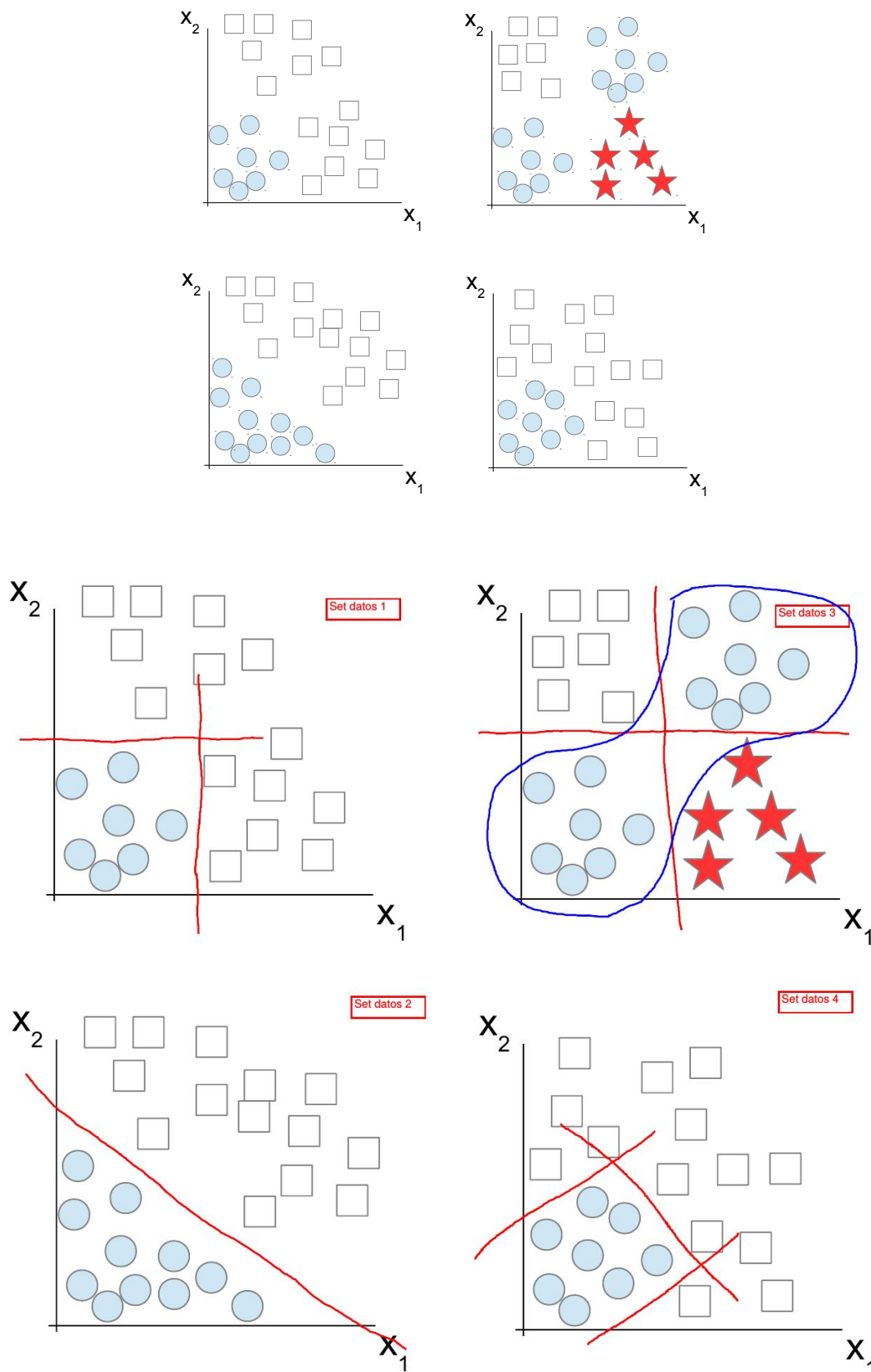
c. (3 pts) Solamente considerando rendimiento de clasificación y que las escalas de ambos gráficos son equivalentes, ¿Cuál clasificador preferiría?. Justifique su respuesta.

Solución: *Según los gráficos es preferible la solución de la red neuronal pues alcanza un menor error en set de test.*

d. (3 pts) Independiente del rendimiento de clasificación, solamente considerando la facilidad de uso de cada clasificador, ¿Cuál clasificador preferiría?. Justifique su respuesta.

Solución: *En términos de facilidad de uso es preferible un SVM pues requiere menos decisiones estructurales y ajuste de hiperparámetros.*

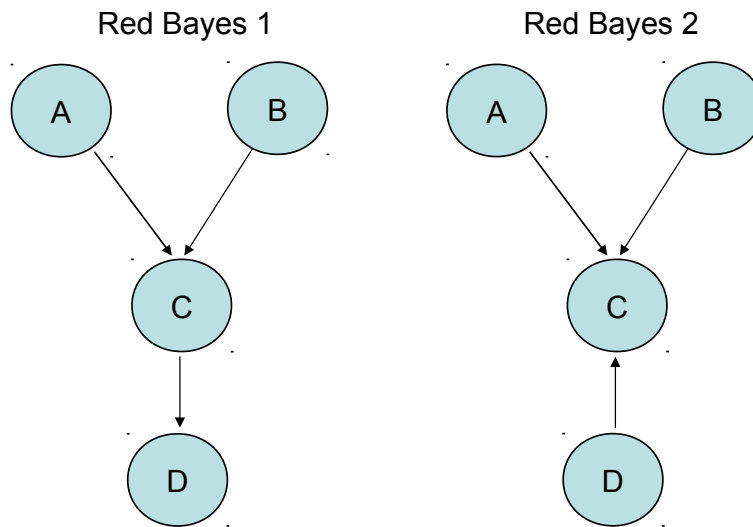
6. (12 puntos) Para cada uno de los set de datos indicados seleccione la red neuronal con menos parámetros que permite separar las clases indicadas. En cada caso justifique brevemente las razones para seleccionar el modelo correspondiente. En su justificación puede utilizar criterios geométricos relacionados con la operación implementada por cada neurona (perceptron) del modelo. (obs: 3 puntos por cada asociación de modelo a gráfico de datos).



Solución:

- Set de datos 1: Red 4. Esta red permite modelar en sus capas ocultas 2 planos separadores que luego pueden ser unidos con un operador tipo “and” (conjunción) en la neurona de salida (ver planos separadores agregados a la figura).
- Set de datos 2: Red 1. En este caso los datos son linealmente separables por tanto pueden ser separados por un perceptron. Dado que el plano separador no pasa por el origen, es necesario tener un bias activo, lo cual implica que la Red 2 no es suficiente para este caso.
- Set de datos 3: Red 5. En esta caso hay 3 clases, las neuronas de la primera capa oculta pueden separar 2 de ellas (planos rojos), pero aun es necesario otro operador que permita clasificar correctamente la clase círculo. Esto es implementado con las neuronas de la parte superior en una segunda capa oculta (ejemplificada en la figura con el lazo azul).
- Set de datos 4: Red 6. Esta red permite modelar en sus capas ocultas 3 planos separadores que luego pueden ser unidos con un operador tipo “and” (conjunción) en la neurona de salida (ver planos separadores agregados a la figura).

7. (12 puntos) Considere las redes de Bayes de la siguiente figura:



a. (3 pts) Indique la factorización de la función de probabilidad conjunta $p(A,B,C,D)$ modelada por la red de Bayes 1.

Solución: $p(A, B, C, D) = p(A)p(B)p(C|A, B)p(D|C)$

b. (3 pts) Indique la factorización de la función de probabilidad conjunta $p(A,B,C,D)$ modelada por la red de Bayes 2.

Solución: $p(A, B, C, D) = p(A)p(B)p(D)p(C|A, B, D)$

c. (3 pts) Indique 2 relaciones de independencia condicional que son válidas para la red de Bayes 1, pero no para la red de Bayes 2.

Solución:

Las relaciones principales son:

- $p(A|C, D) = p(A|C)$. A independiente de D dado C
- $p(B|C, D) = p(B|C)$. B independiente de D dado C

También son válidas las siguientes relaciones:

- $p(A, B|C, D) = p(A, B|C)$. A,B independiente de D dado C
- $p(A|B, C, D) = p(A|B, C)$. A independiente de D dado B,C
- $p(B|A, C, D) = p(B|A, C)$. B independiente de D dado A,C

d. (3 pts) Indique 2 relaciones de independencia condicional que son válidas para la red de Bayes 2, pero no para la red de Bayes 1.

Solución:

Respecto a la red de Bayes 1, las relaciones principales que emergen en la red 2 son:

- $p(D|A) = p(D)$. En ausencia de información sobre C, D independiente de A.
- $p(D|B) = p(D)$. En ausencia de información sobre C, D independiente de B.

También son válidas las siguientes relaciones:

- $p(A|D) = p(A)$. En ausencia de información sobre C, A independiente de D.
- $p(B|D) = p(B)$. En ausencia de información sobre C, B independiente de D.