

Examen – Sistemas Inteligentes (Pauta)

Martes 27 de junio de 2016

Profesor: Alejandro Figueroa

Ayudante: Jean Contreras

- Está prohibido el uso de teléfonos celulares durante el desarrollo de la prueba.
- La prueba debe responderse con un lápiz de tinta indeleble, de lo contrario no hay opción a correcciones.
- Cualquier alumno que sea sorprendido intentando copiar será calificado con una nota 1.
- Está prohibido conversar durante la prueba. Recuerde que su compañero puede estar concentrado y el ruido puede perturbarlo en el desarrollo de su prueba.
- Utilice sólo las hojas entregadas para escribir sus respuestas.
- La nota 4.0 se alcanza con 60 de los 100 puntos que tiene la prueba.

Pregunta 1 (20 puntos)

¿Cuál de los dos siguientes hiperplanos $y=5x$ o $y=-5x$ es más adecuado para un clasificador SVM entrenado en el siguiente data-set?

Feature X	Feature Y	Clase
5	6	P
7	10	P
6	5	P
6	9	P
8	6	P
-5	-5	N
-7	-9	N
-6	-7	N
-8	-8	N
-10	-6	N

Desarrollo:

Utilizando la fórmula de distancia mínima/perpendicular de un punto (x_0, y_0) a una recta $y=ax+b$: $|ax_0 - y_0 + b| / \sqrt{a^2 + 1}$

El denominador para ambas rectas es $\sqrt{26}$ y $b=0$, podemos simplificar el cálculo de la distancia a $|ax_0 - y_0|$ (en términos comparativos). Para ambas rectas, los puntos más cercanos son $(5, 6)$ y $(-5, -5)$. Sin embargo, la distancia

que los separa al plano en mayor cuando se utiliza $y=-5x$ (desde el punto de vista de ambas clases) (10 puntos)

X	Y	$y=5x$	$y=5x$	$y=-5x$	$y=-5x$
5	6	19	3,7262065676	31	6,0796001893
7	10	25	4,9029033785	45	8,8252260812
6	5	25	4,9029033785	35	6,8640647298
6	9	21	4,1184388379	39	7,6485292704
8	6	34	6,6679485947	46	9,0213422164
-5	-5	20	3,9223227028	30	5,8834840541
-7	-9	26	5,0990195136	44	8,6291099461
-6	-7	23	4,5106711082	37	7,2562970001
-8	-8	32	6,2757163244	48	9,4135744866
-10	-6	44	8,6291099461	56	10,982503568

(10 puntos)

Pregunta 2 (20 puntos)

Suponga que un algoritmo genético¹ usa cromosomas de largo fijo (8) y de la forma $x=abcdefgh$. Cada gen puede ser un dígito entre cero y nueve. La función de fitness de un individuo x se calcula mediante la siguiente expresión:

$$f(x)=(a+b)-(c+d)+(e+f)-(g+h)$$

Consideremos además que la población inicial consiste de los siguientes cuatro individuos:

$x_1 = 6\ 5\ 4\ 1\ 3\ 5\ 3\ 2$

$x_2 = 8\ 7\ 1\ 2\ 6\ 6\ 0\ 1$

$x_3 = 2\ 3\ 9\ 2\ 1\ 2\ 8\ 5$

$x_4 = 4\ 1\ 8\ 5\ 2\ 0\ 9\ 4$

a) Calcule el fitness de cada uno de esos individuos, mostrando todos los pasos, y ordénelos del mejor al peor. (4 puntos)

$$f(x_1) = (6 + 5) - (4 + 1) + (3 + 5) - (3 + 2) = 9$$

$$f(x_2) = (8 + 7) - (1 + 2) + (6 + 6) - (0 + 1) = 23$$

$$f(x_3) = (2 + 3) - (9 + 2) + (1 + 2) - (8 + 5) = -16$$

$$f(x_4) = (4 + 1) - (8 + 5) + (2 + 0) - (9 + 4) = -19$$

El orden es x_2 , x_1 , x_3 y x_4 .

b) Realice las siguientes operaciones (4 puntos):

b.1) Cruce los dos individuos más aptos utilizando el cruzamiento de un punto al medio.

$X_1 = 8712 \mid 6601$

$x_2 = 6541 \mid 3532$

1 <http://www.eis.mdx.ac.uk/staffpages/rvb/teaching/BIS3226/sol15.pdf>

O1 = 87123532

O2 = 65416601

b.2) Cruce el segundo y el tercer individuo utilizando un cruzamiento de dos puntos (lugares b y f).

x1= 65|4135|32
X3 = 23|9 2 1 2| 8 5

O3= 65 9212 32

O4= 23 4135 85

c) Suponga que la nueva población consiste de seis hijos producidos por un operador de crossover. Evalúe el fitness de cada individuo mostrando todos los pasos ¿Mejóro el fitness general de la población? (4 puntos)

O1 = 8 7 1 2 3 5 3 2

O2 = 6 5 4 1 6 6 0 1

O3 = 6 5 9 2 1 2 3 2

O4 = 2 3 4 1 3 5 8 5

O5 = 2 7 1 2 6 2 0 1

O6 = 8 3 9 2 1 6 8 5

El fitness general de la población mejoro:

$$f(O1) = (8 + 7) - (1 + 2) + (3 + 5) - (3 + 2) = 15$$

$$f(O2) = (6 + 5) - (4 + 1) + (6 + 6) - (0 + 1) = 17$$

$$f(O3) = (6 + 5) - (9 + 2) + (1 + 2) - (3 + 2) = -2$$

$$f(O4) = (2 + 3) - (4 + 1) + (3 + 5) - (8 + 5) = -5$$

$$f(O5) = (2 + 7) - (1 + 2) + (6 + 2) - (0 + 1) = 13$$

$$f(O6) = (8 + 3) - (9 + 2) + (1 + 6) - (8 + 5) = -6$$

d) Mediante la inspección de la función de fitness y considerando que los genes sólo pueden tomar valores entre cero y nueve, encuentre el cromosoma que representa la solución óptima (4 puntos).

$$x_{\text{optimal}} = 9 \ 9 \ 0 \ 0 \ 9 \ 9 \ 0 \ 0$$

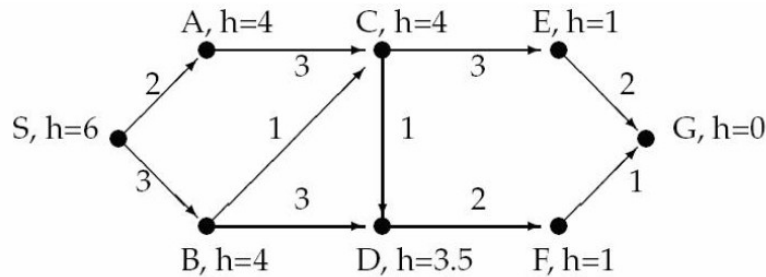
$$f(x_{\text{optimal}}) = (9 + 9) - (0 + 0) + (9 + 9) - (0 + 0) = 36$$

e) Mediante la inspección de la población inicial ¿Podría ud. decir si es posible alcanzar el óptimo sin el operador de mutación? (4 puntos)

No, el algoritmo nunca alcanzaría la solución óptima sin el operador de mutación. Dado que la única forma de cambiar genes es el cruzamiento. Independiente del tipo de cruzamiento, nunca sería posible que apareciera un "9" en la población, ergo alcanzar el óptimo.

Pregunta 3 (20 puntos)

Suponga² que queremos utilizar A* en el siguiente gráfico para encontrar el camino más corto desde el nodo S al nodo G. Cada nodo está etiquetado por una letra mayúscula y el valor de una función heurística. Cada arco está etiquetado por el costo para atravesar el arco. Para este problema: Desarrolle A* en el grafo llenando la tabla adjunta. Indique los valores para f, g y h de cada nodo en la cola en orden. Si encuentra un camino a un nodo que ya está en la cola,



actualice su costo. Muestre el camino más corto encontrado por este algoritmo.

Iteración	Nodo expandido	Cola de prioridad al final de esta iteración
0		$S = 0 + 6 = 6$ (i.e. $S = g(S) + h(S) = f(S)$)
1	S	$A = 2 + 4 = 6$; $B = 3 + 4 = 7$
2	A	$C=9$, $B=7$
3	B	$C=8$, $D=9.5$
4	C	$D=8.5$, $E=8$
5	E	$G=9$, $D=8.5$
6	D	$F=8$, $G=9$
7	F	$G=8$
8	G	

(2,5 puntos cada paso)

Pregunta 4 (15 puntos)

Considere³ un dataset representados con features uni-dimensionales con valores: {2,4,7,8,12,14}. Al final de las tres primeras iteraciones de clustering aglomerativo (distancia euclidiana) obtenemos los siguientes tres clústeres: {2,4}, {7,8} y {12,14} ¿Cuál es la distancia

2 <http://www.cs.virginia.edu/~cs416/2004Fall/Tests/2004midtermAnswers.pdf>

3 <http://pages.cs.wisc.edu/~dyer/cs540/exams/exam1-summer16-sol.pdf>

entre estos tres clústeres utilizando Single Link y Complete Link? (10 puntos) ¿Cuáles son los clústeres formados después de la próxima iteración usando Single Link? (5 puntos)

Single Link (5 puntos):

$$d(\{2,4\}, \{7,8\}) = 7 - 4 = 3$$

$$d(\{2,4\}, \{12,14\}) = 12 - 4 = 8$$

$$d(\{7,8\}, \{12,14\}) = 12 - 8 = 4$$

Complete Link (5 puntos):

$$d(\{2,4\}, \{7,8\}) = 8 - 2 = 6$$

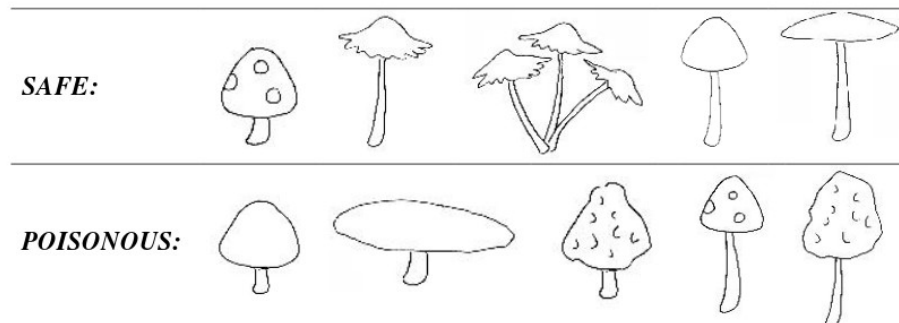
$$d(\{2,4\}, \{12,14\}) = 14 - 2 = 12$$

$$d(\{7,8\}, \{12,14\}) = 14 - 7 = 7$$

El par de clústeres más cercanos es $\{2,4\}$ y $\{7,8\}$ con distancia tres. Estos son los clústeres a unir: $\{2,4,7,8\}$ y $\{12,14\}$. (5 puntos)

Pregunta 5 (25 puntos)

Considere⁴ la tarea de aprender a identificar hongos que son seguros (safe) o venenosos (poisonous) para comer basándose en un conjunto de features físicos. Utilice cuatro features: Tronco= {corto, largo}, campana= {redonda, aplanada}, textura= {plana, puntos, dispareja,



rugosa} y número= {única, múltiple}. Considere el siguiente conjunto de entrenamiento:

¿Cómo clasificaría un clasificador bayesiano el siguiente ejemplo? (5 puntos)

⁴ <http://pages.cs.wisc.edu/~bsettles/cs540/exams/final-soln.pdf>



Sistemas Inteligentes - 1er semestre 2017 - Solemne 2

Seguro es la clase más probable:

$$P(\text{safe}) \times P(\text{long} | \text{safe}) \times P(\text{flat} | \text{safe}) \times P(\text{spots} | \text{safe}) \times P(\text{single} | \text{safe}) =$$

$$(.5)(.8)(.6)(.2)(.8) = 0.0384$$

$$P(\text{poison}) \times P(\text{long} | \text{poison}) \times P(\text{flat} | \text{poison}) \times P(\text{spots} | \text{poison}) \times P(\text{single} | \text{poison}) =$$

$$(.5)(.4)(.2)(.2)(1.0) = 0.008$$

Utilizando la ganancia de información entre textura y número dibuje el árbol de decisión que da al comparar ambos features. (10 puntos)

InfoGain(TEXTURE)

$$= \text{Entropy}(S) - (.4)\text{Entropy}(S_{\text{plain}}) - (.2)\text{Entropy}(S_{\text{spots}}) - (.2)\text{Entropy}(S_{\text{bumpy}}) - (.2)\text{Entropy}(S_{\text{ruffles}})$$

$$= 1 - (.4)(0) - (.2)(0) - (.2)(1) - (.2)(1)$$

$$= 0.4$$

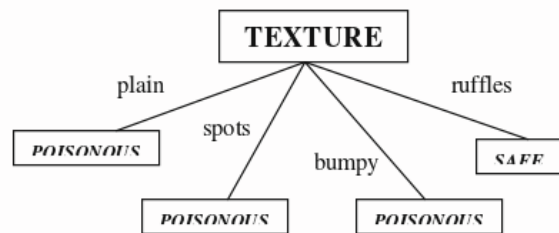
InfoGain(NUMBER)

$$= \text{Entropy}(S) - (.9)\text{Entropy}(S_{\text{single}}) - (.1)\text{Entropy}(S_{\text{multiple}})$$

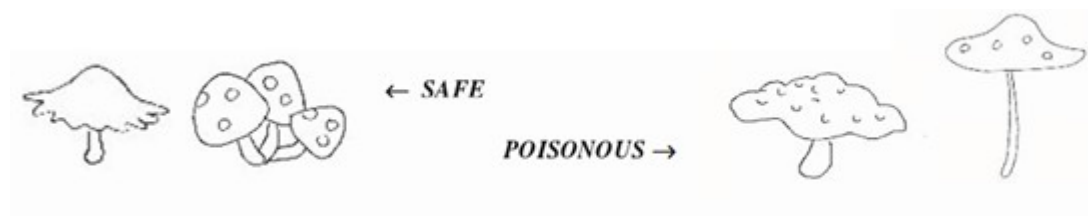
$$= 1 - (.9)(.991) - (.1)(1)$$

$$= 0.108$$

TEXTURE is the winner:



Estime la accuracy del árbol de decisión resultante utilizando el siguiente conjunto de prueba (10 puntos):



La accuracy es 75% ya que clasifica mal el segundo ejemplo seguro.