## **Pregunta 1**

**¿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(x0,y0) a una recta y=ax+b: |ax0-y0+b|/sqrt(a\*a+1)

El denominador para ambas rectas es sqrt(26) y b=0, podemos simplificar el cáculo de la distancia a |ax0-y0| (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) 

## **Pregunta 2**

**Suponga que un algoritmo genético[[1]](#footnote-1) 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:**

**x1= 6 5 4 1 3 5 3 2**

**x2= 8 7 1 2 6 6 0 1**

**x3= 2 3 9 2 1 2 8 5**

**x4= 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.**

f(x1) = (6 + 5)−(4 + 1) + (3 + 5)−(3 + 2) = 9

f(x2) = (8 + 7)−(1 + 2) + (6 + 6)−(0 + 1) = 23

f(x3) = (2 + 3)−(9 + 2) + (1 + 2)−(8 + 5) = −16

f(x4) = (4 + 1)−(8 + 5) + (2 + 0)−(9 + 4) = −19

El orden es x2, x1, x3 y x4.

**b) Realice las siguientes operaciones:**

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

X1= 8712 | 6601

x2 = 6541 | 3532

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. Evalué el fitness de cada individuo mostrando todos los pasos ¿Mejoró el fitness general de la población?**

**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.**

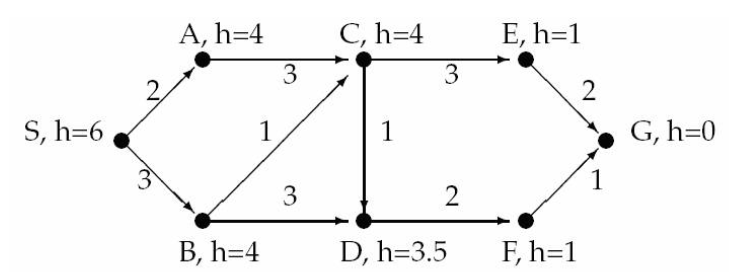
xoptimal= 9 9 0 0 9 9 0 0

f(xoptimal) = (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?**

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**

**Suponga[[2]](#footnote-2) 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 |  |

## **Pregunta 4**

**Considere[[3]](#footnote-3) 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 entre estos tres clústeres utilizando Single Link y Complete Link? ¿Cuáles son los clústeres formados después de la próxima iteración usando Single Link?**

Single Link:

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:

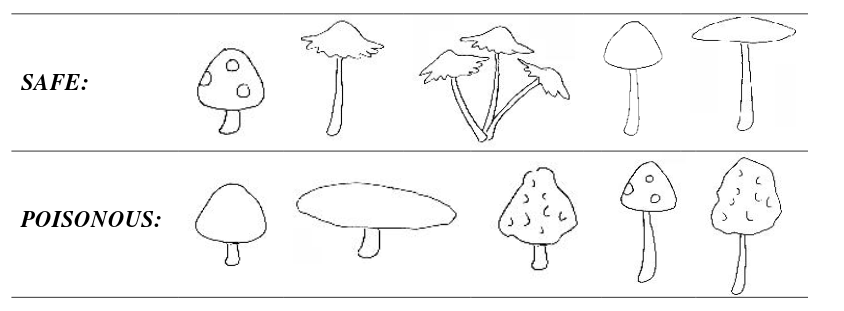
d({2,4},{7,8}) = 8 – 2 = 6

d({2,4},{12,14}) = 14 – 2 = 12

d({7,8), {12,14}) = 14 – 7 = 7

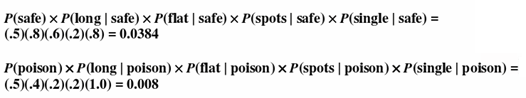
El par de clústeres más cercanos es {2,4} y {7,8} con distancia tres. Estos son los clústeres a unir: {2,4,7,8} y {12,14}.

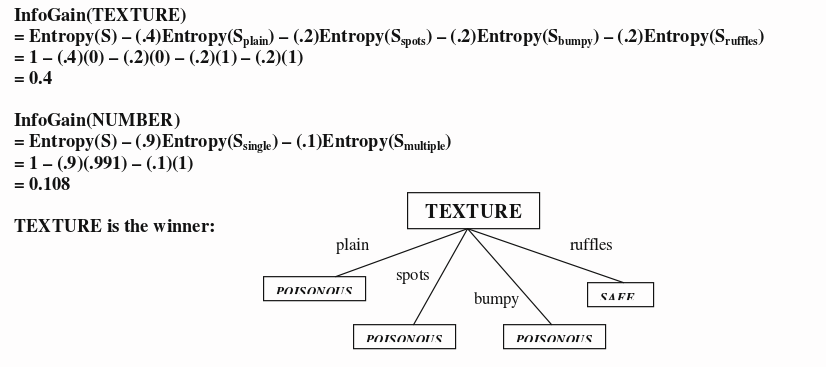
## **Pregunta 5**

**Considere[[4]](#footnote-4) 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:**

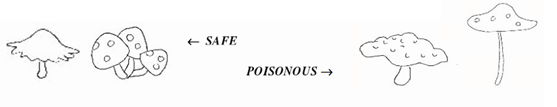
**¿Cómo clasificaría un clasificador bayesiano el siguiente ejemplo?**

Seguro es la clase más probable:

****

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

**Estime la accuracy del árbol de decisión resultante utilizando el siguiente conjunto de prueba:**

****

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

1. [http://www.eis.mdx.ac.uk/staffpages/rvb/teaching/BIS3226/sol15.pdf](# http://www.eis.mdx.ac.uk/staffpages/rvb/teaching/BIS3226/sol15.pdf) [↑](#footnote-ref-1)
2. [http://www.cs.virginia.edu/~cs416/2004Fall/Tests/2004midtermAnswers.pdf](# http://www.cs.virginia.edu/~cs416/2004Fall/Tests/2004midtermAnswers.pdf) [↑](#footnote-ref-2)
3. [http://pages.cs.wisc.edu/~dyer/cs540/exams/exam1-summer16-sol.pdf](# http://pages.cs.wisc.edu/~dyer/cs540/exams/exam1-summer16-sol.pdf) [↑](#footnote-ref-3)
4. [http://pages.cs.wisc.edu/~bsettles/cs540/exams/final-soln.pdf](# http://pages.cs.wisc.edu/~bsettles/cs540/exams/final-soln.pdf) [↑](#footnote-ref-4)