



Solemne 2 - Sistemas Inteligentes

viernes 24 de junio de 2015

Profesor: Alejandro Figueroa

Ayudante: Alexander Espina

- 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.

Sección 1 (40 puntos)

Indique si cada una de las siguientes aseveraciones es verdadera(V) o falsa(F). No es necesario justificar las falsas. Cada respuesta correcta vale 2 punto, mientras que las incorrectas descuentan 1 puntos.

1. ..F.. K-Means termina siempre en la misma solución cada vez que se ejecuta.
2. ..F.. El centroide de los siguientes puntos (7,8) (4,3) (5,6) (4,3) es: (16,17).
3. ..V.. En un algoritmo de aprendizaje no-supervisado, es altamente probable que se obtenga un resultado o modelo distinto si cambiamos los features.
4. ..V.. En K-Means puede quedar varios clústeres vacíos al final del proceso iterativo.
5. ..V.. K-Means es simple, eficiente, fácil de entender y de implementar.
6. ..F.. Aprendizaje semi-supervisado utiliza un gran número de datos etiquetados y no-etiquetados.
7. ..F.. Las estrategias aglomerativas de agrupamiento va dividiendo el conjunto de datos en cada iteración.
8. ..F.. Complete-link se caracteriza por tener problemas con los datos ruidosos.
9. ..V.. Con clustering aglomerativo se puede inspeccionar cualquier nivel de granularidad en los datos.
10. ..F.. Las técnicas de escalamientos se utilizan cuando los datos tienen una alta dimensionalidad.
11. ..F.. La distancia de Chebychev entre (1,2) y (2,3) es 2.
12. ..V.. El Rand Index está en el intervalo [0,1] y toma en cuenta todas las celdas de la matriz de confusión binaria respectiva en su cálculo.
13. ..V.. La idea detrás del Índice de Dunn es que los elementos dentro de un clúster deben estar más cerca, y que los centroides de estos clústeres deben estar lejanos.
14. ..F.. Cuando hacemos el 2-opt tenemos al menos una solución nueva para explorar.
15. ..V.. 3-opt tiene, en general, mejor desempeño que 2-opt en términos de calidad de solución.
16. ..V.. Una forma de lidiar con múltiples funciones objetivos es ponderarlas en una sola.
17. ..F.. ACS refuerza la feromona con los arcos pertenecientes a la mejor solución encontrada en cada iteración.
18. ..V.. En ACS, las hormigas ven un plano uniforme de feromona al iniciar la búsqueda.
19. ..V.. ANT-Q tiene la propiedad de que utiliza una regla de específica para la actualización local.
20. ..V.. ACS es un algoritmo colaborativo.



Sección 2 (30 puntos)

- 1. Considere los siguientes puntos: (0.1, 100) y (0.3,200); explique el problema de estandarización de datos considerando las métricas de distancia Minkowsky (h=2) y Canberra. (15 puntos)**

Distancias sin estandarizar:

Minkowsky: $((100-0.1)+(200-0.3))^{1/2}=(10000+0.04)^{1/2}= 100,0002$

Canberra: $\frac{|100-200|}{(100+200)} + \frac{|0.1-0.3|}{(0.1+0.3)}=100/300+0.2/0.4=0.3333+0.5=0.83333$

El problema es que la distancia de Minkowsky está sesgada a la componente de mayor magnitud. La contribución de las componentes de menor magnitud es marginal. Esto no sucede con la distancia Canberra, ya que efectúa la división por la suma que actúa como una especie de normalizador. En la distancia canberra vemos que ambas componentes gravitan de forma significativa en la distancia final.

Al standarizar la distancia Minkowsky, en particular la primera componente entre 0 y 1000, tenemos: $((0.1-0.1)+(0.2-0.3))^{1/2}=(0.01+0.04)^{1/2}= 0,2236$. Logrando un efecto similar a la distancia Canberra.

- 2. Un buscador retorna siempre buenos resultados en sus top 10 resultados. Por ejemplo, para un conjunto de diez consultas nueve veces devuelve un documento relevante en la primera posición y en la segunda. Y en la situación restante, devuelve un documento relevante en la tercera y cuarta posición. ¿Cuál es el MRR de este buscador? (15 puntos)**

Las nueve veces que devuelve un documento en la posición uno, aportan 9 unidades. La situación excepcional aporta $\frac{1}{2}$. Por ende, el MRR es $9.3333/10$, lo que da 0.93333.

Sección 3 (30 puntos)

- 1. Enuncie una función de acotamiento para el vendedor viajero con ventanas de tiempo. (15 puntos)**

Dado el principio que al remover restricciones encontraremos cotas superiores al problema original, podemos resolver el problema del vendedor viajero original para encontrar una función de acotamiento. Una vez relajado el problema, podemos utilizar el principio de que tenemos que salir de cada ciudad al menos una vez para encontrar la cota, es decir relajando la restricción que debemos visitar todas las ciudades.

2. Cree ud. que los Algoritmos Genéticos hacen inferencias. Diga explícitamente sí o no, y argumente claramente su posición. Nótese que un “sí” o un “no” sin argumentación sólida no recibirá puntaje. (15 puntos)

Sí, porque a medida que va pasando las iteraciones, el algoritmo genético va encontrando patrones que tienen una alta probabilidad de pertenecer a la mejor solución. La población empieza a ser dominada por patrones que son eficientes en la resolución del problema. Esos patrones son inferidos por el proceso de búsqueda. De la misma forma que la teoría de la evolución natural postula que los seres humanos evolucionamos o “mejoramos la raza”. Atributos mejores o que se adaptan mejor son encontrados (inferidos) durante el proceso.