

## Solemne 1 – Sistemas Inteligentes (Pauta)

Martes 11 de octubre de 2016

**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.
- La nota 4.0 se alcanza con 60 de los 100 puntos que tiene la prueba.

### Pregunta 1 (20 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 1 punto, mientras que las incorrectas descuentan 0.5 puntos. Letras que no sean nítidas serán consideradas malas.**

1. ..F.. SVM clasifica siempre un punto caído en el margen separador como negativo.
2. ..V.. Si tengo un conjunto de buenos (positivo) y malos (negativo) alumnos caracterizados por atributos físicos, al aprender un clasificador SVM, el modelo resultante etiquetará a un alumno con PGA alto como buen alumno.
3. ..F.. En SVM, entre más lejano está un punto del hiper-plano divisor, más representativo es de su clase.
4. ..F.. En SVM, los datos no-etiquetados son inútiles para aprender un modelo.
5. ..V.. El objetivo de la curva de aprendizaje es cuantificar la mejora en la tasa de clasificación con respecto al aumento de los datos de entrenamiento.
6. ..F.. Las probabilidades prior en un clasificador Bayesiano van a ser siempre las mismas antes y después de hacer oversampling.
7. ..F.. Si se entrena un clasificador SVM y uno Bayesiano con el mismo conjunto de datos, los mismos atributos, prediciendo las mismas clases, entonces ambos obtendrán los mismos resultados.
8. ..V.. El área por sobre la curva ROC es un indicador del error de un clasificador.
9. ..F.. Si un sistema obtiene un MRR de uno para un conjunto de prueba, estamos seguros que su F1-Score será de uno también.
10. ..V.. Al utilizar MRR, el score mínimo que puede tener un clasificador en un conjunto de prueba es el inverso del número de clases a predecir.
11. ..V.. La profundidad máxima de un árbol de decisión está dada por el número de atributos.
12. ..V.. En KNN, el utilizar todos los puntos como vecinos, tenemos un clasificador de clase mayoritaria.
13. ..V.. En KNN, la etiqueta de un ejemplo de prueba puede cambiar abruptamente a pesar de que el número de vecinos se aumente suavemente.
14. ..F.. En KNN, es irrelevante la métrica de distancia utilizada.
15. ..F.. En clustering aglomerativo, vamos partiendo el conjunto de datos sistemáticamente.
16. ..V.. En clustering jerárquico gran escala, los nodos hojas pueden representar grupos de datos individuales.
17. ..V.. En DBSCAN, no es necesario especificar el número de clústeres apriori.
18. ..F.. DBSCAN es una estrategia de clustering particional.
19. ..F.. Si el purity de una asignación de clustering aumenta, la entropía también sube.

20. ..V.. En la práctica, K-Means es un algoritmo de complejidad que crece linealmente con el número de datos.

**Pregunta 2 (40 puntos)**

**1. Compare DBSCAN y ST-DBSCAN (4 similitudes y 3 diferencias). (14 puntos)**

DBSCAN	ST-DBSCAN
Utiliza el concepto de vecindad de un punto para organizar sus clústeres.	Utiliza el concepto de vecindad de un punto para organizar sus clústeres.
Tiene una vista de los datos. Ergo, un epsilon y un minPts.	Tiene dos vista de los datos. Ergo, dos epsilon y dos minPts.
No detecta variaciones en la densidad de los clústeres. Lo que lo hace más susceptible a mezclar clústeres que traslapan.	Puede detectar algunas variaciones en la densidad de los clústeres. Lo que lo hace menos susceptible a mezclar clústeres que traslapan.
Utiliza una métrica de distancia.	Puede utilizar métricas de distancia diferentes para cada vista.
Ambos tienen la desventaja de tener que determinar qué puntos caen en la vecindad de un punto.	Ambos tienen la desventaja de tener que determinar qué puntos caen en la vecindad de un punto.
Puede detectar ruido.	Puede detectar ruido.
La métrica de distancia es flexible.	Las métricas de distancia son flexibles.

**2. Compare Single-Link con Complete-Link Ejemplifique ambos grupos. La pauta contempla cuatro diferencias y/o similitudes. (16 puntos)**

Single-Link	Complete -Link
La distancia entre dos clústeres es la distancia de los dos puntos más cercanos.	Une dos clústeres tal que su distancia máxima es la mínima entre todos los clústeres.
Sensible a datos ruidosos	No es sensible a datos ruidosos
Complejidad $O(n)$	Complejidad $O(n^2 \log n)$
No tiene problemas con outliers.	Problemas con outliers.

**3. Considere el problema asignado en las tareas. Diseñe tres estrategias que servirían para combinar los resultados de los diferentes modelos examinados, y que al mismo tiempo, podrían tener el potencial para mejorar los resultados. Nota piense en cosas simples, sin involucrar entrenamientos adicionales. (10 puntos)**

- La etiqueta que asigna la mayoría de los clasificadores.
- Escoger la etiqueta que recibió el puntaje más alto, independiente de si fue mayoritaria o del clasificador. Después de normalizar los puntajes.

c) Normalizar los valores de los puntajes, promediarlos, y asignar la que tenga el promedio más alto.

d) Asignar pesos a cada uno de los clasificadores de acuerdo a su desempeño (en el mismo set de entrenamiento o uno de validación), y después utilizarlos para sopesar sus puntajes normalizados. Asignando el que de más alto en la sumatoria ponderada final.

### Pregunta 3 (40 puntos)

A continuación, se presenta un ejemplo de matriz de confusión generada en la descripción de una patente de un sistema de reconocimiento de emociones<sup>1</sup>:

	Angry	bored	disgusted	Afraid	blissful	sad	neutral
angry	80.31	0.79	3.15	3.94	8.66	0	3.15
bored	0	66.67	6.17	2.47	0	7.41	17.28
disgusted	3.15	17.39	52.17	4.35	4.35	10.87	2.17
afraid	10.24	4.35	5.80	40.58	4.35	10.14	15.94
blissful	18.11	0	4.23	5.63	52.11	0	5.63
sad	0	11.29	0	1.61	0	82.26	4.85
neutral	2.36	17.71	5.06	7.59	0	2.53	63.29

Asuma que las columnas corresponden a la predicción de emoción, y las filas corresponden a las emociones observadas. Los datos están en porcentajes y asuma que son resultados después de hacer undersampling:

*Al ser clases aproximadamente balanceadas en número, se asume que los porcentajes son respecto de esa clase. Una manera muy sencilla de verlo es que cada clase correspondientes a las emociones observadas son equivalentes a 100 datos (total 700). Luego, los porcentajes de cada celda, corresponden a porcentajes de cada emoción observada (sobre ese 100). Con ello cada fila suma 100%, y las columnas no necesariamente.*

**1. Calcule Recall para la clase “bored” (5 pts)**

$$\text{Recall} = 66.67 / (0 + 66.67 + 2.47 + 0 + 7.41 + 17.28) = 66.67 / 100 = 0.66$$

**2. Calcule Precisión para la clase “afraid” (5 pts)**

$$\text{Precision} = 40.58 / (3.94 + 2.47 + 4.35 + 40.58 + 5.63 + 1.61 + 7.59) = 40.58 / 66.17 = 0.613$$

**3. ¿Es representativo el valor de accuracy en esta matriz de confusión? Justifique en base a los datos mostrados. (15 pts)**

<sup>1</sup> Bollano, Gianmario, Donato Ettore, and Antonio Esiliato. "Method and system to improve automated emotional recognition." U.S. Patent Application No. 12/449,298.

Asumiendo que los datos están correctamente balanceados entre clases, se puede asumir. En la respuesta se debe incluir alguna justificación parecida, ya que se asume que los porcentajes son respecto del total del dataset.

**4. ¿Cuál es la clase que tiene una mayor tasa de falsos positivos? ¿En qué afecta este hecho al cálculo del valor de la precisión respectiva para esta clase? (15 pts)**

TFP = FP (suma de la columna de la clase, descartando la diagonal) / (FP+VN [suma de los elementos de la diagonal, descartando el valor VP])

	Sum FP	Sum VN	TFP [FP/(FP+VN)]
Angry	33.86	357.08	0,087
Bored	51.53	370.72	0,122
Disgusted	24.41	385.22	0,060
Afraid	25.59	396.81	0,061
Blissful	17.36	385.28	0,043
Sad	30.95	355.15	0,080
Neutral	49.02	374.1	0,116

La clase que tiene una mayor TFP es Bored. Afecta el valor de la precisión en que al ser una tasa alta, es (entre otros factores), a un alto número de falsos positivos. Si bien eso puede ser sopesado con la suma entre verdaderos negativos y falsos positivos (es decir, todos los que no fueron predichos como la clase), en este caso afecta la precisión, pues depende del valor de los falsos positivos por construcción del indicador. Conclusión: a mayor TFP, menor precisión.