Universidad Rafael Landívar

Inteligencia Artificial

Primer Semestre 2025

Hoja de Trabajo 9

Descripción

Lea detenidamente cada enunciado de la Hoja de Trabajo y resuelva cada uno de los incisos según se le indique.

Serie 1 (10 puntos)

Indique si existe la relación de independencia condicional en la red de Bayes a continuación:

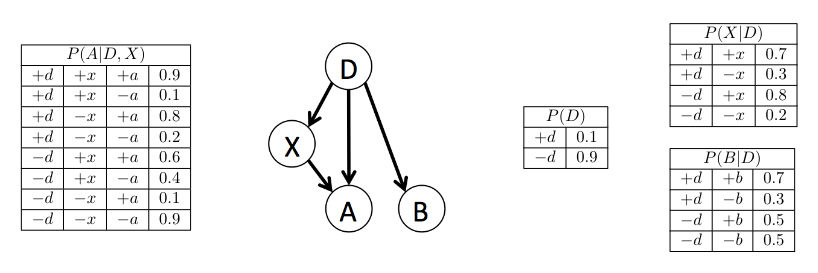
Forma, Círculo

El contenido generado por IA puede ser incorrecto.

* U ⊥ X
* U ⊥ X|T
* V ⊥ W|Y
* V ⊥ W|T

Serie 2 (30 puntos)

La siguiente red bayesiana representa la ejecución de dos pruebas A y B que comprueban la existencia o ausencia de una enfermedad D. Así mismo en pruebas de campo se ha observado que existe un factor ambiental X que afecta la sensibilidad de A.



Utilizando eliminación de variables, responda:

1. ¿Cuál es la probabilidad de tener la enfermedad D y obtener resultado positivo en A?
2. ¿Cuál es la probabilidad de tener la enfermedad D si se obtuvo un resultado positivo en A?
3. ¿Cuál es la probabilidad de tener la enfermedad D si se obtuvo un resultado positivo en B?

Serie 3 (30 puntos)

Ha decidido usar una red bayesiana simple para analizar la probabilidad de que disfrute su experiencia universitaria en la URL. Sin embargo, esto depende de si sus amigos regresaron a la URL y si tiene alguna clase temprano en la mañana. Sabe que la probabilidad de que sus amigos regresen (evento F=+f) es 0.8 y la probabilidad de que tenga que inscribirse en una clase matutina (evento M=+m) es 0.7. La variable L con el valor +l representa si está disfrutando la experiencia universitaria y -l si no la disfruta.

Tabla

El contenido generado por IA puede ser incorrecto.

1. Cuál es la fórmula para la distribución de probabilidad conjunta inducida por la red bayesiana anterior.
2. Calcule la probabilidad de que sus amigos regresen y no tenga que tomar ninguna clase en la mañana.
3. Calcule la probabilidad de que sus amigos regresen dado que le está gustando el semestre.

Serie 4 (20 puntos)

Queremos predecir si un cliente comprará (y=1) o no comprará (y=0) un producto basándonos en una sola característica:

* Ingresos (*X*): Monto de ingresos mensuales del cliente.

Los datos son:

* *β*0​=−58.0264
* *β*1​=0.0116

|  |  |
| --- | --- |
| INGRESOS | COMPRA |
| 2000 | 0 |
| 4000 | 0 |
| 6000 | 1 |
| 8000 | 1 |

Aplicando Regresión Logística:

1. Calcule los log-odds (*z= β0+ β1X*) para cada dato.
2. Aplique la función sigmoide para obtener las probabilidades predichas
3. Haga predicciones para X={3000, 5000, 7000}, usando un umbral de 0.5:
   1. Si P≥0.5, predecimos *y=1*.
   2. Si P<0.5, predecimos *y=0*.

Serie 5 (10 puntos)

Al implementar un clasificador Naïve Bayes para filtrar correos electrónicos (1: spam, 0: no spam), se obtuvieron diferentes matrices de confusión en el conjunto de prueba. Considerando que el modelo debe maximizar la **precisión** para spam (evitar falsos positivos) y mantener buen **recall** (capturar la mayoría del spam), ¿Cuál de las siguientes configuraciones muestra el mejor equilibrio requerido, aplicando los supuestos que caracterizan a Naïve Bayes?

Imagen que contiene Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

1. ¿Cuál de los modelos (A, B, C o D) muestra el mejor equilibrio entre **evitar falsos positivos** (alta precisión para spam) y **capturar la mayoría del spam** (buen recall), según el supuesto de independencia de Naïve Bayes?
2. ¿Cuál de los siguientes modelos detecta **más spam** (mayor número de verdaderos positivos), pero a costa de **más errores graves** (falsos positivos)?