

Instrucciones

- La tarea debe ser subida en Aula el día **Domingo 31** de Octubre.
- Se debe adjuntar un pdf y un notebook con el código de los ejercicios (también puede poner las respuestas en formato latex en el mismo notebook).
- La evaluación será centrada en su informe escrito, sin embargo la presentación de código elegante, documentado y leíble puede darle algunos puntos extras.
- Explique claramente cada una de las conclusiones, incluyendo gráficos y referencias cuando sea necesario.

Pregunta 1 (50%): Programando una red bayesiana

En este ejercicio crearemos una red bayesiana y realizaremos inferencia sobre ella. Para esto se le recomienda utilizar alguna librería que automatice el proceso. Aunque usted puede elegir el lenguaje y librería, se le recomienda trabajar con pgmpy (puede encontrar notebooks muy útiles en este link). Además se estudiará en ayudantía.

Creación de la red

Para crear la red utilizaremos los datos historicos adjuntos de evaluaciones de riesgo de créditos. Esta evaluación se basa tanto en características del cliente cómo del crédito.

1. Para comenzar estudie el método de aprendizaje de estructuras de redes Hill-Climbing y explíquelo brevemente aquí.
2. Utilice este método (utilice BICScore) para aprender una estructura para la red y grafíquela.
3. Comente sobre la estructura obtenida y proponga una mejora, cree la red utilizando pgmpy y luego grafíquela.
4. Aprenda los parámetros de cada CPD de la red utilizando el método de máxima verosimilitud (utilizando el método MaximumLikelihoodEstimator).

Independencias

Liste todas las independencias locales y todas las independencias markovianas globales para la variable Riesgo.

Consultas

Conteste las siguientes consultas:

1. Dado que conocemos que la edad del cliente es 21, ¿cuál es la probabilidad de riesgo alto?
2. Liste la probabilidad condicional de riesgo dado evaluación crediticia histórica.
3. Dado que el crédito es de duración de 3 años (36 meses), ¿cuál es la probabilidad de riesgo alto?

Pregunta 2: (50%): Eliminación de Variables

1. Estudie el método de eliminación de variables (Variable elimination) para inferencia exacta en redes bayesianas. Explíquelo con sus palabras.
2. Considere la siguiente red que modela la probabilidad de tener un ataque cardíaco antes de los 70 (alta o baja) dado diferentes factores: si fuma o no fuma, si ejerce o no ejerce, el nivel de colesterol (alto o bajo) y la presión sanguínea (alta o baja). Calcule las siguientes probabilidades utilizando el algoritmo de eliminación de variables, explique cada paso del algoritmo (debe resolverlo con lápiz y papel, sin utilizar código).
 - (a) Dado que el paciente es fumador ¿Cuál es la probabilidad de que tenga un infarto?
 - (b) Dado que el paciente tuvo un infarto, ¿Cuál es la probabilidad de que tuviese la presión arterial alta?
 - (c) Dado que el paciente es fumador y tiene el colesterol alto, ¿Cuál es la probabilidad de que tenga un infarto?

