

# Práctica Obligatoria de Redes Bayesianas

**Esta práctica es OBLIGATORIA. La fecha límite de entrega es el domingo 9 de diciembre de 2018 a las 23:55**

Esta práctica se compone de dos bloques distintos. El primer ejercicio se valorará en una escala de 0 a 10, siendo obligatorio obtener una calificación de al menos 5 puntos para tener la práctica aprobada. Por otra parte, el segundo se valorará en una escala de 0 a 5. **La nota final de la práctica obligatoria será la suma de ambos.** Esto quiere decir que el segundo ejercicio es opcional, ya que se puede obtener un 10 con el primero.

En caso de obtener más de 10, la nota extra computará en la parte de Práctica Voluntarias.

## Ejercicio obligatorio:

Un granjero llama al veterinario para que determine si una de sus vacas está preñada o no (P). Para determinar esto, lo primero que hace el veterinario es preguntar si hay toros (HT) en la finca o no (el 60% de los granjeros los tienen), y en caso de haberlos, si estos se mezclan (MT) con las vacas (suele ocurrir en el 30% de los casos). Si los hay, y además se mezclan con las vacas, la probabilidad de que esté preñada es del 30%, mientras que si no se mezclan es del 10%. En caso de no haber toros, entonces sólo hay un 1% de probabilidades de que la vaca esté preñada (hay otros toros en otras fincas).

A continuación, para afinar con el diagnóstico, el veterinario lleva a cabo tres tests: una ecografía (E), un análisis de sangre (AS) y un análisis de orina (AO). En el caso de la ecografía, ésta mostrará una anomalía con una probabilidad del 90% en caso de que la vaca esté preñada y con un 15% en caso de que no lo esté.

Por otra parte, el test de sangre puede ser engañoso, ya que hay vacas con un tipo de sangre específico (SR) que hace que el resultado del test de sangre sea siempre anómalo. Si bien es cierto que sólo un 5% de las vacas tienen este tipo de sangre. No obstante, existe una prueba (PS) para comprobar si una vaca tiene la sangre de este tipo “raro” o no, aunque no es infalible. Si la vaca tiene sangre “rara”, el test lo detecta correctamente un 95% de las veces, y si su sangre es normal, el test falla un 10% de las veces.

En cualquier caso, tanto el test de sangre como el de orina dependen en gran medida del nivel hormonal (NH) de la vaca (alto o bajo), el cual depende a su vez de si la vaca está realmente preñada. Si está preñada, entonces el nivel hormonal es alto un 60% de

las veces (en circunstancias normales está alto en 2 de cada 10 vacas), y en ese caso la probabilidad de anomalías en la muestra de orina es del 75%. En el caso del test de sangre, éste aparecerá con anomalías en un 40% de los casos en los que la vaca tiene el nivel hormonal alto (siempre que no tenga la sangre “rara”). Cuando el nivel hormonal es bajo, la probabilidad de encontrar una anomalía en una muestra de orina es del 15%, y de un 10% en el caso de la sangre (de nuevo, si no es “rara”).

A partir de la situación descrita, se deberán resolver los siguientes apartados:

- Implementa en Netica, la red bayesiana derivada de la información anterior.
- Introduce todas las probabilidades necesarias.
- Utilizando la red que has creado, calcula los siguientes datos:
  - Si el granjero tiene toros, la vaca no tiene sangre “rara” y su muestra de sangre es anómala, ¿cuál es la probabilidad de que la vaca esté preñada?
  - ¿Y si no sabemos nada acerca de la sangre rara, pero el test dice que sí la tiene?
- Apoyándote en la ventana de mensajes de Netica, crea la distribución conjunta para el subconjunto de variables (MT, E, AS)
- Haz una **inferencia exacta** sobre la red:
  - Elige un valor para cada una de las variables y calcula a mano la probabilidad (transparencia 42 de teoría):
 
$$P(HT, \neg MT, P, \neg E, \neg NH, \neg AO, AS, \neg SR, \neg PS)$$
  - Explica cómo sería el procedimiento si lo que quisieses calcular fuese la probabilidad condicionada (transparencia 43 de teoría):
 
$$P(AS|AO, \neg P)$$
- Deberás hacer las siguientes comprobaciones de independencia de variables:
  - Siguiendo la condición de Markov (transparencia 32 de teoría), ¿qué valores debería conocer si quiero afirmar que E y PS son independientes? ¿Y para afirmar que MT y AS lo son?
  - Apoyándote en el concepto de D-separación, ¿crees que P y AS son independientes? Explica por qué
  - Apoyándote en el concepto de D-separación, ¿crees que P y PS son independientes? Explica por qué. ¿Y si conociésemos AS?
- Dar un traza de ejemplo de cómo generar una posible muestra estocástica para calcular  $P(MT|E, \neg AS)$  (puedes hacer uso de la página <https://www.random.org/integers/> para crear los números aleatorios). Deberás indicar, en cada paso, el número aleatorio que ha salido y por qué has asignado cada valor a cada variable.
- Repetir el apartado anterior, pero utilizando el método de **ponderación de la verosimilitud**. Se deberá utilizar la misma secuencia de números aleatorios que en el caso anterior.
- Por último, indica qué ventajas tiene el método de ponderación de la verosimilitud sobre el muestreo estocástico.

**Se deberá entregar un fichero comprimido en la tarea del Campus Virtual habilitada para tal efecto. Dicho fichero comprimido contendrá la red creada (en el**

formato de Netica), y un informe en formato pdf dando respuesta a todos los ejercicios planteados.

## Extensión Opcional:

Se trata de construir una red bayesiana en Netica. El tema de la red es libre, puedes hacerla sobre cualquier tema que te parezca interesante.

Se valorará la originalidad del tema, el interés o utilidad de la red y también en cierta medida su complejidad (esto no significa que la hagas demasiado complicada sin necesidad, sólo que no sea excesivamente simple).

**Es obligatorio que la red diseñada tenga un número de nodos de como mínimo 8 y como máximo 12.** Si para el problema que quieres resolver necesitas más nodos, o menos, consulta al profesor. De todas formas, recuerda que Netica (en su versión gratuita) no permite guardar redes de más de 15 nodos, por lo que la red que desees crear no puede ser mucho mayor.

**Además de implementar la red, se escribirá un breve informe sobre la red creada.** Dicho informe deberá contener los siguientes elementos:

- Una **explicación** de en qué consiste la red implementada, para qué sirve, y quizás algún detalle o curiosidad adicional que hayas comprobado utilizándola.
- Cómo se han estimado las probabilidades introducidas en las tablas.
- Dar algún ejemplo de cálculo sobre la red que sea útil para el problema a resolver.
- Deberás hacer tres comprobaciones de independencia de variables:
  - En una de ellas debes comprobar que son independientes utilizando la condición de Markov (ver diapositiva 32 de teoría)
  - Para las otras dos debes usar el método de D-separación, pero haz que en un caso sean independientes y en el otro no. Elige las variables que desees, y también las variables conocidas que desees.

**Se deberá entregar un fichero comprimido en la tarea del Campus Virtual habilitada para tal efecto.** Dicho fichero comprimido contendrá la red creada (en el formato de Netica), y el informe correspondiente en formato pdf. Además, si la red que desees crear quieres que aprenda datos de algún fichero, deberás incluir también dicho fichero en el fichero comprimido. La parte del informe relativa a esta tarea, se incluirá en una sección nueva detrás de la resolución de los ejercicios obligatorios.