



# GUÍA DE ESTUDIO DE LA ASIGNATURA MÉTODOS PROBABILISTAS

2ª Parte

Manuel Luque Gallego,  
Manuel Arias Calleja  
y Francisco Javier Díez Vegas

MÁSTER UNIVERSITARIO EN  
INTELIGENCIA ARTIFICIAL AVANZADA:  
FUNDAMENTOS, MÉTODOS Y APLICACIONES

Curso 2017-2018



# Índice general

<b>1. Plan de trabajo</b>	<b>3</b>
1.1. Secuenciación de contenidos . . . . .	3
1.2. Actividades propuestas . . . . .	3
1.3. Planificación temporal . . . . .	4
<b>2. Orientaciones para el estudio</b>	<b>7</b>
2.1. Fundamentos de redes bayesianas . . . . .	8
2.1.1. Objetivos . . . . .	8
2.1.2. Orientaciones específicas . . . . .	8
2.2. Inferencia en redes bayesianas . . . . .	9
2.2.1. Objetivos . . . . .	9
2.2.2. Orientaciones . . . . .	9
2.3. Construcción de redes bayesianas . . . . .	9
2.3.1. Objetivos . . . . .	9
2.3.2. Orientaciones . . . . .	10
2.4. Análisis de decisiones . . . . .	10
2.4.1. Objetivos . . . . .	10
2.4.2. Orientaciones . . . . .	10
2.5. Aplicaciones . . . . .	11
2.5.1. Orientaciones . . . . .	11
<b>3. Presentación de las prácticas</b>	<b>13</b>
3.1. Recomendaciones sobre la presentación de las prácticas . . . .	13
3.2. Máquina virtual con herramientas básicas . . . . .	14



# Introducción

Ésta es la parte 2 de la guía de estudio de la asignatura “*Métodos Probabilistas*” (código 31101199). Esta guía amplía la información de la parte 1 de la guía de estudio que se encuentra disponible públicamente en la web de la UNED; recomendamos al alumno que antes de comenzar a leer este documento lea la parte 1 de la guía: Parte 1.

Este documento trata de ayudar al alumno a conseguir una organización eficaz del estudio de esta asignatura. Le recomendamos que lo lea con paciencia y que consulte regularmente el curso virtual en Alf, ya que constituye el espacio común de trabajo de alumnos y profesores.

Vamos a comenzar dando una visión general del plan de trabajo. Seguidamente, proporcionaremos las orientaciones para el estudio. Finalmente, describiremos unos consejos sobre la entrega de las prácticas.

Es importante señalar que en esta asignatura existen dos opciones para cursarla: *general* (modalidad A) y *orientada a la medicina* (modalidad B). Para la mayor parte de los alumnos recomendamos la A, que estudia con mayor profundidad los aspectos matemáticos y algorítmicos de los modelos gráficos probabilistas, mientras que para aquéllos más interesados en la aplicación de la inteligencia artificial a la medicina recomendamos la B, que trata algunos temas específicos del análisis de decisiones sanitarias, tales como el análisis de coste-efectividad.

Este documento corresponde fundamentalmente a la modalidad A. Los documentos de la modalidad B, incluyendo una guía de estudio específica, se encuentran en una carpeta propia dentro del área de documentos. Los ejercicios de evaluación también son (parcialmente) diferentes. No es necesario que cada alumno/a indique explícitamente qué opción desea cursar: basta que entregue los ejercicios de la modalidad que prefiera.



# Capítulo 1

## Plan de trabajo

Vamos a describir el plan de trabajo básico que el equipo docente seguirá a lo largo del curso.

### 1.1. Secuenciación de contenidos

Los contenidos de esta asignatura se han agrupado en cinco temas, impartidos en este orden:

1. Fundamentos de redes bayesianas
2. Inferencia en redes bayesianas
3. Construcción de redes bayesianas
4. Análisis de decisiones
5. Aplicaciones

### 1.2. Actividades propuestas

La sección “*Orientaciones para el estudio*” describe los resultados de aprendizaje esperados. Se proponen los siguientes tipos de actividades:

- la lectura y estudio de contenidos teóricos y artículos de bibliografía
- la discusión activa, en los foros, de aquellas dudas que el alumno tenga acerca del material anterior
- la realización de tareas consistentes en entregar en formato electrónico distintas prácticas con ejercicios

- el uso de un software especializado: OpenMarkov

Para una mejor organización del curso, el profesor Manuel Arias Calleja se encargará de responder dudas y corregir las actividades de los temas 1 y 3, el profesor Manuel Luque Gallego hará lo mismo pero con los temas 2 y 4, y el tema 5 será cubierto por los tres profesores del equipo docente.

### **1.3. Planificación temporal**

Ésta es una asignatura con 6 créditos ECTS reconocidos, es decir, con una dedicación de unas 150 horas por parte del estudiante. Recomendamos seguir el ritmo sugerido en el curso virtual para entregar las 5 prácticas de evaluación continua con las que se calcula la nota final de la asignatura.

Como referencia, adjuntamos la propuesta de hitos y de dedicación de la tabla 1.1.



Tabla 1.1: Plan de trabajo de la asignatura. La dedicación se indica en número de horas de trabajo del alumno.

Hito	Tarea	Dedicación
<b>Fundamentos de redes Bayesianas</b>		
1	Estudiar los contenidos teóricos de este tema	15
2	Realizar ejercicios no evaluables	5
3	Realizar la 1ª Práctica de Evaluación Continua	10
<b>Inferencia en redes Bayesianas</b>		
4	Estudiar los contenidos teóricos de este tema	15
5	Realizar ejercicios no evaluables	5
6	Realizar la 2ª Práctica de Evaluación Continua	10
<b>Construcción de redes Bayesianas</b>		
7	Estudiar los contenidos teóricos de este tema	11.25
8	Realizar ejercicios no evaluables	3.75
9	Realizar la 3ª Práctica de Evaluación Continua	7.5
<b>Análisis de decisiones</b>		
10	Estudiar los contenidos teóricos de este tema	15
11	Realizar ejercicios no evaluables	5
12	Realizar la 4ª Práctica de Evaluación Continua	10
<b>Aplicaciones</b>		
13	Estudiar artículos sobre aplicaciones	18.75
14	Realizar la 5ª Práctica de Evaluación Continua	18.75
<b>TOTAL</b>		150



## Capítulo 2

# Orientaciones para el estudio

Tal y cómo hemos mencionado anteriormente, la asignatura se compone de 5 temas. Las orientaciones generales, comunes a todos los temas son:

- El texto básico de estudio es (Díez, 2007), que está disponible para libre descarga en el siguiente enlace: <http://www.ia.uned.es/~fjdíez/libros/intro-mgp.html>.
- Habrá ejercicios que el alumno tenga que realizar a mano y con calculadora.
- El software básico a utilizar es OpenMarkov, que se puede descargar desde: <http://www.openmarkov.org/usuarios.html>. Habrá ejercicios en que se pedirá expresamente al alumno que utilice OpenMarkov. Además, siempre que el alumno tenga que realizar un ejercicio manualmente conviene que utilice OpenMarkov (si es posible) para verificar los resultados. Es por ello que le recomendamos se instale cuanto antes el software y se familiarice con su interfaz gráfica. Un tutorial de OpenMarkov está disponible en <http://www.openmarkov.org/docs/tutorial/tutorial.html>; estamos seguros de que le ayudará a sacarle más partido a OpenMarkov.
- Cada tema tiene una práctica de evaluación continua asociada, cuyo enunciado se encuentra para descarga en el apartado “Tareas” del curso virtual en Alf. El alumno tendrá que cumplir los plazos de entrega que se establecen ahí. La entrega se realizará también en ese mismo apartado “Tareas” adjuntando el archivo preparado por el alumno.

Pasamos ahora a ver los objetivos y orientaciones específicas de cada tema.

## **2.1. Fundamentos de redes bayesianas**

### **2.1.1. Objetivos**

El objetivo de este tema es exponer los fundamentos axiomáticos de las redes bayesianas. Para ello se hace un repaso de los conceptos fundamentales de la teoría de la probabilidad y la teoría de grafos y se introduce el método bayesiano ingenuo. El alumno debe comprender la definición matemática de red matemática y las implicaciones que ésta tiene en cuanto representación de independencias probabilistas y en cuanto representación causal del mundo real.

Los objetivos concretos son:

1. Recordar las definiciones y teoremas básicos de la teoría de la probabilidad.
2. Comprender el método bayesiano ingenuo, conocer sus ventajas y sus limitaciones y saber en qué casos puede ser aplicado y en cuáles no.
3. Recordar los conceptos básicos de la teoría de grafos, especialmente los grafos dirigidos acíclicos, que constituyen la base de los modelos gráficos probabilistas.
4. Aprender la definición axiomática de red bayesiana, que se basa en la factorización de la probabilidad.
5. Comprender el significado de los grafos de dependencias e independencias y sus propiedades básicas, que son la separación direccional y la propiedad de Markov, y cómo éstas se relacionan con la factorización de la probabilidad en redes bayesianas.
6. Distinguir entre la interpretación probabilista de un grafo y su interpretación causal, y la relación que existe entre ambas.

### **2.1.2. Orientaciones específicas**

El alumno tiene que leer el tema 1 de (Díez, 2007). A medida que el alumno estudia los diferentes algoritmos, conviene que piense cómo los implementaría en su lenguaje de programación preferido (por ejemplo, Java) y trate de analizar cuál es la complejidad espacial y temporal de cada uno de ellos para distintos tipos de redes.

## **2.2. Inferencia en redes bayesianas**

### **2.2.1. Objetivos**

El objetivo de este tema es conocer los principales algoritmos de inferencia en redes bayesianas, que permiten calcular la probabilidad a posteriori de cada variable o de un conjunto de variables dada cierta evidencia.

Los objetivos concretos son:

1. Conocer los métodos exactos de inferencia en redes bayesianas más utilizados, principalmente la eliminación de variables, los métodos de agrupamiento y la inversión de arcos.
2. Conocer algunos de los métodos estocásticos de inferencia en redes bayesianas más relevantes, tales como el muestreo lógico, la ponderación por verosimilitud y el muestreo por importancia.

### **2.2.2. Orientaciones**

El alumno tiene que leer el tema 2 de (Díez, 2007). Conviene que realice los ejercicios no evaluables propuestos a final de dicho tema. Finalmente, tendrá que realizar y entregar en el curso virtual la 2ª Prueba de Evaluación Continua.

## **2.3. Construcción de redes bayesianas**

### **2.3.1. Objetivos**

El objetivo de este tema es conocer los dos métodos principales de construcción de redes bayesianas. El primero de ellos, que se aplica cuando no existe una base de datos de suficiente tamaño o calidad, consiste en construir primero el grafo de la red con la ayuda de un experto humano y luego añadir las probabilidades numéricas. El segundo consiste en aplicar un algoritmo de aprendizaje a una base de datos, con lo cual se obtiene de forma automática la red completa, es decir, el grafo y las probabilidades. Los objetivos concretos son:

1. Ser capaz de (1) construir un grafo causal a partir de una descripción de un problema, (2) determinar en qué nodos pueden aplicarse modelos canónicos que simplifican la obtención de las probabilidades y (3) interrogar a un experto humano con el fin de obtener estimaciones de las probabilidades numéricas.

2. Conocer los principales algoritmos de aprendizaje automático de redes bayesianas y ser capaz de aplicarlos a una base de datos mediante el programa OpenMarkov.

### **2.3.2. Orientaciones**

El alumno tiene que leer el tema 3 de (Díez, 2007). Conviene que realice los ejercicios no evaluables propuestos a final de dicho tema. Finalmente, tendrá que realizar y entregar en el curso virtual la 3ª Prueba de Evaluación Continua.

## **2.4. Análisis de decisiones**

### **2.4.1. Objetivos**

El objetivo de este tema es aprender a resolver problemas de decisiones mediante modelos gráficos probabilistas. De forma más específica, se trata de que el alumno/a aprenda a construir diagramas de influencia y a evaluarlos mediante la expansión de un árbol de decisión equivalente, mediante el método de eliminación de variables o mediante inversión de arcos. Los objetivos concretos son:

1. Comprender los fundamentos de teoría de la decisión, especialmente los conceptos de valor esperado y utilidad esperada.
2. Comprender qué es un diagrama de influencia (DI) y un árbol de decisión (AD), cómo construir un AD para un DI dado y cómo evaluar el AD.
3. Conocer otros métodos de evaluación de DI's más eficientes que la conversión en AD, tales como la eliminación de variables y la inversión de arcos.
4. Ser capaz de construir DI's para aplicaciones reales: medicina, seguridad informática, etc.

### **2.4.2. Orientaciones**

El alumno tiene que leer el tema 4 de (Díez, 2007). Conviene que realice los ejercicios no evaluables propuestos a final de dicho tema. Finalmente, tendrá que realizar y entregar en el curso virtual la 4ª Prueba de Evaluación Continua.

Conviene construir con el programa OpenMarkov los DI's propuestos en los ejemplos de este tema y expandir el árbol de decisión asociado a cada uno de ellos, para comprobar la equivalencia de ambas representaciones.

## 2.5. Aplicaciones

El objetivo de este tema es conocer cómo los modelos gráficos probabilistas (MGPs) se están utilizando para resolver algunos problemas del mundo real. Los objetivos concretos son:

1. Comprender cómo los MGPs resuelven problemas de diagnóstico y planificación terapéutica en medicina.
2. Comprender cómo los MGPs se utilizan en informática educativa y en la construcción de interfaces inteligentes, especialmente como modelos del usuario.
3. Comprender cómo los MGPs se aplican en seguridad informática y vigilancia: filtrado de correo basura, detección de intrusos, etc.
4. Comprender cómo los MGPs se aplican en ingeniería (diagnóstico y reparación de averías) y en visión artificial.
5. Comprender cómo los MGPs se aplican en otros campos, tales como la recuperación de información, el comercio electrónico, el análisis de riesgos financieros, etc.

### 2.5.1. Orientaciones

El alumno tiene que realizar y entregar en el curso virtual la 5ª Prueba de Evaluación Continua. Como orientación de dicho trabajo, conviene que el alumno/a lea varios de los artículos recomendados en la bibliografía de este tema, especialmente aquellos más relacionados con el área en que esté más interesado/a, con el fin de que en el futuro pueda construir MGPs aplicados a los problemas que encuentre en su práctica profesional.

A lo largo del desarrollo del curso, el equipo docente y los alumnos podrán proponer otras actividades orientadas a ampliar algunos de los aspectos teóricos o algorítmicos expuestos en el temario, o a implementar mediante ordenador los métodos aprendidos.





## Capítulo 3

# Presentación de las prácticas

### 3.1. Recomendaciones sobre la presentación de las prácticas

Queremos comenzar dejando claro que el equipo docente da mucha importancia a la calidad de la presentación de los trabajos de cara a la calificación de las prácticas. Esta importancia se debe a que en cursos pasados hemos encontrado que algunos compañeros vuestros han presentado sus actividades redactándolas a mano y después escaneándolas, o a veces en formato digital pero de no muy fácil lectura. Es por ello que queremos dar las siguientes directrices para la calidad de la presentación de los trabajos sea adecuada.

Es importante que los trabajos tengan una estructura clara y que la información esté bien presentada y organizada. Os dejo como ejemplo de una posible forma de presentar el trabajo el siguiente informe técnico: <http://www.cisiad.uned.es/techreports/luque08b.pdf>. En él se pueden distinguir elementos como el nombre y apellidos, los títulos del documento y de secciones y subsecciones, y cómo al final del mismo aparece una lista de las referencias bibliográficas que son citadas en el texto; el resumen (abstract) y el índice (Content) son opcionales.

No vamos a aceptar trabajos escritos a mano. El formato en que se entreguen las prácticas debe ser PDF y en él el alumno se tiene que asegurar de que todas las tablas, figuras y texto relevantes para la práctica se pueden leer adecuadamente.

Aconsejamos el uso del sistema de tipografía  $\text{\LaTeX}$  para escribir los trabajos. El principal motivo por el que lo sugerimos es porque nuestra asignatura presenta un alto contenido matemático y  $\text{\LaTeX}$  tiene muy buenas cualidades para la edición de texto matemático. Además,  $\text{\LaTeX}$  es altamente recomendable cuando se presenta una tesis de máster o una tesis doctoral en Ciencias

de la Computación debido a todo el trabajo automático que es capaz de realizar, dejando que el autor sólo tenga que preocuparse de escribir el contenido y no de su apariencia. En caso de que utilizar  $\text{\LaTeX}$ , recomendamos emplear el sistema Bibtex para la gestión de las referencias bibliográficas; también dispone de un editor de interfaz amigable para Bibtex como es Jabref.

Si el alumno decide utilizar  $\text{\LaTeX}$  entonces le recomendamos emplear el programa LyX, que es un editor del tipo WYSIWYM (*What you see is what you mean*) que le evita dedicar tiempo en aprender la sintaxis de  $\text{\LaTeX}$ . LyX es multiplataforma, por lo que funciona tanto en Windows, Mac como Linux.

Aunque recomendamos utilizar  $\text{\LaTeX}$ , seguiremos aceptando trabajos editados con programas como Microsoft Word, OpenOffice, LibreOffice, etc. No obstante, la experiencia nos ha enseñado que los trabajos que han presentado los alumnos que han utilizado  $\text{\LaTeX}$  han tenido generalmente mejor calidad en apariencia.

Finalmente, dejamos como opcional que el estudiante presente las prácticas redactadas en inglés en lugar de en español y lo valoraremos positivamente. A día de hoy el idioma inglés sigue teniendo una enorme importancia en la comunidad científica.

Esperamos que estas directrices sean fáciles de cumplir. Además, estamos seguros que con el paso del tiempo redundarán claramente en vuestro beneficio.

## 3.2. Máquina virtual con herramientas básicas

Para ahorrar al alumno el tiempo de instalación del software OpenMarkov y otros programas de edición para la realización de las prácticas, el equipo docente ha preparado una máquina virtual sobre VirtualBox que se puede descargar y trabajar con ella. El sistema operativo instalado es Ubuntu 13.10, que es una versión de Ubuntu muy ligera en uso de memoria y apropiada para su uso en máquinas virtuales.

Los programas que hemos instalado en la máquina virtual, además de OpenMarkov, son:

- LyX: Editor WYSIWYM que emplea  $\text{\LaTeX}$  para generar documentos en PDF y otros formatos.
- Jabref: Gestor de bibliografía para Bibtex.
- $\text{\LaTeX}$ ila: Editor de texto plano para  $\text{\LaTeX}$ .

- Editores de imágenes:
  - Qtikz: Editor de gráficos con el paquete Tikz para uso con  $\text{\LaTeX}$ .
  - Inkscape: Editor de gráficos vectoriales.
  - Dia: Editor de diagramas.
- Dropbox: Programa de compartición de archivos en la nube para poder pasar fácilmente archivos entre vuestro ordenador y la máquina virtual.

Para instalar la máquina virtual en vuestro ordenador tenéis que seguir los siguientes pasos:

1. Descargar un cliente de VirtualBox de la página: <https://www.virtualbox.org/wiki/Downloads> (Recomendamos instalarlo en inglés).
2. Descargar el archivo de la máquina virtual, que es: <https://drive.google.com/file/d/0BxTvhfWHr0AfrXlpa0lacUdkdDQ/edit?usp=sharing>
3. Abrir VirtualBox, pulsar en "File" -> "Import Appliance" -> "Open appliance", seleccionar el archivo descargado, pulsar en "Next" y en "Import".

La máquina virtual ya estaría instalada en el equipo. Para arrancarla sólo hay que seleccionarla en VirtualBox y pulsar en "Start". El usuario en el sistema es "student" (sin comillas) y la contraseña es "student". Al entrar en el sistema se encontrará que en el escritorio hay enlaces a los programas indicados arriba. Además, como cualquier sistema operativo, tiene programas básicos como el navegador (Firefox), explorador de archivos, visor de PDFs, etc.

Esperamos que la máquina virtual resulte muy útil.



# Bibliografía

Díez, F. J., 2007. Introducción a los modelos gráficos probabilistas. UNED, Madrid.