

62234 - Machine learning for Big Data

Información del Plan Docente

Año académico	2016/17
Centro académico	110 - Escuela de Ingeniería y Arquitectura
Titulación	534 - Máster Universitario en Ingeniería Informática
Créditos	3.0
Curso	2
Periodo de impartición	Primer Semestre
Clase de asignatura	Optativa
Módulo	---

1. Información Básica

1.1. Recomendaciones para cursar esta asignatura

No existe ningún requisito ni recomendación especial para cursar la asignatura.

1.2. Actividades y fechas clave de la asignatura

El calendario de clases, prácticas y exámenes, así como las fechas de entrega de trabajos de evaluación, se anunciará con suficiente antelación.

2. Inicio

2.1. Resultados de aprendizaje que definen la asignatura

El estudiante, para superar esta asignatura, deberá demostrar los siguientes resultados...

- Conocer técnicas avanzadas de aprendizaje automático y reconocimiento de patrones.
- Comprender los fundamentos de varios algoritmos de aprendizaje automático para grandes volúmenes de datos.
- Ser capaz de analizar los resultados de un algoritmo de aprendizaje automático, relacionarlos con los parámetros e hiperparámetros del algoritmo, comprender si es posible mejorar dichos resultados y rediseñar el algoritmo para alcanzar dicha mejora.
- Comprender los ámbitos de aplicación y limitaciones de los algoritmos de aprendizaje automático; y ser capaz de caracterizar cada problema y escoger el algoritmo más adecuado.
- Comprender de manera detallada y precisa los aspectos teóricos y prácticos de las redes neuronales profundas.
- Implementar aplicaciones de aprendizaje automático utilizando grandes volúmenes de datos.
- Conocer algunos de los resultados más recientes, los problemas abiertos y las oportunidades que ofrece el aprendizaje automático con grandes volúmenes de datos.

2.2. Introducción

Breve presentación de la asignatura

El análisis y la extracción de patrones en grandes volúmenes de datos ocupan actualmente un lugar fundamental en diversas aplicaciones científicas y tecnológicas y tienen gran importancia económica y social. La inmensa cantidad de datos disponibles en la actualidad, provenientes de tecnologías diversas como sensores o la Web 2.0, ha producido un salto cualitativo en numerosos problemas como el reconocimiento de escenas, caras u objetos en imágenes, sistemas de recomendación, o lenguaje natural. En estos y otros campos, la disponibilidad de grandes conjuntos de datos ofrece varias oportunidades, como modelar patrones complejos o mejorar el desempeño de algoritmos sencillos.

62234 - Machine learning for Big Data

El objetivo de este curso es introducir al estudiante en el aprendizaje automático con grandes volúmenes de datos, exponer el estado del arte, limitaciones y ejemplos de aplicación, y profundizar en los algoritmos más extendidos. En concreto se profundizará sobre todo en redes neuronales, que es el algoritmo con más potencial para capturar patrones complejos a partir de datos. También se estudiará el aprendizaje secuencial para el caso de flujos de datos y aproximaciones eficientes al problema del vecino más próximo.

3.Contexto y competencias

3.1.Objetivos

La asignatura y sus resultados previstos responden a los siguientes planteamientos y objetivos:

El objetivo de la asignatura es dotar al estudiante de la capacidad de diseñar sistemas de aprendizaje automático para extraer patrones de grandes volúmenes de datos. En la asignatura se ofrece una visión muy especializada del estado del arte, para que el estudiante sea capaz de comprender las técnicas utilizadas en las aplicaciones actuales y pueda entender los desarrollos futuros. El curso se focaliza en su mayor parte en los aspectos teóricos y prácticos de las redes neuronales y tratará con menor detalle otros algoritmos. La asignatura se plantea con una elevada carga práctica y de trabajo individual del estudiante, para que se aprendan los conceptos fundamentales con una mayor profundidad.

3.2.Contexto y sentido de la asignatura en la titulación

Esta asignatura está relacionada con otras asignaturas del máster que tratan temas computacionales, en concreto Sistemas Inteligentes, Manipulación y análisis de grandes volúmenes de datos y Computación gráfica. A diferencia de las anteriores, ofrece una formación especializada para el caso concreto de extracción de información con grandes volúmenes de datos. La asignatura responde a la necesidad creciente del mercado laboral de incorporar expertos especialistas en datos (data scientists) en un entorno donde inmensas cantidades de datos poco estructurados son fácilmente accesibles pero las técnicas para extraer información útil son complejas y están en constante evolución.

3.3.Competencias

Al superar la asignatura, el estudiante será más competente para...

Afrontar con éxito los siguientes desempeños transversales:

1. Utilizar modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática.
2. Aplicar de los conocimientos adquiridos y resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinarios, siendo capaces de integrar estos conocimientos.
3. Adquirir conocimientos avanzados y demostrados, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en uno o más campos de estudio.
4. Aplicar e integrar sus conocimientos, la comprensión de estos, su fundamentación científica y sus capacidades de resolución de problemas en entornos nuevos y definidos de forma imprecisa, incluyendo contextos de carácter multidisciplinar tanto investigadores como profesionales altamente especializados.
5. Evaluar y seleccionar la teoría científica adecuada y la metodología precisa de sus campos de estudio para formular juicios a partir de información incompleta o limitada incluyendo, cuando sea preciso y pertinente, una reflexión sobre la responsabilidad social o ética ligada a la solución que se proponga en cada caso.
6. Transmitir de un modo claro y sin ambigüedades a un público especializado o no, resultados procedentes de la investigación científica y tecnológica o del ámbito de la innovación más avanzada, así como los fundamentos más relevantes sobre los que se sustentan.
7. Desarrollar la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro su ámbito temático, en contextos interdisciplinarios y, en su caso, con una alta componente de transferencia del conocimiento.
8. Asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de

62234 - Machine learning for Big Data

estudio.

9. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
10. Aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
11. Integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
12. Comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
13. Poseer las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Afrontar con éxito los siguientes desempeños relacionados con la Ingeniería Informática:

1. Aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

3.4.Importancia de los resultados de aprendizaje

El aprendizaje a partir de grandes volúmenes de datos es una pieza fundamental en multitud de sistemas tecnológicos y científicos: buscadores, sistemas de recomendación, sistemas financieros y de toma de decisiones, redes sociales, reconocimiento de lenguaje natural oral y escrito o reconocimiento visual de objetos, escenas o acciones. La importancia de este tipo de sistemas es creciente, no sólo desde el punto de vista económico sino también social.

4.Evaluación

El estudiante deberá demostrar que ha alcanzado los resultados de aprendizaje previstos mediante las siguientes actividades de evaluación

Realización y presentación de un caso práctico . Selección de un caso de aplicación de las técnicas vistas en clase, estudio de la bibliografía existente, selección del algoritmo más adecuado, implementación de la solución, análisis de resultados, rediseño si fuese necesario, presentación del trabajo y conclusiones. [90%]. Resultados de aprendizaje: 1, 2, 3, 4, 5 y 6

Realización de las sesiones prácticas de la asignatura . Será obligatoria la asistencia, realización, y redacción de un breve informe de cada una de las 3 sesiones prácticas de la asignatura. [10%]. Resultados de aprendizaje: 1, 2, 3, 4, 5 y 6

El estudiante que no opte por el procedimiento de evaluación descrito anteriormente, no supere dichas pruebas durante el periodo docente o que quisiera mejorar su calificación tendrá derecho a realizar una prueba global que será programada dentro del periodo de exámenes correspondiente a la primera o segunda convocatoria.

5.Actividades y recursos

5.1.Presentación metodológica general

El proceso de aprendizaje que se ha diseñado para esta asignatura se basa en lo siguiente:

Las actividades de enseñanza y aprendizaje presenciales se basan en:

1. **Clase presencial**. Exposición de contenidos mediante presentación o explicación por parte del profesor.
2. **Charlas de expertos**. Exposición de contenidos mediante presentación o explicación por parte de un experto externo a la Universidad.

62234 - Machine learning for Big Data

3. **Laboratorio** . Actividades desarrolladas en espacios especiales con equipamiento especializado (laboratorio, aulas informáticas).
4. **Tutoría** . Período de instrucción realizado por un tutor con el objetivo de revisar y discutir los materiales y temas presentados en las clases. Varias de ellas (al menos 4 horas) serán obligatorias para informar del progreso del trabajo de asignatura.
5. **Evaluación** . Presentación oral del trabajo de asignatura.

Las actividades de enseñanza y aprendizaje no presenciales se basan en:

1. **Trabajos prácticos**. Preparación de actividades para exponer o entregar en las clases prácticas.
2. **Estudio teórico**. Estudio de contenidos relacionados con las "clases teóricas".
3. **Estudio práctico**. Estudio relacionado con las "clases prácticas"
4. **Realización del trabajo de asignatura** . Realización individual tutorada de un trabajo relacionado con los conceptos de la asignatura, bajo la supervisión del profesor.

5.2.Actividades de aprendizaje

El programa que se ofrece al estudiante para ayudarle a lograr los resultados previstos comprende las siguientes actividades...

Trabajo del estudiante

La asignatura consta de 3 créditos ECTS que suponen una dedicación estimada por parte del alumno de 75 horas (35 horas presenciales y 40 horas no presenciales) distribuidas del siguiente modo:

- 30 horas de actividades presenciales (clases magistrales incluyendo seminarios profesionales, resolución de problemas y casos, y prácticas de laboratorio).
- 35 horas de trabajo personal para la realización del trabajo práctico de asignatura.
- 5 horas para la redacción de informes de las sesiones prácticas y preparación de la presentación del trabajo práctico.

5.3.Programa

1. Introducción y resumen de conocimientos previous (fundamentos de aprendizaje automático y redes neuronales).
2. Redes neuronales profundas (Deep Learning).
3. Entrenamiento de redes neuronales profundas. Retropropagación.
4. Redes convolucionales.
5. Aspectos prácticos del aprendizaje con redes profundas (entre otros: gradiente numérico, sobreajuste, regularización, funciones de activación y gradiente descendiente estocástico).
6. Aplicaciones: Procesamiento de lenguaje y clasificación visual.
7. Herramientas software y hardware para redes neuronales profundas.
8. Aprendizaje secuencial a partir de flujos de datos.
9. Vecino más próximo aproximado.

5.4.Planificación y calendario

Calendario de sesiones presenciales y presentación de trabajos

La organización docente prevista de las sesiones presenciales en el campus Río Ebro es la siguiente:

- Clases magistrales.
- Resolución de problemas y casos.
- Prácticas de laboratorio.

Los horarios de todas las clases y fechas de las sesiones de prácticas se anunciarán con suficiente antelación a través de las webs del centro y de la asignatura.

Los proyectos propuestos serán entregados al finalizar el cuatrimestre, en las fechas que se señalen.

5.5. Bibliografía y recursos recomendados