

Máster en Ingeniería Informática

62234 - Machine learning for Big Data

Guía docente para el curso 2015 - 2016

Curso: , Semestre: , Créditos: 3.0

Información básica

Profesores

- **Javier Civera Sancho** jcivera@unizar.es

Recomendaciones para cursar esta asignatura

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

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.

Inicio

Resultados de aprendizaje que definen la asignatura

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

1:

Conocer técnicas avanzadas de aprendizaje automático y reconocimiento de patrones.

2:

Comprender los fundamentos de varios algoritmos de aprendizaje automático para grandes volúmenes de datos.

3:

Ser capaz de analizar los resultados de un algoritmo de aprendizaje automático y relacionarlos con los parámetros del algoritmo.

4:

Comprender los ámbitos de aplicación y limitaciones de los algoritmos de aprendizaje automático; y es capaz de caracterizar cada problema y escoger el algoritmo más adecuado.

5:

Implementar aplicaciones de aprendizaje automático utilizando las librerías de software existente y el

hardware adecuado (CPU, GPU o clúster).

6:

Conocer algunos de los resultados más recientes, los problemas abiertos y las oportunidades que ofrece el aprendizaje automático con grandes volúmenes de datos.

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. Por otro lado plantea retos importantes, como el elevado coste computacional de ciertos métodos.

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.

Contexto y competencias

Sentido, contexto, relevancia y objetivos generales de la asignatura

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. 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.

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.

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

1:

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 multidisciplinares, 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 multidisciplinario 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 interdisciplinares 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 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.

2:

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.

Importancia de los resultados de aprendizaje que se obtienen en la asignatura:

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.

Evaluación

Actividades de evaluación

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

1:

Realización y presentación de trabajos. Estudio de un tema relacionado con la asignatura, elaboración de un informe sobre el mismo y su presentación en clase. [100%]. Resultados de aprendizaje: 1, 2, 3, 4, 5 y 6

Actividades y recursos

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.
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.

Actividades de aprendizaje programadas (Se incluye programa)

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

1:
Contenidos a desarrollar

1. Introducción.
2. Resumen: Fundamentos del aprendizaje automático.
3. Redes neuronales. Deep Learning.
4. Gradiente Descendiente Estocástico.
5. Computación distribuida. MapReduce.
6. Aprendizaje secuencial a partir de flujos de datos.
7. Approximate Nearest Neighbour. Locality Sensitive Hashing.

2:

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, aproximadamente, de actividades presenciales (clases magistrales incluyendo seminarios profesionales, resolución de problemas y casos, y prácticas de laboratorio).
- 30 horas de trabajo en grupo.
- 10 horas de trabajo y estudio individual efectivo.
- 5 horas dedicadas a distintas pruebas de evaluación.

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.

Referencias bibliográficas de la bibliografía recomendada