

GUIA DOCENTE

MÁSTER UNIVERSITARIO FINANZAS Y BANCA

Ciencia De Datos Masivos Para Las Finanzas

Profesor: Luz Mary Pinzon, Ferran Carrascosa

E-mail: luzmary.pinzon@upf.edu; Ferran.carrascosa@upf.edu

Descripción de la asignatura

- **Tipo de asignatura:** OBLIGATORIA
- **Créditos:** 4 ECTS

Este curso es una introducción al análisis de datos y el aprendizaje automático, con aplicaciones para las finanzas. Primero nos enfocamos en problemas que son independientes del contexto o área de aplicación específicos; por ejemplo: cómo determinar qué modelo funciona mejor entre un conjunto de candidatos, cómo estimar correctamente la precisión de un modelo, cómo hacer que nuestros modelos se generalicen bien a nuevos datos.

Presentamos un marco estadístico mínimo utilizado para comprender el aprendizaje y luego profundizamos en dos problemas principales del aprendizaje supervisado: regresión (predicción de números) y clasificación (predicción de etiquetas no numéricas).

OBJETIVOS

- Competencias generales:
 - Analizar y manipular datos: lectura, limpieza, visualización y exploración de datos.
 - Aprender desde una perspectiva matemática y estadística; Problemas de aprendizaje supervisado, no supervisado y por refuerzo.
 - Modelado: selección, evaluación y comparación de modelos. Sesgo y varianza de un estimador.
 - Problemas de regresión y clasificación desde el punto de vista del Machine Learning.
 - Machine Learning con pipelines usando Python.
- Métodos y conceptos específicos:
 - Estimación de una relación: modelos, funciones de hipótesis y pérdidas, sesgo y varianza, sobreajuste y desajuste.
 - Entrenamiento de un modelo: descenso de gradiente

La información recogida en este documento es a título informativo y puede sufrir modificaciones en las adaptaciones de cada curso académico. La Guía definitiva estará a disposición de los estudiantes matriculados en el entorno virtual antes del inicio de cada asignatura

- Realización de la selección del modelo mediante retención, validación cruzada y bootstrap. Ajuste de hiperparámetros.

- Herramientas:

- Usar Python and Jupyter notebooks
- Familiaridad con data libraries, incluidos pandas, pyplot, seaborn.
- Familiaridad con las machine Learning libraries, incluido sklearn.

Prerequisitos:

Haber superado el curso introductorio de Python(Brush up)

Datos masivos

Contenidos

- Introducción a la analítica de datos y el aprendizaje automático.
 - estimadores, hipótesis y función de pérdida, error reducible e irreducible, entrenamiento y prueba de un modelo.
 - Sesgo y varianza, sobreajuste y desajuste, ejemplos.
- Leer y manipular datos, imputación de datos.
- Clasificación
 - Árboles de decisión. Caso de uso: segmentación de carteras
 - Modelos predictivos con Random Forest i Gradient Boosting Machine. Casos de uso: propensión de contratación, modelos de ingresos

Metodología docente

Las clases son tanto conferencia por parte del profesor como laboratorios prácticos, en una proporción aproximada del 50-50%. Las clases de laboratorio complementan las de teoría, familiarizando a los estudiantes con el lenguaje de programación Python y sus principales bibliotecas de ciencia de datos.

Las competencias, los resultados del aprendizaje, los elementos de evaluación y la calidad del proceso de aprendizaje incluidos en este Plan Docente no se verán afectados si durante el trimestre académico el modelo docente tiene que cambiar a un modelo híbrido (combinación de presencial y sesiones en línea) o a un modelo completo en línea.

Evaluación

- Quizes.
- Proyecto de trabajo (40%)
- Examen final (60%). Contiene preguntas sobre teoría únicamente.

Solo el examen final participación se transfiere a una eventual repetición. No hay repetición para los quizes o proyectos

La información recogida en este documento es a título informativo y puede sufrir modificaciones en las adaptaciones de cada curso académico. La Guía definitiva estará a disposición de los estudiantes matriculados en el entorno virtual antes del inicio de cada asignatura

Evaluación regular

A los estudiantes que no superen la evaluación regular se les permite hacer un segundo examen de recuperación cuya nota reemplaza la nota del examen.

En caso de inconveniente para presentarse al examen, deberá informar al profesor correspondiente y al (los) director (es) del programa para que estudien la posibilidad de reprogramarlo.

Se requiere que los estudiantes asistan al 80% de las clases como mínimo.

PROFESORADO

Luz Mary Pinzón S.

Doctorado en Estadística, Universidad de Salamanca-España. Magíster en Estadística, Universidad Nacional de Colombia. Licenciada en Matemáticas, Universidad Pedagógica Nacional de Colombia.

Experta en modelos estadísticos aplicados, análisis e inferencia multivariada.

Aplicaciones de Machine Learning in Python y R.

Consultora Senior en Estadística aplicada con experiencia en medio ambiente, educación, investigación social y marketing.

Profesora de estadística y matemáticas aplicadas a la ingeniería, economía, contaduría, marketing, finanzas.

Ferran Carrascosa Mallfrè

Gerente Data Scientist del equipo de Risk Analytics en CaixaBank. Apasionado del análisis de datos con veinte años de experiencia laboral, inicialmente 5 años en Estudios de Mercado y posteriormente 10 años en la consultora AIS de Aplicaciones de Inteligencia Artificial en los sectores de Banca y Retail. Actualmente como gerente metodológico en el equipo de innovación en Big Data en Banca. Quince años de experiencia como Profesor Asociado en el departamento de Economía y Empresa en la Universitat Pompeu Fabra.

BIBLIOGRAFIA (obligatoria/ recomendada)

“Introducción al aprendizaje estadístico”(“An introduction to Statistical Learning”), disponible gratuitamente en línea.

En el libro encontrará la mayoría de los temas teóricos cubiertos por este curso y muchos más.

“Elements of Statistical Learning” más Avanzado, también disponibles gratuitamente en línea.

Ian Goodfellow and Yoshua Bengio and Aaron Courville. Deep Learning. deeplearningbook.org. Un libro mucho más avanzado y también disponible en línea.

La información recogida en este documento es a título informativo y puede sufrir modificaciones en las adaptaciones de cada curso académico. La Guía definitiva estará a disposición de los estudiantes matriculados en el entorno virtual antes del inicio de cada asignatura