Guía docente
200614 - MCI - Métodos de Computación Intensiva

Última modificación: 10/06/2016

Unidad responsable: Facultad de Matemáticas y Estadística
Unidad que imparte: 1004 - UB - Universitat de Barcelona.
715 - EIO - Departamento de Estadística e Investigación Operativa.

Titulación: MÁSTER UNIVERSITARIO EN ESTADÍSTICA E INVESTIGACIÓN OPERATIVA (Plan 2013). (Asignatura optativa).

Curso: 2016 **Créditos ECTS:** 5.0 **Idiomas:** Castellano

PROFESORADO

Profesorado responsable: JORDI OCAÑA REBULL

Otros: Segon quadrimestre:
PEDRO FRANCISCO DELICADO USEROS - A
JORDI OCAÑA REBULL - A

CAPACIDADES PREVIAS

Familiaridad con conceptos básicos de cálculo en una y varias variables. Formación de nivel medio en probabilidades e inferencia. Dominio del entorno de trabajo estadístico y programación R (material para preparación previa: cualquier buen curso de autoaprendizaje de R, como http://www.ub.edu/stat/docencia/EADB/Curso%20basico%20de%20R.htm).

REQUISITOS

"Fundamentos de Inferencia Estadística" o "Inferencia Estadística Avanzada"
"Computación en Estadística y en Optimización"

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:
MESIO-CE2. CE-2. Capacidad para dominar la terminologia propia de algún ámbito en el que sea necesaria la aplicación de modelos y métodos estadísticos o de investigación operativa para resolver problemas reales.
MESIO-CE3. CE-3. Capacidad para formular, analizar y validar modelos aplicables a problemas de índole práctica. Capacidad de seleccionar el método y/o técnica estadística o de investigación operativa más adecuado para aplicar dicho modelo a cada situación o problema concreto.
MESIO-CE6. CE-6. Capacidad para utilizar el software más adecuado para realizar los cálculos necesarios en la resolución de un problema.
MESIO-CE8. CE-8. Capacidad de discutir la validez, el alcance y la relevancia de estas soluciones y saber presentar y defender sus conclusiones.
MESIO-CE9. CE-9. Capacidad para implementar algoritmos de estadística e investigación operativa.
MESIO-CE4. CE-4. Capacidad de utilizar los diferentes procedimientos de inferencia para responder preguntas, identificando las propiedades de los diferentes métodos de estimación y sus ventajas e inconvenientes, adaptados a una situación concreta y con un contexto específico.

Transversales:
CT1a. EMPRENDIMIENTO E INNOVACIÓN: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio.

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

METODOLOGÍAS DOCENTES

Enseñanza presencial
==========
La enseñanza presencial se estructura en sesiones teórico-prácticas. Se prevé que un 50% de las sesiones se impartirán en el aula de teoría dotada de medios informáticos y de proyección, mientras que el otro 50% tendrá el carácter de prácticas dirigidas o talleres, y se tendrá que impartir en un aula de informática.

En la vertiente teórica de las sesiones se presentan y discuten los conceptos teóricos acompañados de ejemplos prácticos, utilizando diapositivas que previamente se pondrán a disposición del alumno.

El entorno de trabajo fundamental de las sesiones prácticas será R, del cual se supondrán conocimientos medios (uso del entorno y programación básica). También se introducirá otro soporte lógico cuando se considere adecuado.

Horas presenciales por tema: el contenido de los diversos temas se ha procurado que sea balanceado, de forma que se prevé invertir 12 horas (6 en aula de teoría + 6 en aula de ordenadores) por tema.

Enseñanza no presencial
==========
Consiste en el estudio y en la resolución de problemas teóricos y prácticos que el alumno deberá entregar a lo largo del curso.
Concretamente, las actividades desarrolladas serán:
- Estudio de materiales docentes, previo y/o posterior a cada sesión presencial.
- Análisis detallado de diversos conjuntos de datos (personalizados, específicos para cada alumno/a). Se procurará que cada conjunto de datos sirva como base de un caso de estudio de diversos métodos (p.e. cálculo de diversos intervalos de confianza bootstrap y realización de pruebas de permutaciones).
- Realización de ejercicios teóricos y prácticos sobre los métodos tratados. Los ejercicios prácticos requerirán la realización de tareas de programación en R.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

Capacidades a adquirir:
Adquisición de habilidades para estudiar mediante simulación toda clase de situaciones inferenciales y de modelización estadística.
Entender los fundamentos de la metodología bootstrap y saberla aplicar a la resolución de problemas estadísticos diversos.
Conocer los principales métodos de construcción de intervalos de confianza bootstrap.
Entender los fundamentos de los tests de permutaciones y adquirir las habilidades necesarias para aplicar estos conceptos en diversas situaciones de interés práctico.
Entender los fundamentos de los Métodos de Montecarlo basados en Cadenas de Markov, y adquirir las habilidades necesarias para aplicarlos, especialmente en la determinación de distribuciones posteriores bajo un enfoque Bayesiano.

Objetivo transversal fundamental: interiorizar adecuadamente y saber aplicar dos ideas fundamentales: a) el "Método de Montecarlo" como herramienta para estudiar las propiedades de los métodos estadísticos; b) el "Método de Montecarlo" como base de algunos métodos estadísticos.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo pequeño	15,0	12.00
Horas aprendizaje autónomo	80,0	64.00
Horas grupo grande	30,0	24.00

Dedicación total: 125 h

CONTENIDOS

-Tema 1. Metodología bootstrap

Descripción:
El principio "plug-in". Bootstrap y el método de Montecarlo.
Validez y fuentes de error. Bootstrap no paramétrico y paramétrico.
Estimación bootstrap del sesgo y de la varianza.
Intervalos de confianza bootstrap: Intervalos bootstrap-t o estudentizados; bootstrap-t simetrizado. Intervalos percentil y BCa.
Orden de error de los diversos intervalos bootstrap.

Objetivos específicos:
Ante un conjunto de datos obtenidos a partir de un diseño experimental u observacional concreto, se tiene que adquirir la capacidad de realitzar un análisis estadístico inferencial basado en la metodologia bootstrap.

Actividades vinculadas:
Exposición de conceptos teórico-práctica en clase, combinada con ejercicios prácticos con datos reales, a realitzar tanto en classe como en casa.

Competencias relacionadas:
MESIO-CE9. CE-9. Capacidad para implementar algoritmos de estadística e investigación operativa.
MESIO-CE8. CE-8. Capacidad de discutir la validez, el alcance y la relevancia de estas soluciones y saber presentar y defender sus conclusiones.
MESIO-CE6. CE-6. Capacidad para utilizar el software más adecuado para realizar los cálculos necesarios en la resolución de un problema.
MESIO-CE3. CE-3. Capacidad para formular, analizar y validar modelos aplicables a problemas de índole práctica. Capacidad de seleccionar el método y/o técnica estadística o de investigación operativa más adecuado para aplicar dicho modelo a cada situación o problema concreto.

Dedicación: 11h 15m
Grupo grande/Teoría: 7h 30m
Grupo mediano/Prácticas: 3h 45m

-Tema 2. Tests de permutaciones y de aleatorización

Descripción:
Suficiencia del estadístico ordinal. Tests condicionales exactos.
Aproximación de Montecarlo. Determinación del número de permutaciones aleatorias.
Algunos tests de permutaciones importantes a la práctica. Test de Mantel. Tests de permutaciones y tests bootstrap.
Tests de permutaciones basados en distancias. PERMANOVA.

Objetivos específicos:
Dada una hipòtesis nul·la, adquirir la capacidad de identificar qué queda invariante respecto de permutaciones y saber implementar el correspondiente test de permutaciones.

Actividades vinculadas:
Exposición de conceptos teórico-práctica en clase, combinada con ejercicios prácticos con datos reales, a realitzar tanto en classe como en casa.

Competencias relacionadas:
MESIO-CE9. CE-9. Capacidad para implementar algoritmos de estadística e investigación operativa.
MESIO-CE8. CE-8. Capacidad de discutir la validez, el alcance y la relevancia de estas soluciones y saber presentar y defender sus conclusiones.
MESIO-CE6. CE-6. Capacidad para utilizar el software más adecuado para realizar los cálculos necesarios en la resolución de un problema.
MESIO-CE3. CE-3. Capacidad para formular, analizar y validar modelos aplicables a problemas de índole práctica. Capacidad de seleccionar el método y/o técnica estadística o de investigación operativa más adecuado para aplicar dicho modelo a cada situación o problema concreto.

Dedicación: 11h 15m
Grupo grande/Teoría: 7h 30m
Grupo mediano/Prácticas: 3h 45m

- Tema 3. Especialización de algoritmos de optimización en la estimación de parámetros

Descripción:
1. Introducción a los métodos de optimización.
2. Optimización cíclica por coordenadas. Estimadores Lasso en regresión.
3. Método de descenso del gradiente. Algoritmo de backpropagation en redes neuronales.
4. Métodos de Newton-Raphson y Fisher Scoring. Estimación máximo verosímil de parámetros y de sus matrices de varianzas.
4.1 Estimación máximo verosímil en el modelo de regresión lineal generalizado (GLM). Iteratively Reweighted Least Squares.
4.2 Versión no paramétrica del GLM mediante máxima verosimilitud local.

Dedicación: 11h 15m
Grupo grande/Teoría: 7h 30m
Grupo mediano/Prácticas: 3h 45m

- Tema 4. Algoritmo EM

Descripción:
1. Estimación máximo verosímil en modelos con variables latentes.
2. Ejemplos: Datos faltantes, estimación de modelos de mixtura, variables truncadas o censuradas.

Dedicación: 11h 15m
Grupo grande/Teoría: 7h 30m
Grupo mediano/Prácticas: 3h 45m

SISTEMA DE CALIFICACIÓN

La evaluación del alumnado se realizará en función de:
- Ejercicios realizados y entregados durante el curso (50%)
- Examen práctico sobre los conceptos teóricos trabajados durante el curso (50%)

NORMAS PARA LA REALIZACIÓN DE LAS PRUEBAS.

Para ser evaluados se requerirá al alumnado una asistencia mínima al 80% de las clases.

BIBLIOGRAFÍA

Básica:

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York [etc.]: Springer, cop. 2009. ISBN 978-0-387-84857-0.
- Gamerman, Dani; Lopes, Hedibert F. Markov chain Monte Carlo: stochastic simulation for bayesian inference. 2nd ed. Chapman & Hall/CRC, 2006. ISBN 1-58488-587-4.
- Manly, Bryan F.J. Randomization, bootstrap and Monte Carlo methods in biology. 3rd ed. Chapman & Hall/CRC, 2007.
- Robert, Christian P.; Casella, George. Monte Carlo statistical methods. 2nd ed. Springer, 2004. ISBN 0387212396.
- Santner,Thomas J.; Williams, Brian J.; Notz, William. The Design and analysis of computer experiments. Springer (Springer Series in Statistics), 2003. ISBN 0387954201.
- Kroese, Dirk P.; Taimre, Thomas; Botev, Zdravko I. Handbook of Monte Carlo Methods. New Jersey: John Wiley & Sons, 2011. ISBN 978-0-470-17793-8.
- Efron, B.; Tibshirani, R. An introduction to the bootstrap. Chapman & Hall, 1993.
- Lange, Kenneth. Numerical analysis for statisticians. New York [etc.]: Springer, cop. 1999. ISBN 0-387-94979-8.
- Hastie, Trevor; Tibshirani, Robert; Wainwright, Martin. Statistical learning with sparsity: The Lasso and Generalizations [en línea]. Boca Raton, FL: Chapman & Hall/CRC, 2015Disponible a: http://site.ebrary.com/lib/upcatalunya/detail.action?docID=11057047. ISBN 978-1-4987-1216-3.

Complementaria:

- Gentle, J. E. Elements of computational statistics [en línea]. Springer, 2002Disponible a: http://link.springer.com/book/10.1007/b97337. ISBN 0387954899.
- Gentle, J. E.; Härdle, W.; Mori, Y. (editors). Handbook of computational statistics: concepts and methods. Springer, 2004. ISBN 3540404643.
- Hjorth, J. S. Urban. Computer intensive statistical methods validation model selection and bootstrap. Chapman and Hall, 1994. ISBN 0412491605.
- Rao, C.R. (Editor). Computational statistics. North-Holland, 1993.
- Chernick, Michael R. Bootstrap methods: a guide for practitioners and researchers. 2nd ed. J. Wiley & Sons (Wiley Series in Probability and Statistics), 2008. ISBN 9780471756217.
- Good, Philip I. Permutation, parametric, and bootstrap tests of hypotheses [en línea]. New York: Springer (Springer Series in Statistics), 2005Disponible a: http://dx.doi.org/10.1007/b138696. ISBN 9780387271583.
- Chen, Ming-Hui; Shao, Qi-Man; Ibrahim, Joseph G. Monte Carlo methods in bayesian computation. Springer (Springer Series in Statistics), 2000. ISBN 0387989358.

RECURSOS

Otros recursos:
Campus virtual