



Universidad de Granada

GRADO EN ESTADÍSTICA
PROPUESTA DE TEMA PARA TRABAJOS FIN DE GRADO
CURSO ACADÉMICO 2014/2015

Título del trabajo: <i>Método estadístico para inferir el mejor número de clusters usando validación biológica</i>
Tutor/a: <i>Rocío Romero Zaliz</i>
Cotutor/a: -
Departamento responsable: <i>Ciencias de la Computación e Inteligencia Artificial</i>
Perfil y número de estudiantes al que va dirigido (máximo 3): <i>1</i> <i>(Rellenar sólo en caso de que la propuesta esté realizada a través de estudiante)</i>
Estudiante que propone el trabajo (Nombre, Apellidos, DNI):
Tipo de trabajo <i>(consultar (*)) Creación y/o empleo de herramientas informáticas para su uso en Estadística</i>
Competencias <i>(estas son la mínimas; consultar (**)) si se considera añadir otras</i> <i>Competencias generales:</i> G02, G03, G04, G05, G06, G07, G08, G09. <i>Competencias específicas:</i> E01, E02, E03, E04, E05, E08, E09, E10.
<ul style="list-style-type: none">• Resultados de aprendizaje (estos son los mínimos; añadir otros si se considera)• Adquirir competencias globales ligadas al desarrollo y aplicación de los conocimientos del Grado.• Adquirir competencias ligadas a la búsqueda y organización de información y documentación relevante sobre el tema objeto de estudio.• Aplicar el "pensamiento estadístico" y tener capacidad para enfrentarse a las distintas etapas de un estudio estadístico (desde el planteamiento del problema hasta la exposición de resultados).• Saber presentar, de forma escrita y oral, la memoria, los resultados y las conclusiones del trabajo realizado.• Aplicar las nociones de informática adquiridas en la resolución de un problema estadístico.
Antecedentes y resumen del tema propuesto: Determinar el número de grupos en un conjunto de datos es un problema frecuente en la agrupación de datos, y es una cuestión distinta del proceso de resolver realmente el problema de la agrupación. Para una cierta clase de algoritmos de agrupación (como ser, k-medias), hay un parámetro comúnmente conocida como k que especifica el número de grupos de detectar. La correcta elección de k es a menudo ambigua, con interpretaciones dependiendo de la forma y la escala de la distribución de puntos en un conjunto de datos. Además, el aumento de k sin penalidad siempre será reducir la cantidad de error en la agrupación resultante, para el caso extremo de error cero si cada punto de datos se considera su propio clúster (es decir, cuando k es igual al número de puntos de datos, n). Intuitivamente entonces, la elección óptima de k lograr un equilibrio entre la máxima compresión de los datos con un solo grupo, y la máxima precisión asignando a cada punto de datos a su propio grupo. En problemas de biología computacional es necesario tener en cuenta datos externos para realizar una correcta agrupación de los datos. Esta información externa puede provenir de diferentes fuentes, tales como repositorios biológicas.



Breve descripción de las actividades presenciales y no presenciales a realizar:

Actividades presenciales (15-30%)	Planteamiento, orientación y supervisión	80 horas
	Exposición del trabajo	10 horas
	Otras:	
Actividades no presenciales (70-85%)	Preparación del trabajo	150 horas
	Elaboración de la memoria	60 horas
	Otras:	
Total (12 ECTS)		300 horas

Objetivos que se pretenden alcanzar:

El objetivo de este proyecto fin de grado es implementar un módulo en R (Bioconductor) que permita seleccionar el número de clúster óptimo teniendo en cuenta la anotación biológica del conjunto de datos base.

Bibliografía básica para la puesta en marcha del trabajo:

1. Clustering in Bioinformatics and Drug Discovery. MacCuish, John David. Chapman & Hall/CRC Mathematical & Computational Biology. 2010.
2. R for Statistics. Pierre-Andre Cornillon, Arnaud Guyader, Francois Husson, Nicolas Je- gou, Julie Josse, Maela Kloareg, Eric Matzner-Lober, Laurent Rouvière. Chapman and Hall/CRC. 2012.

Tipo de trabajo (*):

1. Estudio de profundización en algún tema concreto de Estadística, o como proyecto de aplicación de la misma a estudios o problemas de otros ámbitos científicos o sociales.
2. Realización completa de todas las fases de un proyecto estadístico, bien con auxilio de prácticas en empresas o con prácticas propuestas y dirigidas por el tutor.
3. Estudio de casos, teóricos o prácticos, relacionados con la Estadística.
4. Elaboración de un informe o un proyecto de naturaleza profesional.
5. Elaboración de un plan de empresa.
6. Simulación de encargos profesionales.
7. Trabajos bibliográficos sobre el estado actual de una temática relacionada con la Estadística.
8. Creación y/o empleo de herramientas informáticas para su uso en Estadística.
9. Trabajos de inicio a la investigación.
10. Trabajos cuya finalidad sea la divulgación de la Estadística en diversos contextos.
11. Trabajos sobre Historia de la Estadística.
12. Trabajos relacionados con la docencia de la Estadística.

Competencias ()**

Competencias generales:



- G01.** Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.
- G02.** Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.
- G03.** Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- G04.** Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.
- G05.** Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- G06.** Saber utilizar herramientas de búsqueda de recursos bibliográficos.
- G07.** Poder comunicarse en otra lengua de relevancia en el ámbito científico.
- G08.** Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.
- G09.** Fomentar y garantizar el respeto a los Derechos Humanos, a los principios de accesibilidad universal, igualdad, y no discriminación; y los valores democráticos, de la cultura de la paz y de igualdad de género.

Competencias específicas:

- E01.** Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.
- E02.** Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.
- E03.** Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.
- E04.** Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.
- E05.** Comprender la importancia de la Investigación Operativa como metodología de optimización, toma de decisiones y diseño de modelos particulares para la resolución de problemas en situaciones específicas.
- E06.** Comprender y utilizar básicamente el lenguaje matemático.
- E07.** Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.
- E08.** Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.



Universidad de Granada

**GRADO EN ESTADÍSTICA
PROPUESTA DE TEMA PARA TRABAJOS FIN DE GRADO
CURSO ACADÉMICO 2014/2015**

E09. Conocer los conceptos básicos y habilidades propias de un ámbito científico o social en el que la Estadística o la Investigación operativa sean una herramienta fundamental.

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.