



Aprendizaje estadístico y métodos de ensamblado

(Fecha última actualización: 04/10/2024)

Máster Universitario en Estadística Aplicada. Curso académico 2024-2025	
Título	Aprendizaje estadístico y métodos de ensamblado
Tipo	INVESTIGACIÓN <input checked="" type="checkbox"/> ORIENTACIÓN PRÁCTICA <input type="checkbox"/>
Número de alumnos admitidos	1
Profesor(es)/ email	María Dolores Martínez Miranda (mmiranda@ugr.es)
Descripción	<p>El aprendizaje estadístico (statistical learning) se refiere a un conjunto de herramientas para modelizar y comprender datos complejos. Tales herramientas suelen clasificarse en supervisadas o no supervisadas. El aprendizaje estadístico supervisado engloba métodos predictivos, en los que una de las variables que intervienen en el problema se define como variable respuesta. Dependiendo del tipo de respuesta se diferencia entre clasificación (respuesta categórica) o regresión (respuesta cuantitativa). El aprendizaje no supervisado busca descubrir relaciones y estructuras en los datos y, aunque intervienen variables (inputs) no se considera ninguna variable de respuesta (output). Con la explosión del denominado "Big Data", el aprendizaje estadístico se ha convertido en tema de gran actualidad e interés en muchas disciplinas, siendo sus herramientas muy demandadas en el mundo empresarial.</p> <p>Independientemente del método de aprendizaje estadístico que nos planteemos utilizar, la elección del modelo concreto y su nivel de complejidad es crucial. En todo problema de aprendizaje estadístico se busca reducir el sesgo y la varianza. Una mayor flexibilidad en el modelo, así como la incorporación de diversas variables relevantes al problema, permitirán reducir el sesgo, sin embargo, esto conllevará un incremento de la varianza, implicando posiblemente un sobreajuste no deseado. Entender y encontrar el equilibrio entre varianza y sesgo es clave en la aplicación de este tipo de métodos. Los denominados métodos de ensamblado hacen uso de distintas técnicas estadísticas y matemáticas, como el bootstrap o los algoritmos de optimización, para mejorar las capacidades predictivas de los modelos de aprendizaje estadístico, permitiendo a su vez encontrar un equilibrio óptimo entre sesgo y varianza. En términos generales, estos métodos buscan acoplar, o ensamblar, una serie de modelos de base hasta alcanzar un modelo en el que se consiga el deseado equilibrio entre sesgo y varianza. Se suele usar la terminología modelos de aprendizaje débil para los modelos de base y modelo de aprendizaje fuerte para el modelo final. Entre los métodos de ensamblado más populares se encuentran el denominado bagging, que combina el ensamblado y el bootstrap, el boosting y el stacking.</p> <p>El aprendizaje estadístico no debe usarse como una "caja negra". Por lo general ningún método de aprendizaje estadístico funcionará bien en todos los casos y elegir el más adecuado requerirá un profundo conocimiento de este. Este trabajo supone una introducción al aprendizaje estadístico desde una perspectiva estadística matemática. Tras una adecuada contextualización se hará hincapié en los métodos de aprendizaje supervisados, y en los métodos de ensamblado como herramienta para encontrar un equilibrio óptimo entre flexibilidad y complejidad.</p>
Objetivos particulares	<ul style="list-style-type: none"> • Aprender los conceptos y terminología básica del aprendizaje estadístico, así como las diferencias entre aprendizaje supervisado y no supervisado.



	<ul style="list-style-type: none"> • Describir las bases y fundamentos de algunos de los métodos de aprendizaje supervisado más relevantes. • Comprender el problema de equilibrar sesgo y varianza y su relación con la complejidad y flexibilidad de los modelos de aprendizaje estadístico. • Estudiar medidas para la evaluación y validación de modelos, incluyendo validación cruzada y bootstrap. • Comprender la motivación y objetivos generales de los métodos de ensamblado en el contexto de los modelos de aprendizaje estadístico supervisado. • Conocer y describir algunos de los métodos de ensamblado más relevantes. • Realizar aplicaciones en R y/o Python con datos reales y/o simulados.
<p>Prerrequisitos y recomendaciones</p>	<p>Haber cursado alguna asignatura donde se incluyan contenidos de modelos de regresión y clasificación, desde la perspectiva formal de la estadística matemática. Además, se requiere destreza en el uso del entorno de programación y análisis estadístico R (y/o Python), incluyendo habilidad para estudiar paquetes específicos disponibles en CRAN y realizar tareas de programación de funciones.</p> <p>Para una introducción en el tema de esta propuesta se recomienda comenzar con una lectura de los libros de Hastie et al. (2009) y James et al. (2013) que aparecen en la bibliografía.</p>
<p>Plan de trabajo</p>	<ol style="list-style-type: none"> 1. Revisión de los aspectos teóricos. 2. Implementación computacional: estudio de herramientas disponibles R y creación de funciones originales en su caso. 3. Aplicaciones con datos reales y datos simulados. 4. Reflexión sobre limitaciones y extensiones de la metodología considerada. 5. Elaboración de las conclusiones del trabajo.
<p>Competencias generales y específicas</p>	<p>CG1 - Los titulados han de saber 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.</p> <p>CG7 - Los titulados han de realizar una contribución a través de una investigación original que amplíe las fronteras del conocimiento desarrollando un corpus sustancial, del que parte merezca la publicación referenciada a nivel nacional o internacional.</p> <p>CG8 - Los titulados deben ser críticos en el análisis, evaluación y síntesis de ideas nuevas y complejas.</p> <p>Competencias específicas:</p> <p>CE1 - Conocer métodos para el Análisis de Datos</p> <p>CE5 - Adquirir conocimientos avanzados en Inferencia Estadística</p> <p>CE17 - Adquirir capacidades de elaboración y construcción de modelos y su validación</p> <p>CE18 - Ser capaz de realizar un análisis de datos</p> <p>CE20 - Ser capaz de realizar una correcta representación gráfica de datos</p> <p>CE21 - Conocer, identificar y seleccionar fuentes estadísticas</p> <p>CE22 - Ser capaz de interpretar resultados a partir de modelos estadísticos</p> <p>CE24 - Ser capaz de extraer conclusiones y redactar informes</p> <p>CE25 - Ser capaz de identificar relaciones o asociaciones</p> <p>CE26 - Saber utilizar con destreza entornos de programación y análisis estadístico</p>
<p>Bibliografía</p>	<ol style="list-style-type: none"> 1. Bühlmann, P. and van de Geer, S. (2011). Statistics for High-Dimensional Data: Methods, Theory and Algorithms. Springer Series in Statistics. 2. Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning. Springer. 3. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer. 4. Joshi, A.V. (2023). Machine Learning and Artificial Intelligence. Springer. 5. Kuncheva, L. (2014). Combining Pattern Classifiers: Methods and Algorithms. Wiley 6. Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018). Foundations of Machine Learning. MIT Press. 7. Rhys, H. (2020). Machine Learning with R, the tidyverse, and mlr. New York, Manning.