

# Estudio de problemas de clasificación supervisada y de localización en redes mediante optimización matemática

Marta Baldomero Naranjo

Departamento de Estadística y Ciencia de los Datos

Universidad Complutense de Madrid

[martbald@ucm.es](mailto:martbald@ucm.es)

Director: Antonio Manuel Rodríguez Chía

Departamento de Estadística e Investigación Operativa

Universidad de Cádiz

[antonio.rodriguezchia@uca.es](mailto:antonio.rodriguezchia@uca.es)

**Palabras clave:** Máquina de Vectores Soporte, Optimización Matemática, Problemas de cubrimiento, Teoría de Localización.

**MSC Clasificación por Temas:** 90C11, 90C27, 90B10.

Los avances tecnológicos desarrollados en los últimos años han permitido desarrollar dispositivos y herramientas capaces de almacenar y organizar enormes volúmenes de datos cada segundo. Sin embargo, las técnicas necesarias para extraer conocimiento de esos grandes volúmenes están un paso por detrás. Uno de los campos que responde a esta necesidad es la Clasificación. El objetivo de esta disciplina es proporcionar una regla de decisión para clasificar a los individuos de una población en diferentes clases. Estos problemas pueden abordarse desde dos enfoques principales: Clasificación supervisada y Clasificación no supervisada. En concreto, la Clasificación supervisada considera que las clases en las que deben agruparse los individuos se conocen de antemano. Además, se sabe la clase a la que pertenece un conjunto de individuos (conjunto de entrenamiento). Mientras que la Clasificación no supervisada no requiere un conocimiento previo de las clases a las que pertenecen los individuos.

Por otra parte, la Teoría de Localización consiste en identificar la ubicación óptima de una o varias instalaciones en una zona determinada para cubrir la demanda de un conjunto dado de clientes, teniendo en cuenta las distancias/costes entre las instalaciones y los clientes, la demanda de los clientes y las restricciones existentes. Estos modelos pueden aplicarse en un amplio abanico de campos, como la cadena de suministro humanitario, los servicios de emergencia, los servicios públicos, la logística, las telecomunicaciones, etcétera.

Esta tesis doctoral ha abordado el estudio de varios problemas en los campos de la Clasificación supervisada y la Teoría de la Localización, empleando herramientas y técnicas propias de la Optimización Matemática. A continuación, se hace una breve descripción de estos problemas y de las metodologías propuestas para su análisis y resolución.

En el primer capítulo se discuten en detalle los fundamentos de la Clasificación supervisada y de la Teoría de la Localización, enfatizando los aspectos estudiados en esta tesis.

En los dos capítulos siguientes se analizan problemas de Clasificación. En particular, en el segundo capítulo, se proponen procedimientos exactos de resolución para varios modelos de Máquinas de Vectores de Soporte (SVM) con *ramp loss*. Este es un modelo conocido por limitar la influencia de los valores anómalos presentes en los datos. La principal aportación de este capítulo son resultados teóricos que permiten mejorar la formulación de los modelos existentes en la literatura y como consecuencia resolver de manera exacta problemas de mayor tamaño. Para ello, los modelos resultantes se analizan para obtener una cota inicial de los parámetros *M grande* incluidos en la formulación. Posteriormente, se proponen enfoques de resolución basados en tres estrategias para obtener valores más ajustados de dichos parámetros. Dos de ellas requieren resolver una secuencia de problemas de optimización continuos, mientras que la tercera utiliza el método de la relajación lagrangiana. Los procedimientos de resolución derivados son válidos para las formulaciones *ramp loss* con norma  $\ell_1$  y norma  $\ell_2$ . Estos algoritmos se han probado y comparado con los procedimientos de resolución existentes, tanto en conjuntos de datos simulados como reales, mostrando la eficacia de la metodología desarrollada. Los resultados obtenidos se encuentran publicados en Baldomero-Naranjo, Martínez-Merino y Rodríguez-Chía (2020).

El tercer capítulo presenta un nuevo clasificador basado en SVM que simultáneamente aborda la limitación de la influencia de los valores atípicos y la selección de características o atributos. La influencia de los valores atípicos se controla mediante el criterio *ramp loss*, mientras que el proceso de selección de características se lleva a cabo incluyendo una nueva familia de variables binarias y varias restricciones. El modelo resultante se formula como un modelo de programación entera mixta con parámetros *M grande*. En este capítulo, se analizan las características del modelo y se proponen dos algoritmos de resolución diferentes (exacto y heurístico). El rendimiento del clasificador obtenido se compara con varios clasificadores en diversos conjuntos de datos, mostrándose la efectividad del modelo propuesto, así como, la eficiencia del algoritmo heurístico desarrollado. Este trabajo se encuentra publicado en Baldomero-Naranjo, Martínez-Merino y Rodríguez-Chía (2021).

Los dos capítulos siguientes abordan diferentes versiones del problema de localización de máximo cubrimiento (MCLP) en redes. La versión clásica de este problema tiene como objetivo determinar la ubicación óptima de un conjunto de servicios para maximizar la demanda cubierta. Un cliente se considera cubierto si su distancia a un servicio es menor o igual a un radio fijado. En particular, en esta tesis se abordan dos variantes, una considerando que la longitud de las aristas de la red puede modificarse satisfaciendo una restricción de presupuesto y otra en la que hay incertidumbre en la demanda de los clientes.

En primer lugar, se presenta el MCLP en redes con mejoras permitiéndose la reducción de la longitud de las aristas. Este problema tiene como objetivo ubicar  $p$  instalaciones en los nodos de la red para maximizar la cobertura, considerando que la longitud de las aristas puede reducirse dentro de un presupuesto. Por lo tanto, tenemos que decidir la ubicación óptima de las  $p$  instalaciones y las reducciones óptimas de la longitud de las aristas. Por lo que sabemos, es la primera vez que este problema es estudiado en la literatura. Para resolverlo, proponemos tres formulaciones enteras mixtas y una fase de preprocesamiento para fijar variables y eliminar algunas de las restricciones. Además, analizamos las características de estas formulaciones para fortalecerlas proponiendo desigualdades

---

válidas. Por último, comparamos las tres formulaciones y sus correspondientes preprocesos y desigualdades válidas, probando su rendimiento en diferentes conjuntos de datos. Pueden encontrarse más detalles en Baldomero-Naranjo, Kalcsics, Marín y col. (2022).

El siguiente capítulo, también considera un MCLP, aunque desde la perspectiva de la incertidumbre. En concreto, este capítulo aborda una versión del MCLP de una sola instalación en una red en la que la demanda está distribuida a lo largo de las aristas y solo se conoce una cota superior y una cota inferior de dicha demanda. Proponemos un modelo que minimiza el peor de los casos (*minmax regret*) en el que la instalación del servicio puede situarse en cualquier punto de la red. Además, presentamos dos algoritmos polinómicos para encontrar la ubicación que minimiza el peor de los casos suponiendo que la realización de la demanda es una función constante desconocida o una función lineal desconocida en cada arista. También incluimos dos ejemplos ilustrativos y un estudio computacional para mostrar el potencial de la metodología propuesta. Los resultados obtenidos se encuentran publicados en Baldomero-Naranjo, Kalcsics y Rodríguez-Chía (2021).

La tesis doctoral finaliza con las conclusiones de la investigación realizada y la presentación de futuras líneas de trabajo.

Esta tesis doctoral se ha desarrollado en el programa de Doctorado en Matemáticas Interuniversitario impartido, conjuntamente, por las Universidades de Almería, Cádiz, Granada, Jaén y Málaga.

## Agradecimientos

Esta investigación se ha financiado con los proyectos: MTM2016- 74983-C2-2-R y RED2018-102363-T (Agencia Estatal de Investigación, Española y el Fondo Europeo de Desarrollo Regional (FEDER)), FEDER-UCA18-106895 y P18-FR-1422 (Junta de Andalucía y FEDER) y el contrato predoctoral UCA/REC01VI/2017 (Universidad de Cádiz). Agradecer también la financiación recibida por la BritishSpanish Society y Telefónica (Scholarship Programme 2018).

## Acerca de la autora



**Marta Baldomero Naranjo** es profesora ayudante doctora en la Universidad Complutense de Madrid. Sus principales intereses son la Clasificación Supervisada y la Teoría de la Localización. Se graduó en Matemáticas por la Universidad de Cádiz en 2016. Obtuvo el Máster en Matemáticas (2017) y el Doctorado en Matemáticas (2021) por las Universidades de Almería, Cádiz, Granada, Jaén y Málaga.

---

## Referencias

- Baldomero-Naranjo, Marta, Jörg Kalcsics, Alfredo Marín y Antonio M. Rodríguez-Chía (2022). «Upgrading edges in the maximal covering location problem». En: *European Journal of Operational Research* 303.1, págs. 14-36. DOI: [10.1016/j.ejor.2022.02.001](https://doi.org/10.1016/j.ejor.2022.02.001).
- Baldomero-Naranjo, Marta, Jörg Kalcsics y Antonio M. Rodríguez-Chía (2021). «Minmax regret maximal covering location problems with edge demands». En: *Computers & Operations Research* 130, pág. 105181. DOI: [10.1016/j.cor.2020.105181](https://doi.org/10.1016/j.cor.2020.105181).
- Baldomero-Naranjo, Marta, Luisa I. Martínez-Merino y Antonio M. Rodríguez-Chía (2020). «Tightening big Ms in integer programming formulations for support vector machines with ramp loss». En: *European Journal of Operational Research* 286.1, págs. 84-100. DOI: [10.1016/j.ejor.2020.03.023](https://doi.org/10.1016/j.ejor.2020.03.023).
- (2021). «A robust SVM-based approach with feature selection and outliers detection for classification problems». En: *Expert Systems with Applications* 178, pág. 115017. DOI: [10.1016/j.eswa.2021.115017](https://doi.org/10.1016/j.eswa.2021.115017).
-