

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

Metodología de clasificación lineal por partes: piecewise Linear Support Vector Machines (PL-SVM)

Tesis para optar al grado de Magíster en Gestión de Operaciones
Memoria para optar al título de Ingeniero Civil Industrial

Por:

Andrés Rodrigo Thiele Horta

Profesor Guía: Richard Weber Haas

Santiago de Chile - Junio 2011

Miembros de la Comisión: Máximo Bosch Passalacqua, Ricardo Montoya Moreira y Sebastián Maldonado Alarcón

Resumen . .	4
Tesis con restricción de acceso en línea, según petición de su autor . .	5

Resumen

El siguiente trabajo de tesis tiene como objetivo plantear una nueva metodología de clasificación binaria que permita obtener resultados predictivos equivalentes a los de los métodos tradicionales, pero salvaguardando caer en el fenómeno de *caja negra*. Este fenómeno se presenta cuando la técnica de clasificación no entrega interpretabilidad de la forma en que se efectúa la clasificación. El paradigma que se presenta es que, a mayor precisión del método, menor es la interpretabilidad que se obtiene, y viceversa.

Una de las técnicas de aprendizaje estadístico con mejores desempeños predictivos es Support Vector Machines. Pueden construir un estimador de clasificación simple e interpretable (SVM-Lineal), o uno complejo (SVM-RBF) con el fenómeno de *caja negra*. El estimador simple asume que la función subyacente que discrimina las clases binarias es lineal, mientras que el otro supone lo contrario, produciendo un estimador que no es directamente interpretable. Para enfrentar este problema, se propone una metodología que construye un estimador lineal por partes. Globalmente, el estimador no es lineal, pero está compuesto por un número fijo de componentes lineales (SVM-Lineal). Cada componente lineal clasifica un subconjunto independiente de puntos del problema global. Conceptualmente, esta metodología trata de trazar un puente entre las dos técnicas tradicionales anteriores, manteniendo niveles de precisión comparables y salvaguardando la interpretabilidad.

En el 100% de las pruebas de precisión de clasificación binaria realizadas, se obtuvo resultados iguales o mejores que el estimador SVM-Lineal. Con la ventaja adicional de que el nuevo estimador entrega una frontera compuesta por distintas funciones lineales independientes entre sí, que a su vez caracterizan de manera distinta a grupos independientes de puntos. En vez de ofrecer un único perfil de comportamiento como lo hace SVM-Lineal, el modelo propuesto ofrece tantos perfiles como componentes integran el estimador lineal por partes.

La principal desventaja del método consiste en la etapa de segmentación de los datos binarios. Esta etapa hace la diferencia con respecto a las técnicas tradicionales, y su orden de complejidad es inclusive mayor al de una técnica de aprendizaje estadístico no lineal. Precisamente, una línea de trabajo a futuro consiste en optimizar el algoritmo de segmentación.

Tesis con restricción de acceso en línea, según petición de su autor

Autorizada por el autor, pero con restricción para ser publicada a texto completo en Cybertesis hasta el año 2016.