

Nova medida de influencia para avaliar a importancia de certos atributos nun problema de clasificación

Laura Davila-Pena¹, Ignacio García-Jurado², Balbina Casas-Méndez¹

¹Departamento de Estatística, Análise Matemática e Optimización. Universidade de Santiago de Compostela.

²Departamento de Matemáticas. Universidade da Coruña.

RESUMO

Un problema de clasificación consiste en predicir o valor dunha variable resposta cualitativa para un ou máis individuos, facendo uso dos valores que xa coñecemos de certas variables categóricas (características ou atributos) de tales individuos. Estas predicións baséanse no coñecemento obtido a través dunha mostra de individuos con valores coñecidos tanto dos atributos como da resposta.

O traballo que se está levando a cabo consiste na proposta dunha medida de influencia que nos permita coñecer cales son as características máis influíntes na clasificación. Esta medida de influencia está baseada no valor de Shapley para xogos cooperativos. No contexto da teoría de xogos, o valor de Shapley é unha regra que distribúe os beneficios da cooperación entre os axentes involucrados. Neste caso particular, emprégase para repartir a importancia entre os distintos atributos.

Nesta comunicación mostraremos unha aplicación da metodoloxía introducida a unha base de datos de pacientes galegos infectados por COVID-19. Con isto pretendemos estudar a influencia que certos factores demográficos e patoloxías previas tiveron en varios eventos de interese relacionados coa evolución da enfermidade.