

# Otimização robusta aplicada ao problema de agrupamento k-médias

Carina Moreira Costa<sup>1</sup>

<sup>1</sup> Trier University

O agrupamento k-médias é um método clássico de aprendizagem não-supervisionada com o objetivo de particionar um dado número de observações em k grupos. Em muitas aplicações modernas, entretanto, esse método sofre de erros de medição pois o agrupamento resultante representa então o agrupamento das observações contaminadas por erros ao invés de recuperar a estrutura verdadeira do agrupamento. Portanto, nós aplicamos técnicas de otimização robusta a fim de proteger o resultado do agrupamento contra erros de medição que podem estar presentes nas observações. Para isso, derivamos as versões estritamente robusta e  $\Gamma$ -robusta do problema de agrupamento k-médias. Visto que o problema nominal já é NP-hard, então geralmente métodos exatos não são viáveis em prática. Como um remédio, desenvolvemos métodos personalizados de direções alternadas para obter soluções viáveis de boa qualidade e rapidamente. Tais métodos são baseados na decomposição do espaço de busca do problema nominal como também das versões robustas. Nossos resultados numéricos revelam um aspecto interessante: a versão mais conservadora—estritamente robusta—teve melhor desempenho do que a versão menos conservadora  $\Gamma$ -robusta. Em particular, a versão estritamente robusta do problema k-médias é capaz de recuperar agrupamentos de dados originais mesmo quando somente observações incorretas são observadas.

Em colaboração com Jan Pablo Burgard e Martin Schmidt.