

Seleção de modelos para segmentação de sequências em alta dimensão

Florença Leonardi¹

¹ USP

Seleção estatística de modelos é uma área ativa de pesquisa em várias sub-áreas da estatística, e tem alcançado um grande interesse pela sua utilidade na análise de grandes conjuntos de dados, como uma forma automática de selecionar um modelo com um bom ajuste mas que não esteja super-dimensionado. De certa forma, há duas culturas dentro da área de seleção de modelos. Uma propõe critérios de ajuste como máxima verossimilhança ou mínimo custo com uma penalidade em relação à dimensão do modelo, o que geralmente são chamados de estimadores regularizados. A outra abordagem, baseada no paradigma de minimização do erro de predição, utiliza validação cruzada para obter o modelo com menor erro preditivo estimado. Nesta palestra faremos uma apresentação geral do problema de seleção de modelos, com ênfase em problemas de alta dimensão, e ilustraremos a aplicação destas duas culturas num problema de segmentação de sequências para estimação de pontos de recombinação em sequências genéticas.