

MODELAGEM FATORIAL PARA DETECTAR PADRÕES EM DADOS DE EXPRESSÃO DE GENES

Vinícius Diniz Mayrink¹

RESUMO DE MINI-CONFERÊNCIAS - JOVENS DOUTORES

Um problema de interesse na análise de dados obtidos a partir de microarrays é a identificação de características que indicam a atividade dos genes. Por exemplo, podemos querer avaliar se um grupo de genes, correlacionados em um conjunto de dados, está ativo e mostrando semelhante associação em outro conjunto de dados. O padrão exibido por uma coleção de probes, representando a seqüência de um gene no microarray, é uma fonte de informação interessante para o problema de classificar genes como presentes ou ausentes. Neste trabalho consideramos um modelo fatorial Bayesiano para identificar padrões de expressão genética exibidos em um conjunto de microarrays construídos para o mesmo tipo de amostra. Dados simulados e reais são usados para comparar o desempenho do modelo fatorial em relação a outros métodos descritos na literatura. Os resultados sugerem que o modelo fatorial supera os demais procedimentos na detecção da atividade dos genes. Este trabalho também aplica a análise fatorial para classificação de grupos de genes; o objetivo aqui será identificar os casos onde existe uma associação entre as expressões genéticas. A alteração no número de cópias (Copy Number Alteration - CNA) em regiões do genoma é um aspecto conhecido e com papel importante em vários tipos de câncer; neste contexto, iremos examinar um conjunto de genes apresentando CNA para tumores do câncer de mama, em seguida, utilizamos o modelo fatorial para identificar a existência ou não de CNA na mesma região do DNA para outros tipos de câncer.

¹Departamento de Estatística Universidade Federal de Minas Gerais