1 Introdução

Modelos sempre procuram aproximar a realidade da melhor maneira possível. Um modelo ideal é aquele que alcançaria esse objetivo com perfeição. Entretanto, essa exatidão é uma utopia. O desafio passa a ser então obter o modelo que melhor se aproxime da situação desejada.

Em um contexto supervisionado, podemos dividir o processo de modelagem em quatro etapas: escolha do modelo, especificação de parâmetros do modelo escolhido, escolha dos atributos e escolha dos dados que serão utilizados durante o treinamento. Todas essas etapas vêm sendo abordadas na literatura, entretanto, a menos explorada de todas é a escolha dos dados.

Selecionar dados específicos da amostra para treinar um modelo visando ganho no conjunto de generalização (fora-da-amostra) é um problema importante. Resultados interessantes de técnicas que envolvem seleção de dados podem ser encontrados em [1], [2], [3], [4] e [5]. Diversos algoritmos têm sido apresentados visando seleção inteligente de dados, ou seja, propor a utilização de subconjuntos do conjunto de treinamento para que o aprendizado seja maximizado e consiga-se encontrar resultados melhores para um determinado problema.

O algoritmo de LVQ*, proposto por Kohonen [8], tornou-se amplamente conhecido pela sua eficiência e facilidade de implementação. A proposta dessa dissertação é apresentar um algoritmo para seleção de dados em LVQ.

Baseado em um resultado mais geral [15], o algoritmo proposto tem como idéia central sortear aleatoriamente diversos modelos para que o erro apresentado

* Ao longo do texto usaremos a sigla LVQ como referência a Aprendizado por Quantização Vetorial. Esta sigla refere-se ao termo em inglês Learning Vector Quantization e é amplamente usada na literatura.

por cada padrão seja correlacionado com o erro de sua respectiva distribuição. Assim esperam-se encontrar os pontos que são realmente relevantes para o aprendizado do modelo. Esta dissertação está organizada como segue:

No capítulo 2 é apresentado o modelo proposto de seleção de dados em modelos LVQ. Os experimentos realizados são mostrados no capítulo 3. Foram realizados cinco experimentos com dados simulados e dois experimentos com bancos reais. Finalmente, a conclusão da dissertação é apresentada no capítulo 4. No apêndice A, a questão de seleção de dados é abordada e no apêndice B, é feita uma revisão teórica do modelo LVQ.