

# Resumo

*Machine Learning* (ML) é uma área da ciência cujo objectivo é o desenvolvimento de algoritmos de aprendizagem, gerando novo conhecimento a partir do já existente. Exemplos de problemas deste género são o diagnóstico automático da situação clínica de um doente ou a navegação autónoma de um *robot* num determinado ambiente. Estas tarefas são complexas já que usualmente quanto mais avançados os algoritmos são, maiores são os requisitos a nível de cálculo exigidos do hardware. Outra razão para a sua complexidade consiste no constante aumento da quantidade de informação a ser processada por estes algoritmos. Sendo muitos dos actuais algoritmos implementados seguindo uma abordagem sequencial, estes dificilmente são capazes de acompanhar as exigências a nível do poder de computação.

Entre os vários algoritmos de aprendizagem supervisionada encontram-se as *Support Vector Machines* (SVMs), conhecidas pelas suas capacidades de generalização. As SVMs são classificadores óptimos utilizados em muitas aplicações científicas nas áreas da engenharia, bioinformática, sistemas de informação, finanças, entre outras.

O princípio no qual as SVMs assentam consiste na pesquisa do hiperplano que maximiza a margem que separa as classes, sendo esta optimização equivalente a minimizar os erros empírico e de generalização. Consequentemente, as SVMs fazem uso do *Structural Risk Minimization* (SRM). As SVMs apresentam um excelente desempenho quer de classificação quer de regressão em problemas não lineares. Para tal são usadas funções (*Kernels*) que obedecendo às condições de *Mercer* permitem o mapeamento do espaço de entrada num espaço de maior dimensão. Neste espaço os dados são classificados de acordo com o hiperplano de separação máxima que, caso seja usado um *Kernel* não linear, corresponde a uma separação não linear no espaço inicial.

As SVMs exigem geralmente uma grande quantidade de operações matemáticas devido ao cálculo da matriz de *Kernel*. Outro ponto a ter em conta é que grande parte das implementações de SVMs existentes não fazem uso das capacidades

quer dos processadores com vários núcleos (*cores*) quer das placas gráficas (GPUs) disponíveis na actualidade.

O objectivo desta dissertação consiste principalmente no desenvolvimento e implementação de Support Vector Machines (SVMs) que façam uso do poder computacional disponível nas placas gráficas. Outro objectivo do trabalho é a criação de uma versão *multi-threaded* para o CPU que serve de referência à implementação para a placa gráfica. Ambas as SVMs são construídas com base no algoritmo SMO.

Embora existam implementações semelhantes, a que apresentamos neste trabalho inclui o novo *Kernel* UKF sendo também *Open Source*. Os resultados dos benchmarks efectuados quer em *datasets* reais quer em *datasets* do repositório de *Machine Learning* da Universidade da Califórnia, Irvine (UCI) demonstram que os classificadores desenvolvidos apresentam uma performance semelhante à conhecida LIBSVM. Outro aspecto positivo dos classificadores desenvolvidos são os tempos mais curtos quer de treino quer de classificação para grandes *datasets*, não esquecendo que a LIBSVM é actualmente o estado da arte da investigação nas SVMs.

Por fim, usámos a implementação para a placa gráfica no estudo de um problema que consiste no reconhecimento *off-line* de assinaturas. Trata-se de um problema complicado com muitas aplicações práticas. Os resultados obtidos na detecção de assinaturas falsas são promissores embora sejam necessários estudos futuros, já que a identificação de assinaturas forjadas é um problema complicado. Ao nível da identificação do autor de uma assinatura foram obtidos resultados excelentes.

No futuro esperamos melhorar o suporte para problemas multi-classe e aperfeiçoar a arquitectura usada no processo de identificação de assinaturas. Outros pontos a otimizar são o uso de heurísticas mais avançadas, *caching* de cálculos do *kernel*, vectorização e suporte para dupla precisão.