|  |
| --- |
| Universidade Federal de Pernambuco  Graduação em Ciência da Computação  Centro de Informática  2010.2 |
| Máquina de Vetores Suporte, estudo e análise de parâmetros para otimização de resultado |
| Proposta de Trabalho de Graduação |
| **Aluno**: Gilson Medeiros de Oliveira Junior ( gmoj@cin.ufpe.br)  **Orientador:** Ricardo Bastos Cavalcanti Prudêncio (rbcp@cin.ufpe.br) |
|  |

1. **Contextualização**

O uso de técnicas de inteligência artificial, mais especificamente algoritmos de aprendizagem para classificação de padrões, tem crescido bastante nos últimos anos pelo falo de auxiliarem na resolução de problemas complexos do mundo real que podem envolver um elevado número de variáveis como: análise de expressões genéticas, reconhecimento de imagens, voz, objetos em três dimensões, categorização de caracteres manuscritos, etc..

Dentre o conjunto existente de técnicas para fins de classificação, algumas se destacam, entre elas estão às máquinas de vetores suporte (SVMs *- Support Vector Machines*), devido principalmente ao resultado satisfatório obtido ao final do processo. De forma simplificada, as SVMs funcionam da seguinte maneira: dadas duas classes e um conjunto de pontos que pertencem a essas classes, o objetivo do algoritmo é encontrar um hiperplano ótimo, ou seja, um hiperplano capaz de separar o maior número de pontos de uma mesma classe do mesmo lado, maximizando a distância entre os pontos e o hiperplano de separação. Para isso a máquina de vetores suporte faz uso de uma função de *kernel*, está possui um conjunto de parâmetros que podem ser manipulados para obtenção de um melhor desempenho do classificador[1].

Para aplicar o classificador SVM, é necessário inicialmente treinar a máquina para que ela possa aprender sobre o domínio desejado, para isso, pode-se simplesmente dividir manualmente o domínio com suas entradas e respectivas saídas em dois subconjuntos, um deles contendo os exemplos de treinamento, este geralmente contém 50% a 90% do total, e o outro com os dados de teste. Contudo, essa divisão do domínio em dois subconjuntos pode não ter sido feita de forma adequada, ocasionando problemas como *overfitting* ou *underfitting*. Para tentar prevenir esse tipo de problema, existem algumas técnicas que podem ser aplicadas, dentre elas uma bastante conhecida é a chamada validação cruzada (*cross-validation*)[2].

Devido ao surgimento de novos problemas e aumento da complexidade dos já existentes, a proposta desse trabalho é avaliar o uso do SVM juntamente com a técnica de validação cruzada, na tentativa de encontrar os parâmetros mais promissores para obtenção do resultado ótimo.

1. **Objetivos**

O objetivo principal desse trabalho é fazer uma série de análises da aplicação do algoritmo SVM (*Support Vector Machine*) usando uma abordagem investigativa. Essa análise será feita aplicando a técnica de classificação escolhida utilizando a função de *kernel* RBF (Radial Basis Function) e variando os parâmetros da função assim como utilizando os seus valores default.

A partir desses experimentos e análise será possível auxiliar pessoas que não estejam familiarizadas com o uso de maquinas de suporte vetorial a escolher os melhores parâmetros para usar, sem que seja necessário testar um conjunto de parâmetros, com intuito de ganhar tempo.

1. **Cronograma**

Nesta seção, é apresentado o cronograma de atividades previsto para o desenvolvimento desse Trabalho de Graduação. Como qualquer atividade, este trabalho é suscetível a imprevistos e, portanto, mudanças poderão ocorrer. O quadro abaixo exibe a lista das atividades a serem realizadas para conclusão do trabalho proposto neste documento.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividades** | **Agosto** | | | | **Setembro** | | | | **Outubro** | | | | **Novembro** | | | | |
| Levantamento Bibliográfico |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Desenvolvimento e execução dos Testes |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Elaboração do Relatório |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Preparação da Apresentação |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. **Referências**

**[1]** A Practical Guide to Support Vector Classification <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.3096&rep=rep1&type=pdf>

[Online; acessado em 18-Agosto-2010].

**[2]** Almeida, F.F.M.Support Vector Machine

<http://www.dsc.ufcg.edu.br/~fabian/projeto/relatorio-svm-final-fabian.pdf>

[Online; acessado em 18-Agosto-2010].

# Assinaturas

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Gilson Medeiros de Oliveira Junior  
 **Orientando**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof.ª Ricardo Bastos Cavalcante Prudêncio  
 **Orientador**