

Exercício LibSVM

Objetivo: Entender o funcionamento da biblioteca LibSVM através da execução de experimentos de classificação.

1) Baixe e instale a libSVM no seu diretório. Disponível nos links abaixo

<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
<http://www.inf.ufpr.br/lesoliveira/download/libsvm-2.91.tar.gz>

2) Compile o código para gerar os executáveis

```
> make
```

3) Ler o documento disponível nos links abaixo

<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
<http://www.inf.ufpr.br/lesoliveira/download/guidesvm.pdf>

4) Baixe as bases de aprendizagem e teste disponíveis no link abaixo

<http://www.inf.ufpr.br/lesoliveira/padroes/data-libsvm.tar.gz>

Essa base contém duas classes rotuladas com +1 e -1. O arquivo treino.vet contém 528 linhas e o arquivo teste.vet contém 12000. O rótulo é dado pelos dois primeiros caracteres de cada linha (+1 ou -1).

5) Execute o script python que acompanha a libSVM, chamado easy.py (localizado no diretório tools). Esse script faz a busca pelos parâmetros do kernel Gaussiano (g) e da variável de custo (C). O script gera alguns arquivos. Liste quais são esses arquivos e explique o conteúdo dos mesmos

```
> ./easy treino.vet teste.vet
```

5.1) Reporte a taxa de reconhecimento no arquivo de teste.

5.2) Existe diferença com a taxa de reconhecimento no arquivo de treinamento? Se sim, explique porque.

5.3) Com base no arquivo de predição (.predict), monte a matriz de confusão.

6) Utilize os parâmetros encontrados pelo easy.py para treinar um svm que estima probabilidades. Para fazer isso você deve usar a opção -b e o comando svm-train. Use também os arquivos normalizados gerados pelo easy.py

```
> ./svm-train -b 1 -C ?? -g ?? TRAINING-SET modelo
```

7) Utilize o modelo aprendido para estimar probabilidade no passo abaixo e utilize o mesmo para reconhecer a base de teste. O arquivo de teste também deve ser o arquivo normalizado.

```
> ./svm-predict -b 1 TESTING-SET modelo output
```

7.1) Compare a taxa de reconhecimento e a matriz de confusão com os resultados do item 5. Alguma diferença relevante?

7.2) Dê uma idéia de como os valores das probabilidades podem ser utilizados para tornar o sistema de classificação mais robusto.