

*Métodos Matemáticos em Engenharia Química*

*Programa de Pós-Graduação em Engenharia Química - PPGEQ*

*Trabalho 03: Máquinas de Vetores de Suporte*

*Prof. Éliton Fontana*

**Parte I - SVM Lineares:** Obtenha um conjunto de dados utilizando o seguinte comando:

```
from sklearn.datasets import make_classification
X, y = make_classification(n_features=2, n_redundant=0, n_informative= 1,
n_clusters_per_class= 1, n_samples = 300, random_state = a )
```

Para o parâmetro  $a$ , utilize algum valor inteiro aleatório com 5 dígitos. Isto irá garantir que o código irá gerar sempre o mesmo conjunto para comparação. Estes dados serão utilizados para um algoritmo de classificação usando Máquinas de Vetores de Suporte.

A classe LinearSVC do sklearn executa um algoritmo de máquina de vetores de suporte linear com margem suave <sup>1</sup>. Este modelo pode ser importado e utilizado usando os comandos:

```
from sklearn.svm import LinearSVC
model = LinearSVC(loss = a, C = b, random_state = c)
```

Assim como para o comando `make_classification`, esta função permite a definição de um `random_state`, portanto pode-se fixar o parâmetro  $c$  como um inteiro qualquer para permitir a comparação dos resultados. Este valor será utilizado para geração de valores pseudo-aleatórios utilizados pelo algoritmo de otimização.

O valor de  $b$  corresponde ao parâmetro de regularização desejado (por padrão igual a 1), enquanto que o argumento `loss` define a função de perda, aceitando os argumentos `"hinge"` (função de articulação) ou `"squared_hinge"` (função de articulação elevada ao quadrado).

a) Fixe a função de perda como `"hinge"` e compare a resposta utilizando  $C = 0.001$  e  $C = 10000$  para o parâmetro de regularização<sup>2</sup>. Avalie mudanças no resultado (acurácia, matrix de confusão, no tamanho da margem, etc.) para os dois casos. Justifique estas mudanças com base na forma como o algoritmo realiza a classificação.

b) Fixe o parâmetro de regularização em  $C = 1$ . Compare a resposta utilizando as duas funções de perda. Avalie mudanças no resultado (acurácia, matrix de confusão, no tamanho da margem, etc.) para os dois casos. Justifique estas mudanças com base na forma como o algoritmo realiza a classificação.

---

<sup>1</sup><https://scikit-learn.org/stable/modules/svm.html>

<sup>2</sup>Caso estes valores gerem resultados onde a margem não pode ser visualizada por estar fora da extensão do gráfico, utilize valores maiores que 0.001 ou menores que 10000.

**Parte II - SVM Não-Lineares:** Obtenha um conjunto de dados utilizando o seguinte comando:

```
from sklearn.datasets import make_circles
X, y = make_circles(n_samples = 400, random_state = a, noise = b )
```

Para o parâmetro  $a$ , utilize algum valor inteiro aleatório com 5 dígitos. Para o parâmetro  $b$ , utilize algum valor real entre 0.05 e 0.15.

A classe SVC do sklearn executa um algoritmo de máquina de vetores de suporte com diferentes kernels não-lineares. Este modelo pode ser importado e utilizado usando os comandos:

```
from sklearn.svm import SVC
```

Os argumentos utilizados na função irão depender do kernel escolhido. Considere, por exemplo, que seja utilizado um kernel polinomial, neste caso, a função pode ser utilizada como:

```
model = SVC(loss = "poly", C = b, random_state = c, gamma =  $\gamma$ , coef0 =  $r$ , degree =  $d$ )
```

onde  $b$  é o parâmetro de regularização,  $c$  é um inteiro aleatório (manter este valor constante) e os parâmetros  $\gamma$ ,  $r$  e  $d$  são os parâmetros presentes na definição do kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + r)^d$$

a) Mantenha  $\gamma = 1$  e utilize os valores “padrão” para  $r$  e  $C$  ( $C = 1$ ,  $r = 0$ ). Compare o desempenho do algoritmo para valores *pares* e *ímpares* de  $d$  (considere valores inteiros entre 1 e 10). Discuta este comportamento com base no funcionamento do algoritmo e nas características dos dados gerados.

b) Mantenha agora os valores  $C = 1$ ,  $r = 0$  e  $d = 2$ . Avalie o desempenho do algoritmo para valores menores de  $\gamma$ , na faixa de 0.05 – 0.5. Como isto afeta o resultado? Por que?

**Obs.:** Este trabalho deverá ser entregue em um arquivo .pdf contendo as respostas de cada item.