

Máquinas de Vetores-Suporte (SVMs)

Parte II

1. Margem Suave (*Soft Margin*)

- Embora a formulação apresentada na primeira parte resolva bem o problema de obter o classificador linear de máxima margem, vigora a hipótese de que os dados são linearmente separáveis. O uso de SVM seria por demais limitado caso essa hipótese fosse incontornável.
- Num primeiro momento, consideremos que se deseje manter a estrutura de separação (classificador linear), mas com certa **tolerância a violações da hipótese de separabilidade**.

- Quantificaremos essa tolerância com a ajuda de um conjunto de variáveis (variáveis de relaxação, *slack variables*) não-negativas $\xi_i \geq 0$, $i = 1, \dots, N$. De posse dessas variáveis, podemos reescrever as restrições que usamos na derivação do hiperplano ótimo da seguinte maneira:

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, \dots, N$$

Isso significa que ocorre separação linear sob certa tolerância.

- Busca-se “minimizar o nível exigido de tolerância”, o que justifica a introdução das variáveis ξ_i num termo de penalização do custo a ser minimizado. Esse termo levará em conta a soma das variáveis (no caso de um problema de “norma 1”) ou de seus quadrados (no caso de um problema de “norma 2”) (CAMPBELL, 2002). O caso de norma-1 poderia ser formulado assim:

$$\begin{cases} \min_{\mathbf{w}, b} \Phi = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{k=1}^N \xi_k \\ \text{s. a. } (\mathbf{w}^T \mathbf{x}_i + b) d_i \geq 1 - \xi_i, i = 1, \dots, N \\ \xi_i \geq 0, i = 1, \dots, N \end{cases}$$

- Já o caso de norma-2 seria:

$$\begin{cases} \min_{\mathbf{w}, b} \Phi = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{k=1}^N \xi_k^2 \\ \text{s. a. } (\mathbf{w}^T \mathbf{x}_i + b) d_i \geq 1 - \xi_i, i = 1, \dots, N \\ \xi_i \geq 0, i = 1, \dots, N \end{cases}$$

- A escolha da norma (1 ou 2) afeta a formulação/solução do problema. Outro ponto que se costuma ressaltar é que o uso da norma-1 pode limitar de maneira mais drástica a contribuição dos *outliers*.
- A constante C é um hiperparâmetro a ser escolhido com muito cuidado. Tipicamente, faz-se uma busca sobre um conjunto de valores tendo por base o desempenho num conjunto de validação (e.g. validação cruzada).

2. Referências bibliográficas

BISHOP, C., *Pattern Recognition and Machine Learning*, Springer, 2006.

CAMPBELL, C., “Kernel Methods: a Survey of Current Techniques”, *Neurocomputing*, Vol. 48, pp. 63 – 84, 2002.

CORTES, C., VAPNIK, V., “Support Vector Networks”, *Machine Learning*, Vol. 20, pp. 273 – 297, 1995.