I ENCONTRO DE PÓS-GRADUANDOS EM MATEMÁTICA NA UFSC

UNIVERSIDADE FEDERAL DE SANTA CATARINA

FUNDAMENTOS DE APRENDIZADO DE MÁQUINA

Análise Funcional voltada a Métodos de Kernel

ALEK FRÖHLICH[1], **DANILO ROYER**[2], **MAURO ROISENBERG**[3]

Universidade Federal de Santa Catarina, SC, Brasil

alek.frohlich@posgrad.ufsc.br[1], danilo.royer@ufsc.br[2], mauro.roisenberg@ufsc.br[3]

Introdução

A troca de representação de conjuntos de dados é uma técnica utilizada no Aprendizado de Máquina. Para alguns algoritmos conhecidos como Métodos de Kernel, há disponível uma técnica de troca de representação poderosa: o Truque do Kernel. Para estes algoritmos, a única informação relevante do conjunto de dados é o valor do produto interno entre pares de vetores de treino. Isto é, a quantidade: $\langle x_i, x_j \rangle$. Por consequência, qualquer transformação $\phi : X \to H$ aplicada sobre o conjunto de dados só aparecerá em expressões da forma: $\langle \phi(x_i), \phi(x_j) \rangle$. É um fato interessante que conseguimos encontrar funções da forma $K(x, y) = \langle \phi(x), \phi(y) \rangle$, onde ϕ e H estão implícitos, que são úteis para a resolução de problemas de Aprendizado de Máquina. Um exemplo é o Kernel Gaussiano, que é altamente usado em conexão com algoritmos como Kernel-SVM e Kernel-PCA:

$$K(x, y) = e^{-\gamma \|x-y\|^2}. \tag{1}$$

É possível se dizer que a relação de kernels com Aprendizado de Máquina começou com a descoberta de que o Perceptron não era capaz de classificar conjuntos de dados não linearmente separáveis [1]. De fato, o Kernel-Perceptron foi o primeiro algoritmo de classificação a ser adaptado para o uso de kernels. Neste caso, o Truque do Kernel consiste em mapear o conjuto de dados para um espaço de dimensão maior, de forma que se torne linearmente separável. Veja a Figura 1. Dentre os vários usuários de kernels, o algoritmo mais proeminente é o Kernel-SVM, introduzido por Vapnik na década de 90 [2].

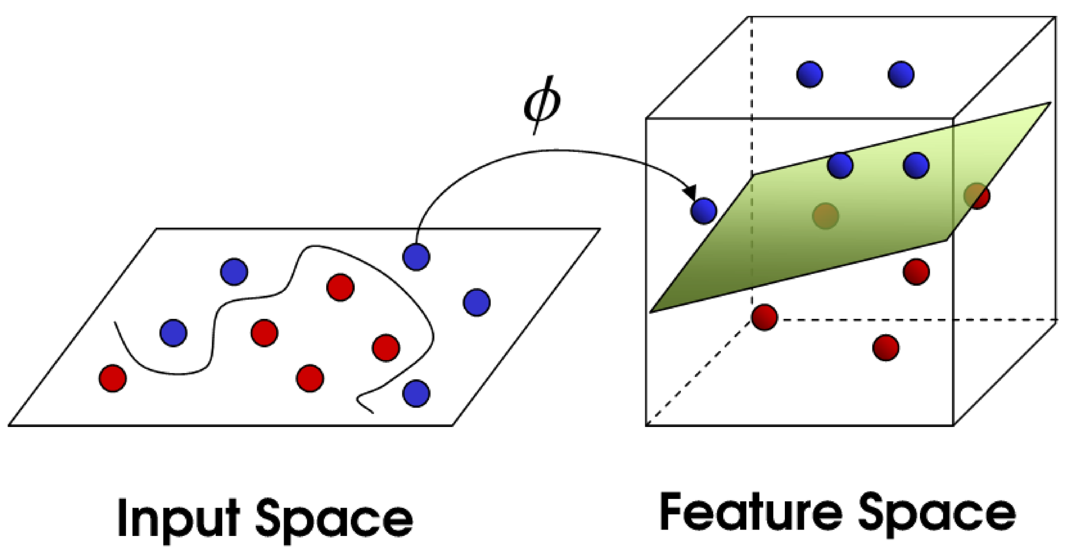

Figura 1: ϕ torna o conjunto de dados linearmente separável. Fonte: VOCATURO et al., 2019.

Objetivos

O principal objetivo deste trabalho é esclarecer as bases teóricas em que estão fundamentados os Métodos de Kernel. Em particular, almejamos encontrar uma caracterização mais simples de kernel reprodutor que nos permita verificar que os principais kernels da literatura de Métodos de Kernel de fato são kernels reprodutores. Ademais, aproveitamos para justificar que o Perceptron é um Método de Kernel; isto é, que depende apenas do valor do produto interno $\langle x_i, x_j \rangle$ entre os vetores de treino. Por fim, nossa discussão permite formular o Teorema do Representante, uma ferramenta teórica importante na demonstração da corretude de outros Métodos de Kernel mais complexos que o Perceptron.

Fundamentação Teórica

Este trabalho se baseou em [3] para os elementos de Espaços de Hilbert e em [5] para Métodos de Kernel. A discussão a respeito do Teorema do Representante pode ser encontrada na versão íntegra do TCC [6] ou, alternativamente, em [4].

Desenvolvimento e Metodologia

Para alcançar esses objetivos, introduzimos elementos da teoria dos Espaços de Hilbert e apresentamos o algoritmo Perceptron. Da Análise Funcional, os resultados mais relevantes abrangem o completamento de um espaço com produto interno, noções de convergência em espaços de funções e a generalização da desigualdade de Cauchy-Schwarz para formas bilineares positivo semidefinidas. Por fim, apresentamos alguns resultados envolvendo funções simétricas e positivo definidas.

Conclusões

O trabalho se propos a esclarecer as bases teóricas dos Métodos de Kernel. Na direção de uma caracterização mais simples para kernels reprodutores, alcançamos o Teorema 1. A respeito da corretude do Kernel-Perceptron, conseguimos obter o Teorema 2. Uma possível continuação deste trabalho consistiria em justificar a corretude de outros Métodos de Kernel, como o K-SVM e o K-PCA.

Teorema 1. ***(Moore-Aronszajn)*** *Seja $K : X \times X \to \mathbb{R}$ uma função simétrica e positivo definida sobre um conjunto X, então existe um único Espaço de Hilbert de Reprodução H_K cujo kernel reprodutor associado é K.*

Teorema 2. ***(Corretude do K-Perceptron)*** *Seja $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ um conjunto de treino linearmente separável pela origem tal que haja w^* que corretamente o classifique com margem γ. Sob estas condições, o número de erros t que o algoritmo comete antes de terminar é limitado superiormente da seguinte forma:*

$$t \leq \left(\frac{R}{\gamma}\right)^2, \tag{2}$$

em que $R := \max_{1 \leq i \leq n} \|x_i\|$.

Referências

[1] M. Minsky. Perceptrons; an introduction to computational geometry. *MIT Press*, Cambridge, 1969.

[2] B. E. Boser, I. M. Guyon, and V. Vapnik. A Training algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, Pennsylvania, 1992.

[3] E. Kreyszig. Introductory Functional Analysis with Applications. *Wiley Classics Library*, 1989.

[4] B. Schlkopf. Learning with Kernels: Support Vector Machines, regularization, optimization, and beyond. *MIT Press*, 2018.

[5] N. Cristianini. An introduction to Support Vector Machines and other kernel-based learning methods. *Cambridge University Press*, 2000.

[6] A. Frohlich. Fundamentos de Aprendizado de Máquina: Análise Funcional voltada a Métodos de Kernel. Trabalho de Conclusão de Curso em Ciência da Computação, *UFSC*, 2022.