



Support Vector Machine

*E-mail: pima.vn@gmail.com

Mô tả Dự án

Support Vector Machines (**SVMs**) là những mô hình học có giám sát được phát minh bởi **Vladimir Naumovich Vapnik** và **Alexey Yakovlevich Chervonenkis** năm **1963**. Mô hình này được dùng rất nhiều trong bài toán phân lớp dữ liệu (được áp dụng trong cả bài toán **phân loại** lẫn **hồi quy**). Ở phiên bản gốc, **SVMs** là một mô hình **nhị phân** gồm một hệ điểm dữ liệu (được gán nhãn 1 hoặc -1) biểu diễn trên một không gian nhiều chiều, **SVMs** hỗ trợ xây dựng một **siêu phẳng** để phân hệ dữ liệu trên thành hai lớp (mỗi lớp tương ứng với một nhãn). Không chỉ dừng lại ở bài toán phân lớp, **SVMs** yêu cầu các điểm dữ liệu nằm càng xa **siêu phẳng** càng tốt.

Yêu cầu

Hãy trình bày mô hình SVM thỏa mãn các yêu cầu sau.

- (1) Phát biểu bài toán phân loại nhị phân theo ngôn ngữ toán học.
- (2) Xây dựng hàm mục tiêu cần tối ưu và điều kiện ràng buộc.
- (3) Trong phần lớn các trường hợp, người ta giải bài toán đối ngẫu thay vì bài toán gốc. Hãy trình bày ý tưởng này và giải thích lý do.
- (4) Áp dụng các mô hình vào một dữ liệu cụ thể. Nhận xét về kết quả và đánh giá mô hình.
- (5) Khi dữ liệu không thể phân cách tuyến tính ngay được hoặc có nhiều thông tin nhiễu, hãy tìm hiểu và trình bày các phương pháp bổ trợ cho SVM trong các trường hợp.
- (6) Trong quá trình trình bày, chỉ rõ ra được ý nghĩa hình học của các thông số, hàm số trong bài làm và đưa ra ví dụ (ở các trường hợp đơn giản) để giải thích các bước thực hiện.

Kiến thức: Ma trận, các phép toán cơ bản trên ma trận, không gian \mathbb{R}^n , vi tích phân nhiều biến, bài toán tìm cực trị, phương pháp nhân tử Lagrange.

Một số từ khóa: SVMs, Supervised Learning, Classification, Regression, Hyperplane, Optimization, Convex Optimization, Linearly Seperable, Linear Classifier, Maximum Margin, Dual Problem.

Tham Khảo

- [1] Các bài giảng PiMA 2018.
- [2] https://en.wikipedia.org/wiki/Support_vector_machine