MỘT SỐ GHI CHÚ VỀ SVM

Các ghi chú bên dưới đi kèm theo slide “SVM\_update.pdf” vốn là bản rút ngắn của tập tin “SVM\_for\_Classification.pdf”.

[#]: slide thứ # trong “SVM\_update.pdf”.

[7] Mục đích chính của thuật toán SVM là phân lớp các input thành 2 tập với nhãn y là 1 và -1.

* Vấn đề phân lớp nhị phân

Để phân lớp thì SVM sẽ đi tìm một siêu phẳng (hyperplane) để chia tập input thành 2 miền đúng với nhãn của nó.

Biên/lề (margin) là khoảng cách ngắn nhất giữa điểm dữ liệu gần siêu phẳng nhất đến siêu phẳng. SVM sẽ đi tìm siêu phẳng có lề cực đại.

Nhiều điểm thuộc margin thì bài toán phụ thuộc vào các điểm đó và đường siêu phẳng có khả năng gây ra nhiễu.

[9] Độ rộng lề: d = |d1 - d2| = 2 / ||w||

w là vectơ pháp tuyến.

Định nghĩa về nhân tử Lagrange (Lagrange multiplier) từ wiki:

*In mathematical optimization, the method of Lagrange multipliers (named after Joseph Louis Lagrange[1]) is a strategy for finding the local maxima and minima of a function subject to equality constraints.*

Có thể nói ngắn gọn là nhân tử Lagrange được dùng để tìm các điểm cực đại và cực tiểu của một hàm theo các ràng buộc cho trước.

Từ bài toán tìm siêu phẳng lề cực đại (tìm max), biến đổi thành bài toán tìm cực tiểu, từ đó dẫn đến việc phải dùng đến công thức Lagrange

[14] điều kiện KKT: Karush-Kuhn-Tucker

[16] Lề mềm (soft margin): bản chất cũng đi tìm lề tối đa nhưng cho phép một số ít nhãn gán sai.

[17] C: misclassification penalty, : càng lớn thì số mẫu bị sai càng lớn

[25] Kernel: biến đổi từ không gian n chiều sang không gina (n+1) chiều. Tuy có thể phân lớp hiệu quả hơn nhưng sẽ bị Curse-of-dimension.

Curse-of-dimension: khi số chiều tăng thì bản chất dữ liệu không còn giữ nguyên => overfitting

[29] Hàm kernel trình bày trong slide và wiki là dạng tuyến tính (linear).

[29] Hàm kernel trong paper LASVM là dạng Gaussian.

Tool cho SVM: LibSVM

Tối ưu lồi: Convex optimization - Stephen Boyd

Học thống kê: Statistical Learning Theory: Vladimir N. Vapnik