1. Nêu khái niệm và ý tưởng thực hiện của Support Vector Machine (SVM) trong classification.

SVM là một thuật toán rất phổ biến trong Machine Learning. Có thể linh hoạt thực hiện nhiều nhiệm vụ như: classification, regression, clustering và outliar detection.

Về ý tưởng, SVM sẻ có gắn tìm ta một decision boundaries mà khoảng cách từ nó tới sample gần nhất của cả hai classes là lớn nhất.

1. Trình bày các phương pháp classification cho nonlinear data sử dụng SVM.

Phương pháp 1: Thêm các feature bậc cao. Khi thêm các feature này ta sẻ học ra được đường decision boundaries có bậc cao hơn thay vì chỉ là một đường thẳng, từ đó có thể thực hiện classification tốt hơn.

Phương pháp 2: Thêm similarity feature. Similarity feature là khoảng cách tính bằng Gausian kernel function tới landmark. Việc thêm này giúp biến dử liệu không gian 1 chiều thành 2 hay nhiều chiều. Từ đó ta có được đường decision boundaries phức tạp hơn mà không mất thời gian như phương pháp 1.