Phân tích dữ liệu thông minh

Support Vector Machine

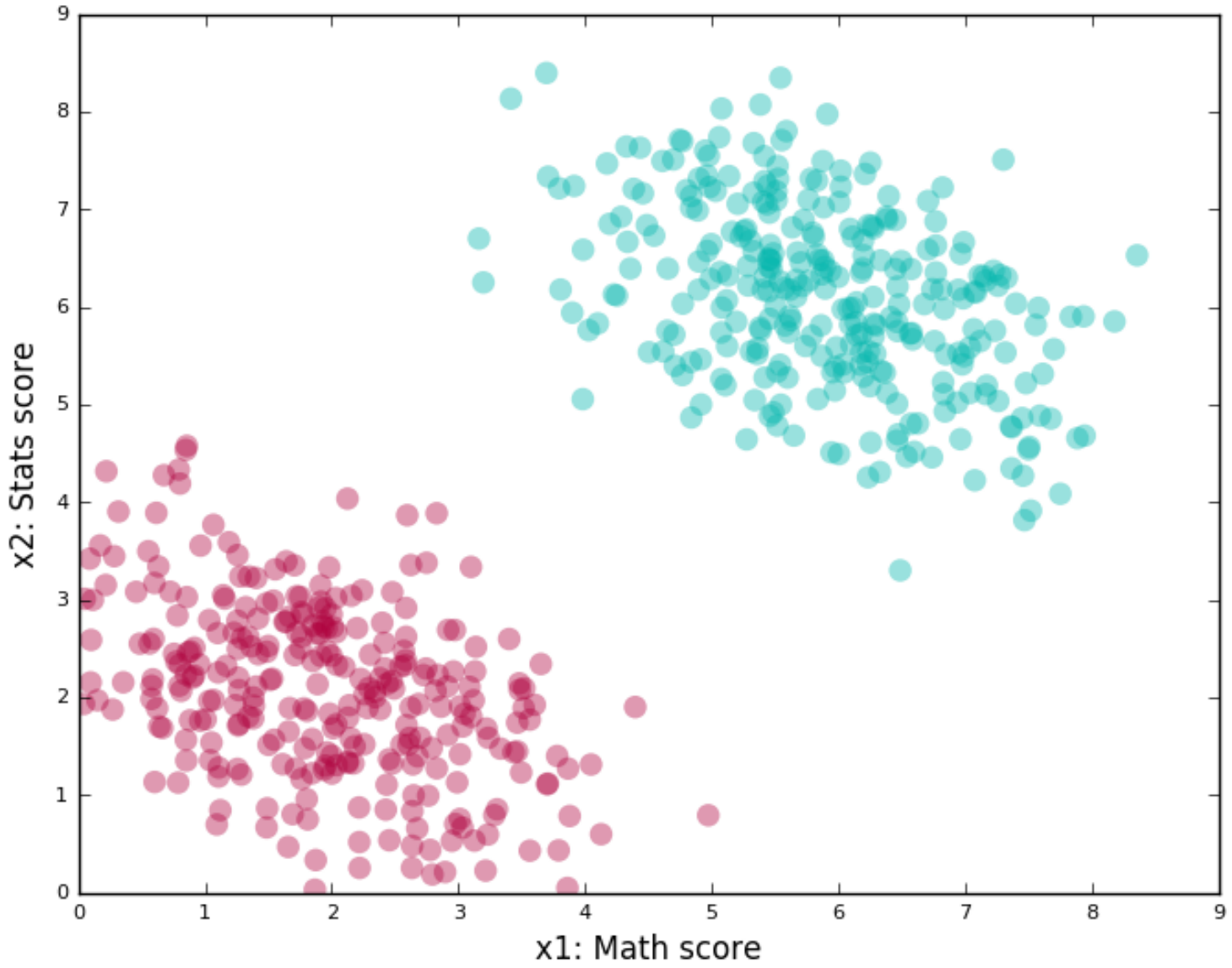
TS. Nguyễn Tiến Huy

ntienhuy@fit.hcmus.edu.vn

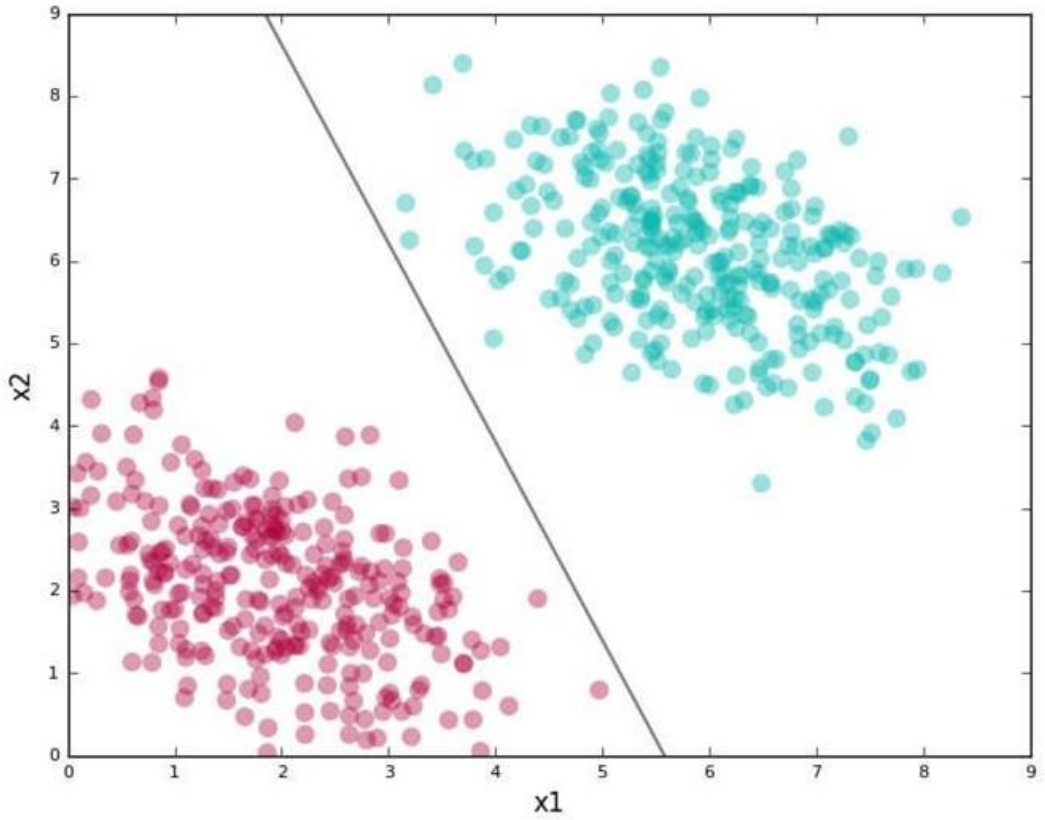
Content

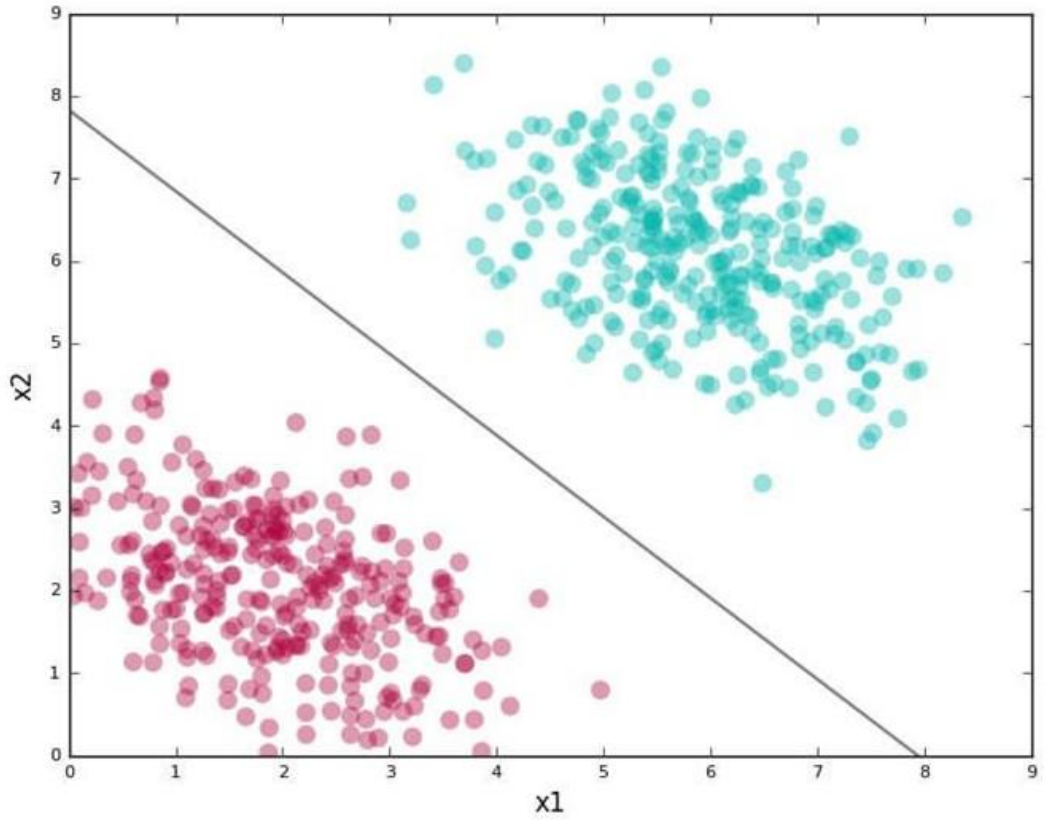
1. Problem of Classification
2. SVM
3. Kernel Trick
4. Multiclass classification

The Problem of Classification

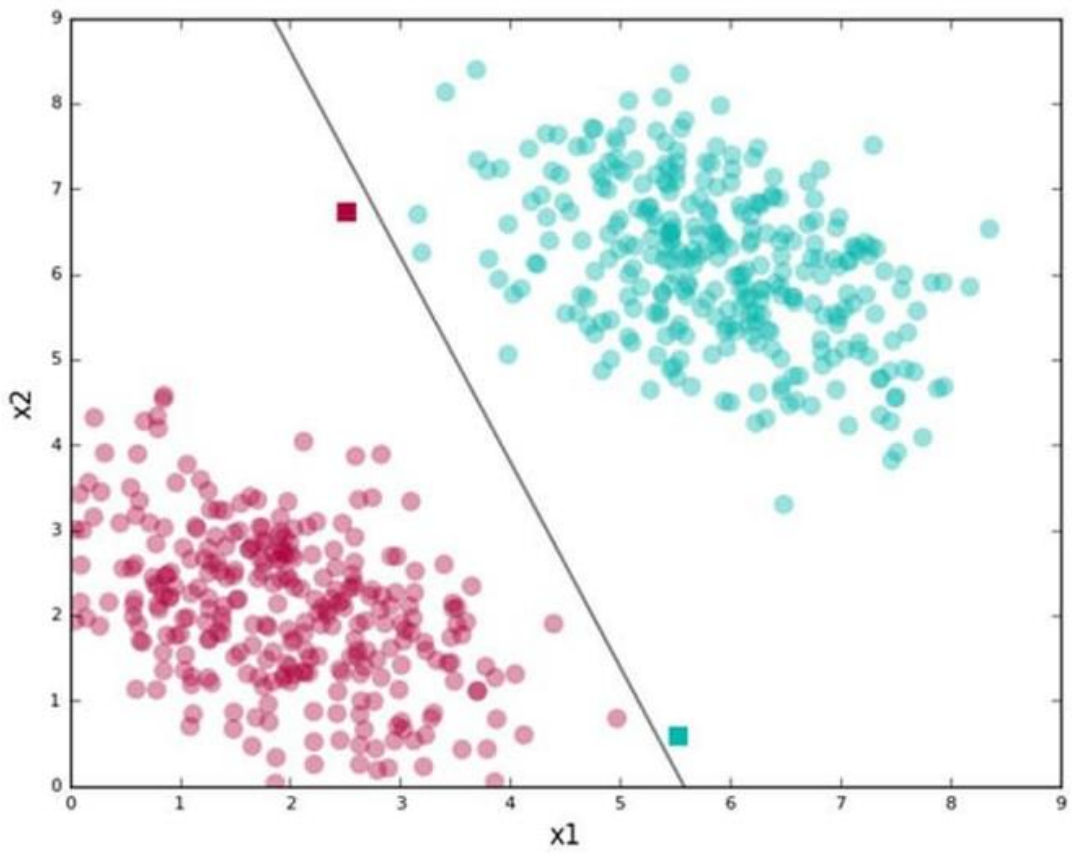


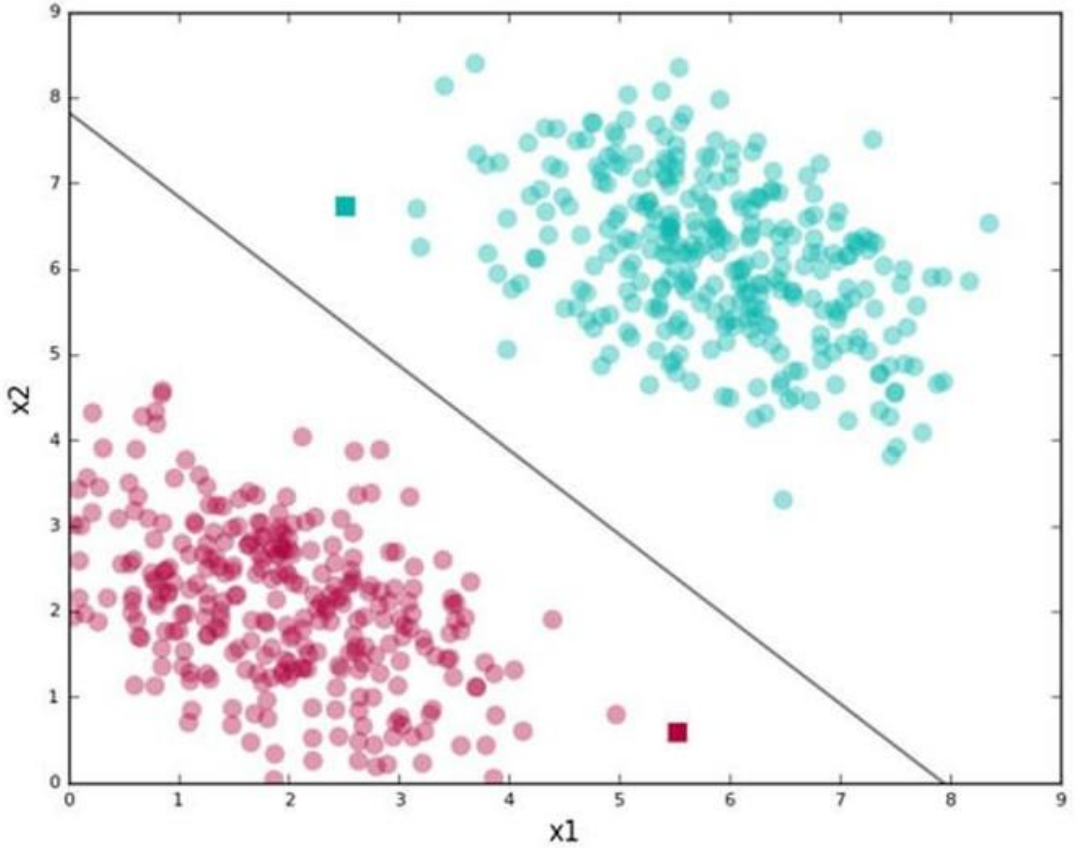
The Problem of Classification





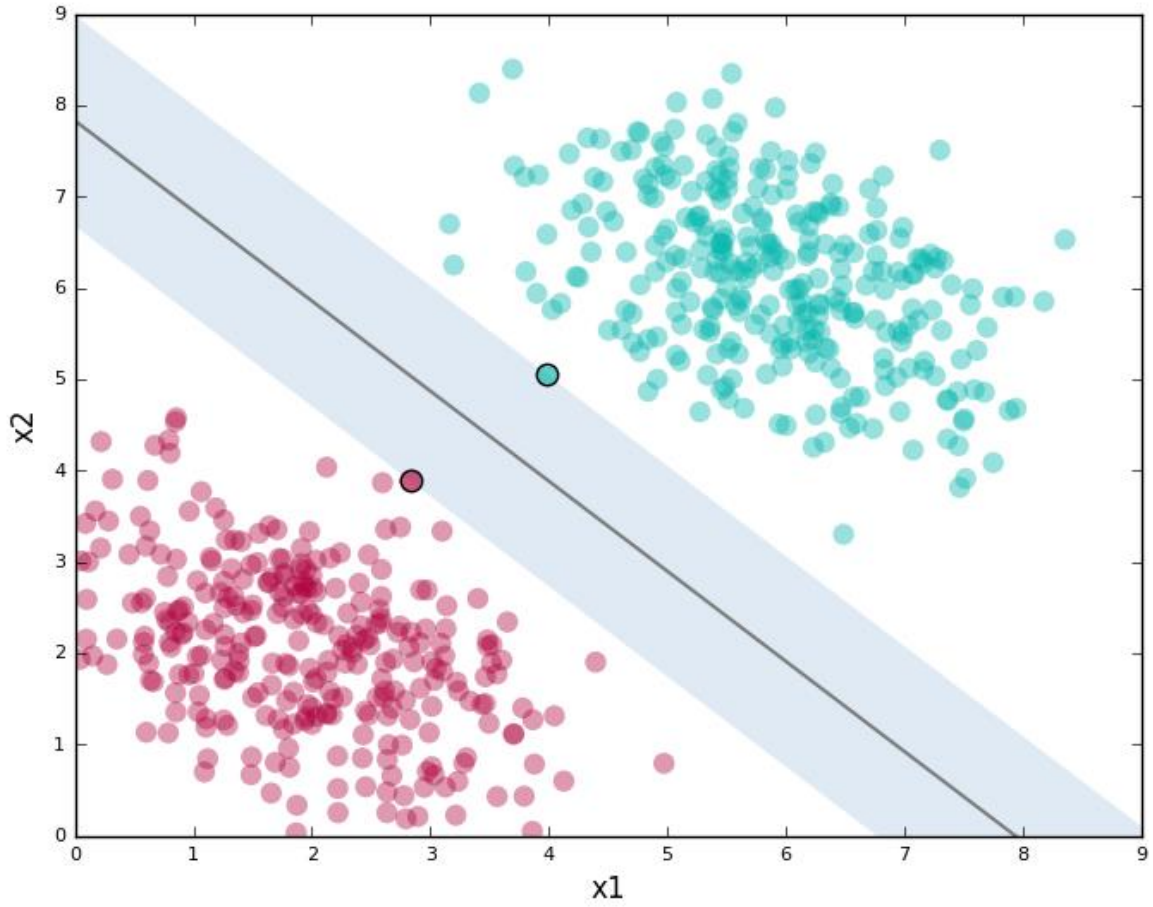
The Problem of Classification



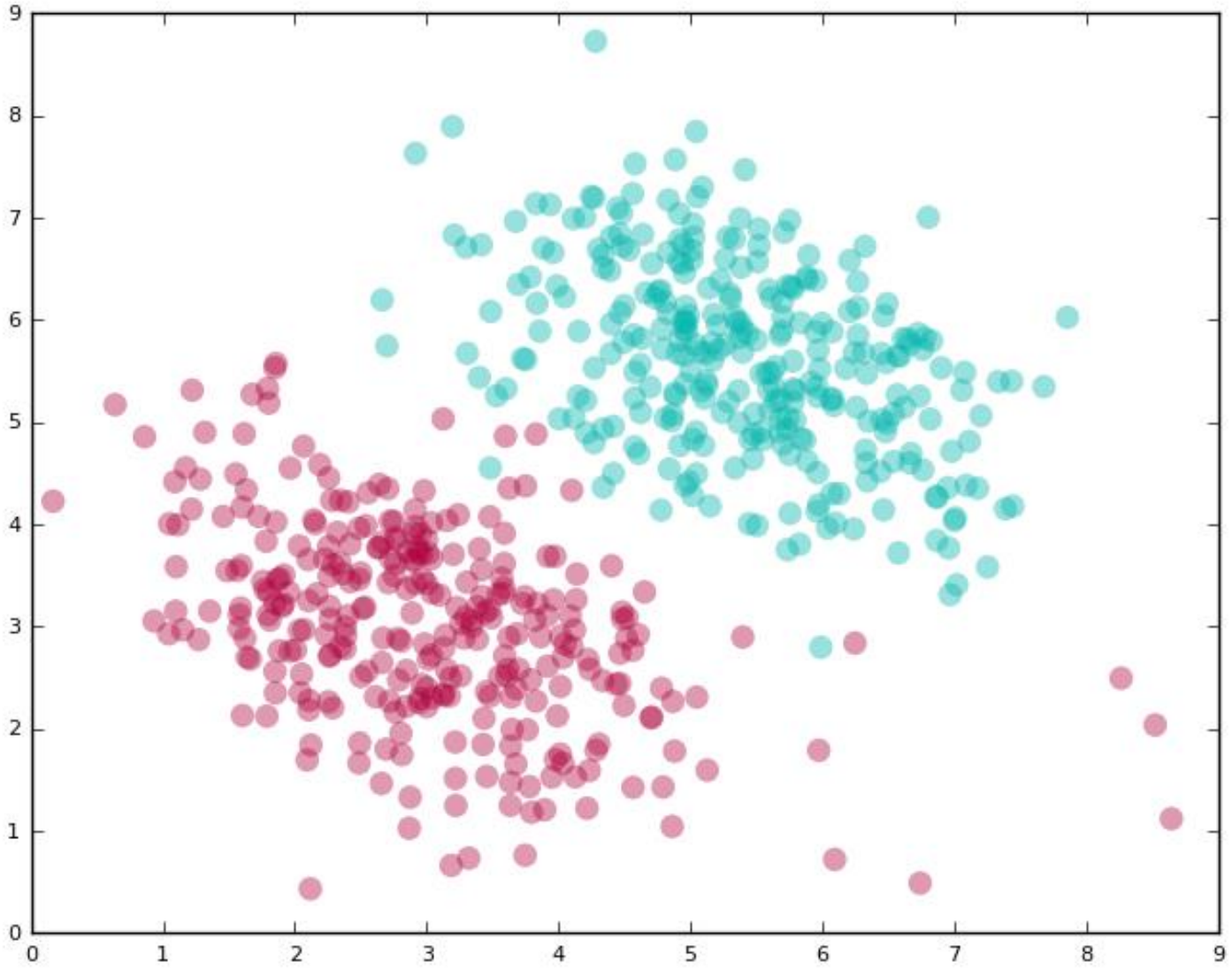


SVM

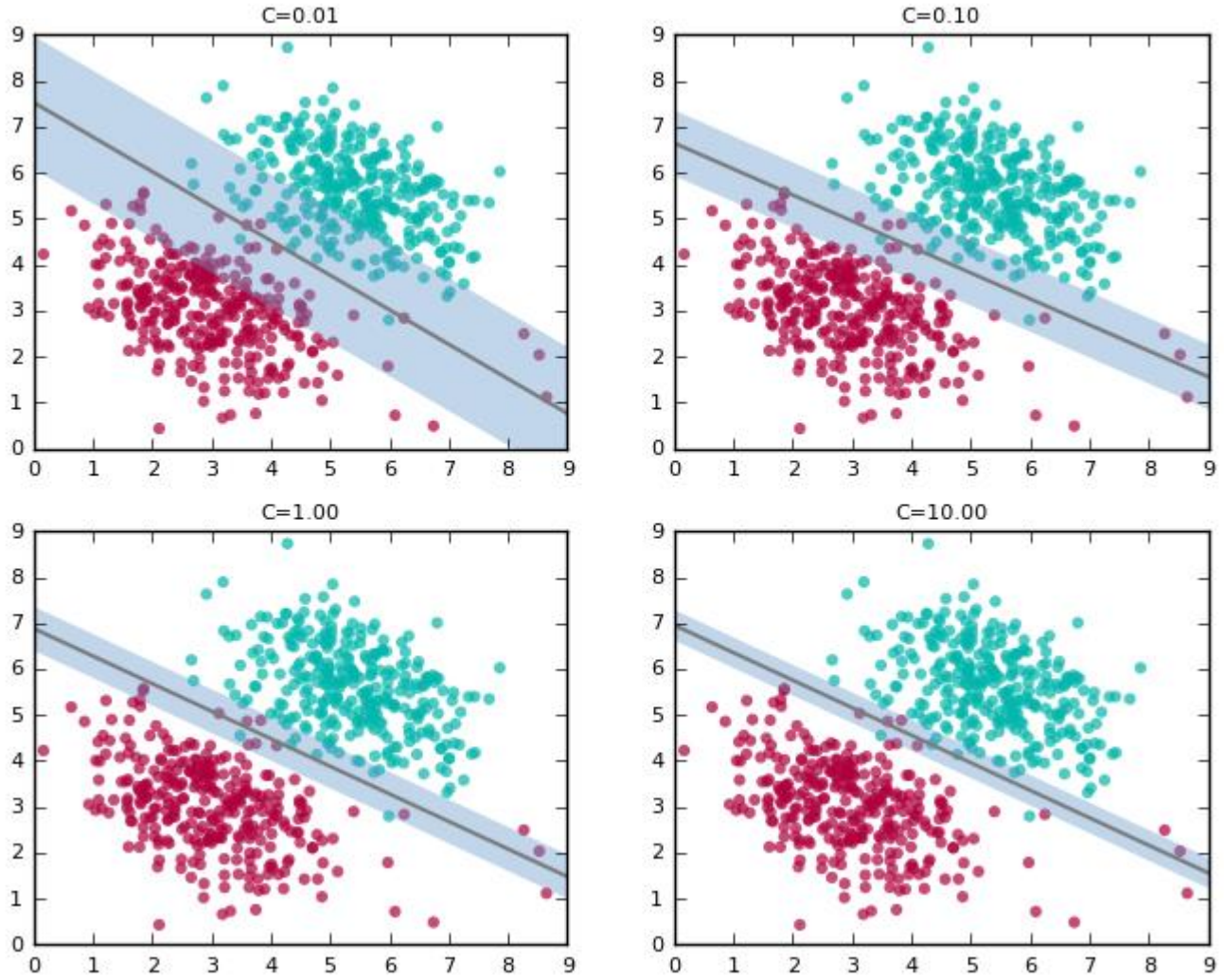
- Find lines that correctly classify the training data
- Among all such lines, pick the one that has the greatest distance to the points closest to it.



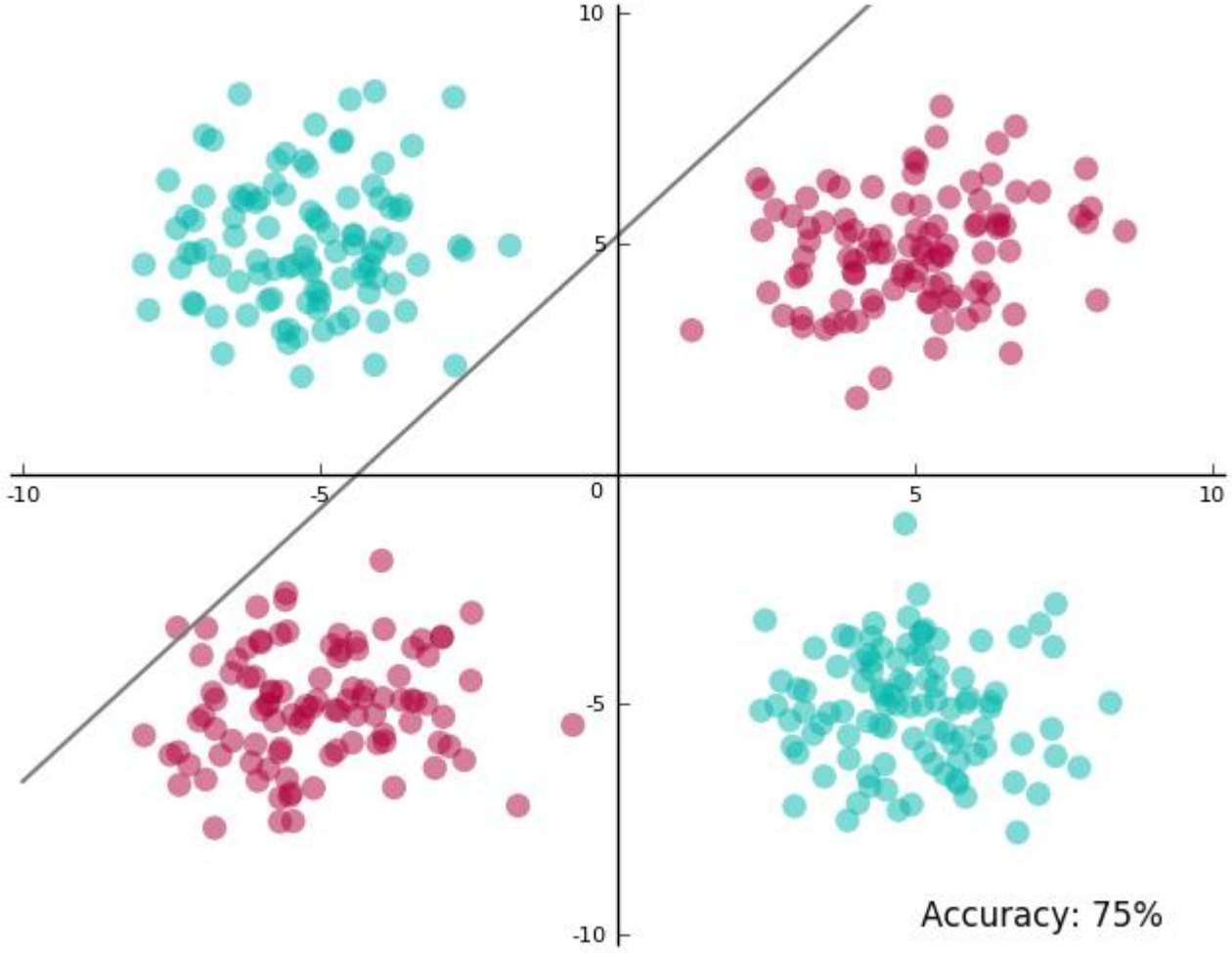
How about errors?



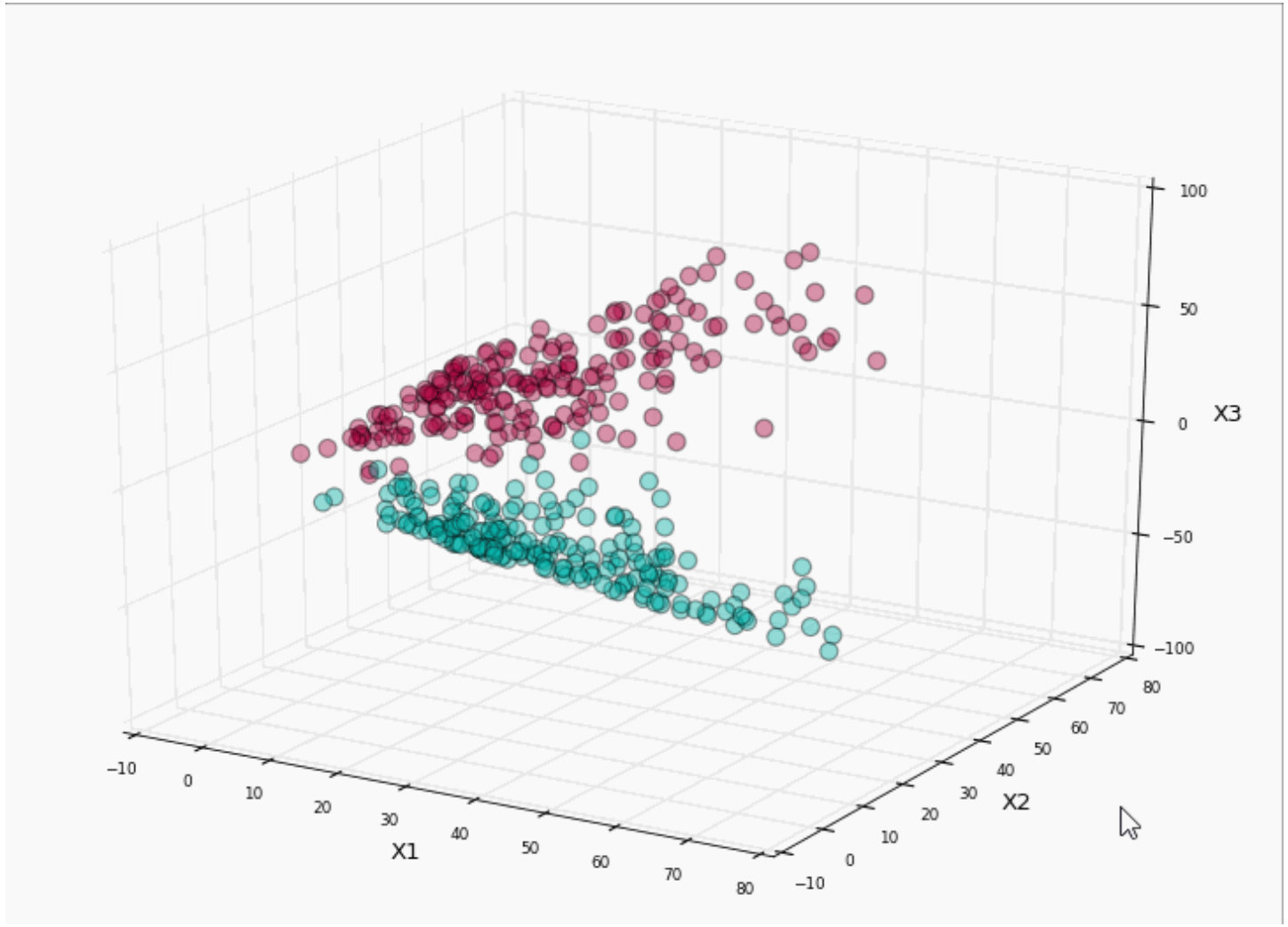
C parameter in SVM



Non-linearly Separable Data

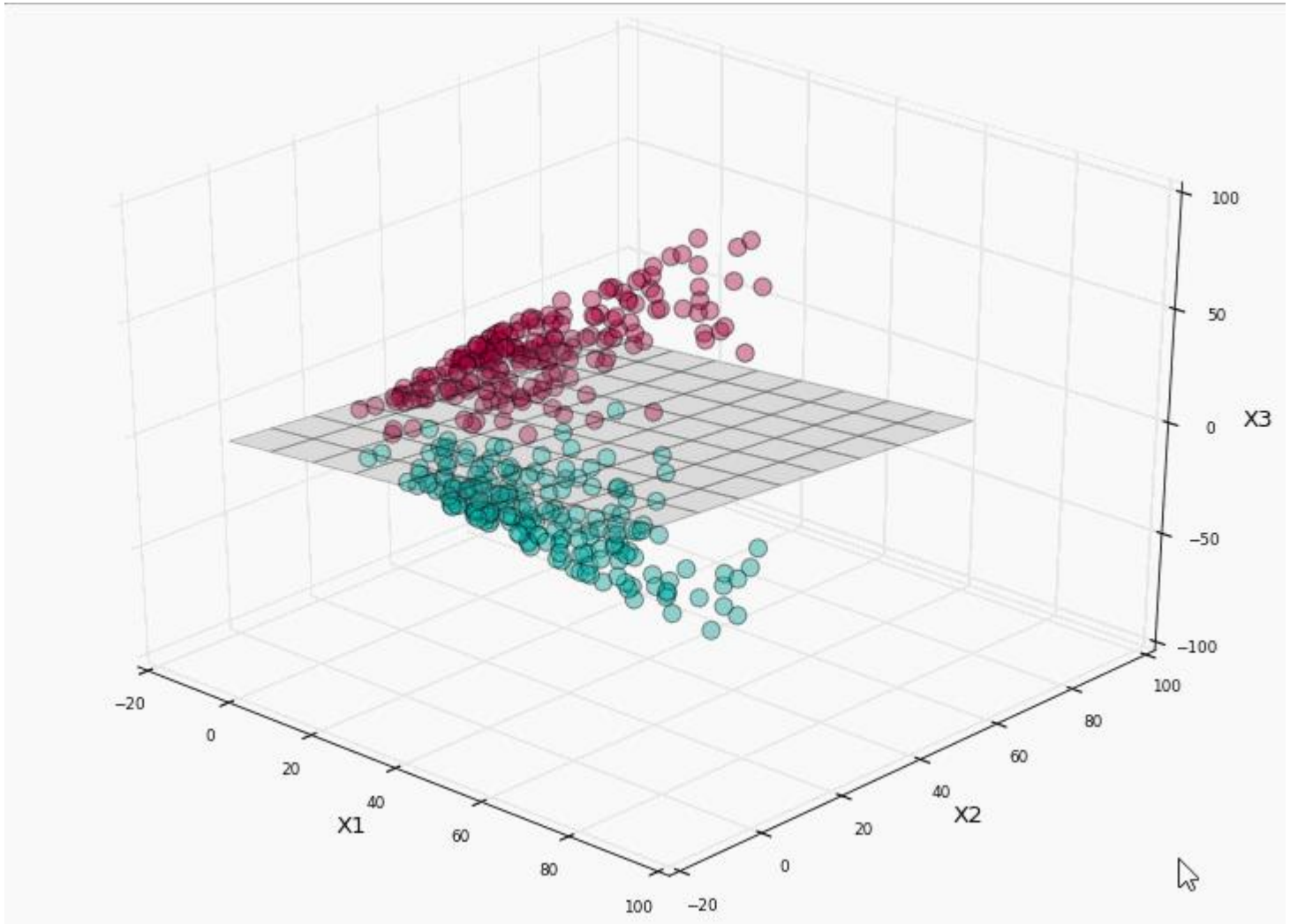


Non-linearly Separable Data

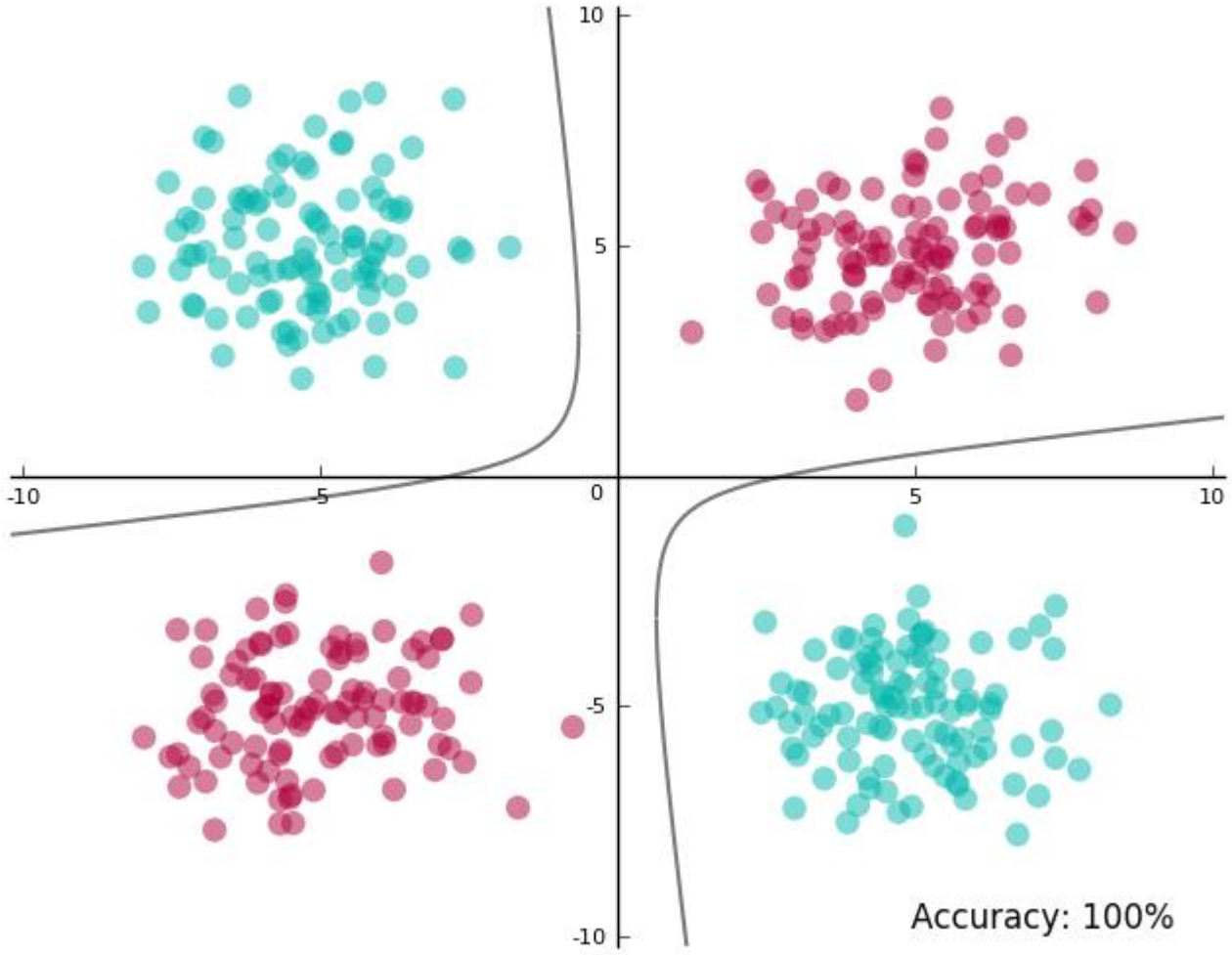


$$X_1 = x_1^2$$
$$X_2 = x_2^2$$
$$X_3 = \sqrt{2}x_1x_2$$

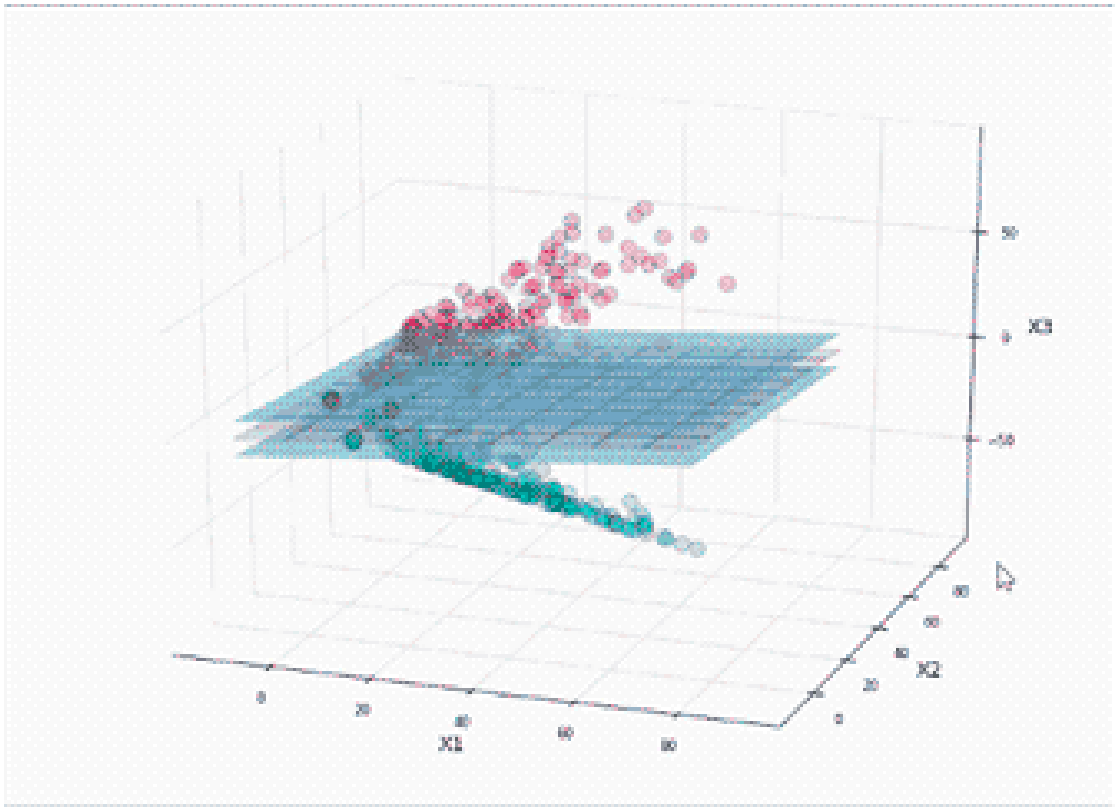
Non-linearly Separable Data

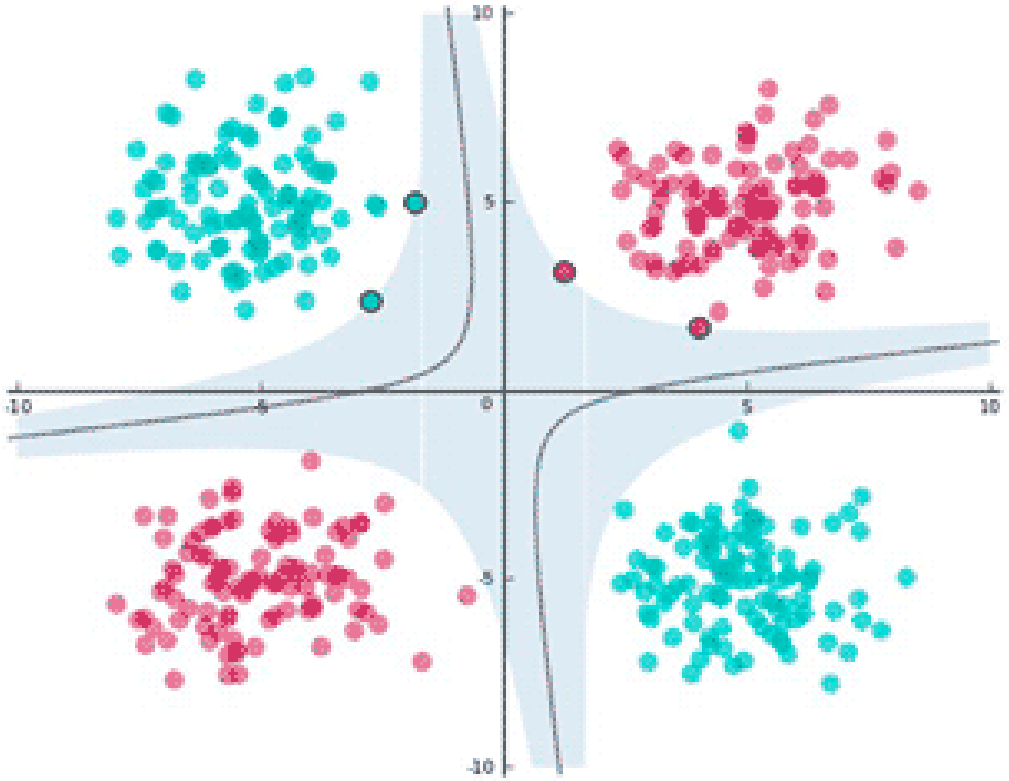


Non-linearly Separable Data



Non-linearly Separable Data





Multiclass classification

- One vs. One
- One vs. Rest

Example

For example consider four class problem having classes A, B, C, and D.

- One vs. Rest
  - Models classifiers_A, classifier_B, classifier_C and classifier_D
  - During prediction here is the probability we get:
    - classifier_A = 40%
    - classifier_B = 30%
    - classifier_C = 60%
    - classifier_D = 50%
  - We assign it class C

Example

- One vs One
  - We train total six classifier with subset of data containing classes involved
    - classifier_AB
    - classifier_AC
    - classifier_AD
    - classifier_BC
    - classifier_BD
    - classifier_CD
  - And during classification
    - classifier_AB assigns class A
    - classifier_AC assigns class A
    - classifier_AD assigns class A
    - classifier_BC assigns class B
    - classifier_BD assigns class D
    - classifier_CD assigns class C
  - We assign it to class A

Conclusion

- SVM for classification
- Kernel tricks
- Multiclass classification

Cảm ơn đã theo dõi!