



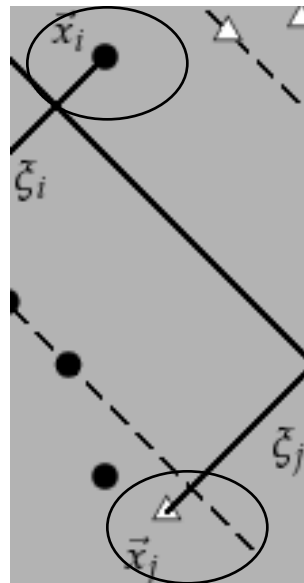
SOFT-MARGIN SVMs

SYRACUSE UNIVERSITY
School of Information Studies

SOFT-MARGIN SVMs

No perfect linear boundary can be found between the two classes due to outliers.

Introduce a slack variable ξ to pay a cost for each misclassified example.



REGULARIZATION IN C-SVC

Tune the regularization parameter C (cost for misclassification).

Default value: $C = 1$

When C is large (high cost), the algorithm tries to build model with fewest training errors, resulting in narrow margin and high chance of overfitting.

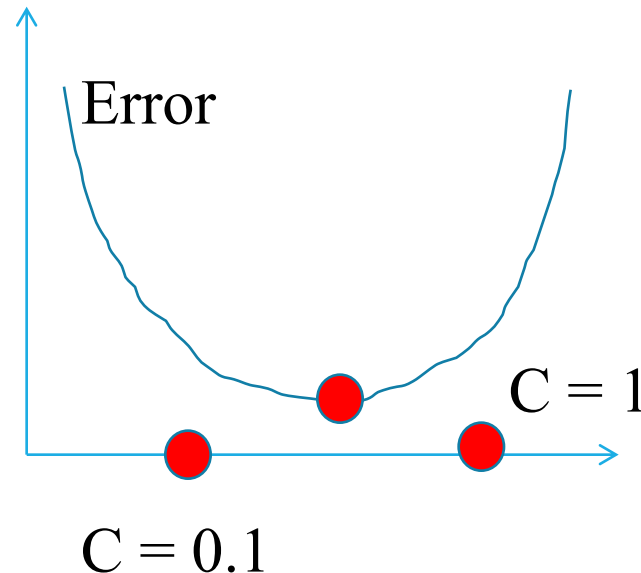
When C is small (low cost), the margin is wider, more robust.

However, C cannot be too small, or else it does not respect the data at all.

REGULARIZATION

Use manual tuning or gradient descent search to find the best C .

E.g., set C 's search range from 0.1 to 1.0 and increase with step size 0.05.



A VISUALIZATION FROM COURSERA

<https://class.coursera.org/ml-003/lecture/72>

07:13–9:00