Foundations of Machine Learning

Module 5: Support Vector Machine
Part D: SVM – Maximum Margin with Noise

Sudeshna Sarkar
IIT Kharagpur

Linear SVM formulation

Find **w** and b such that

$$\frac{2}{||w||} \text{ is maximized}$$

And for each of the m training points (x_i, y_i),

$$y_i(w.x_i + b) \geq 1$$

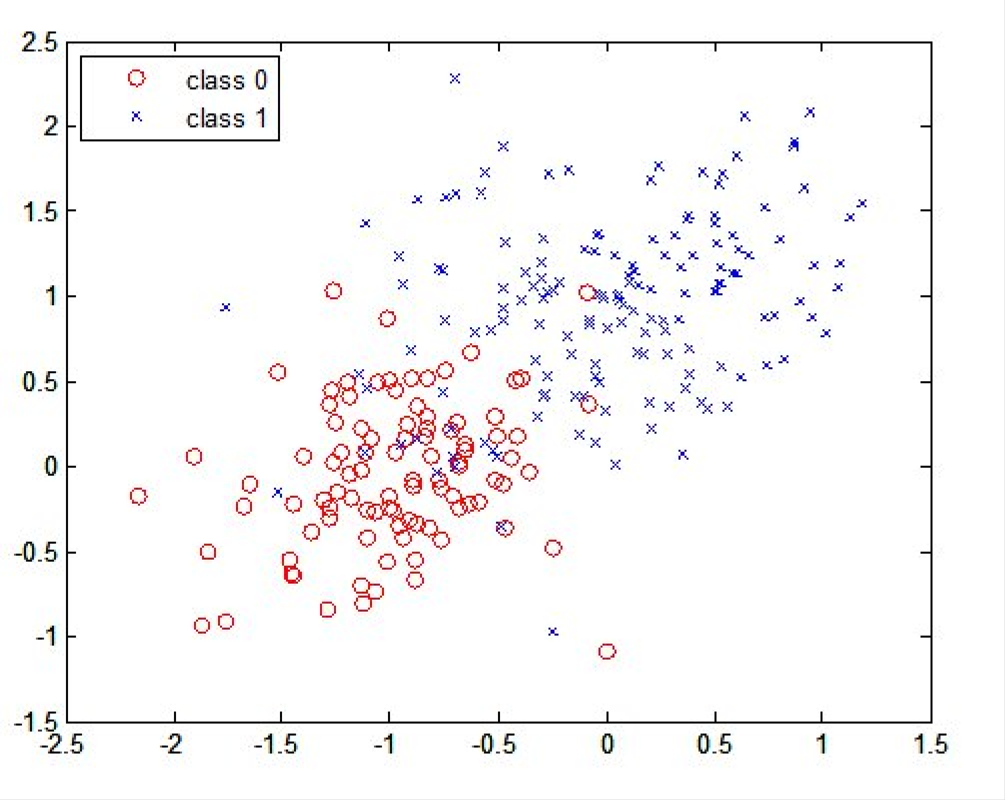
Find **w** and b such that

$$||w||^2 = w.w \text{ is minimized}$$

And for each of the m training points (x_i, y_i),

$$y_i(w.x_i + b) \geq 1$$

Limitations of previous SVM formulation



- What if the data is not linearly separable?
- Or noisy data points?

Extend the definition of maximum margin to allow non-separating planes.

How to formulate?

- Minimize $\|w\|^2 = w.w$ and *number of misclassifications*, i.e., minimize

$$w.w + \#(\text{training errors})$$

- No longer QP formulation

Objective to be minimized

- Minimize

$w.w + C\,(\text{distance of error points to their correct zones})$

- Add slack variables ξ_i

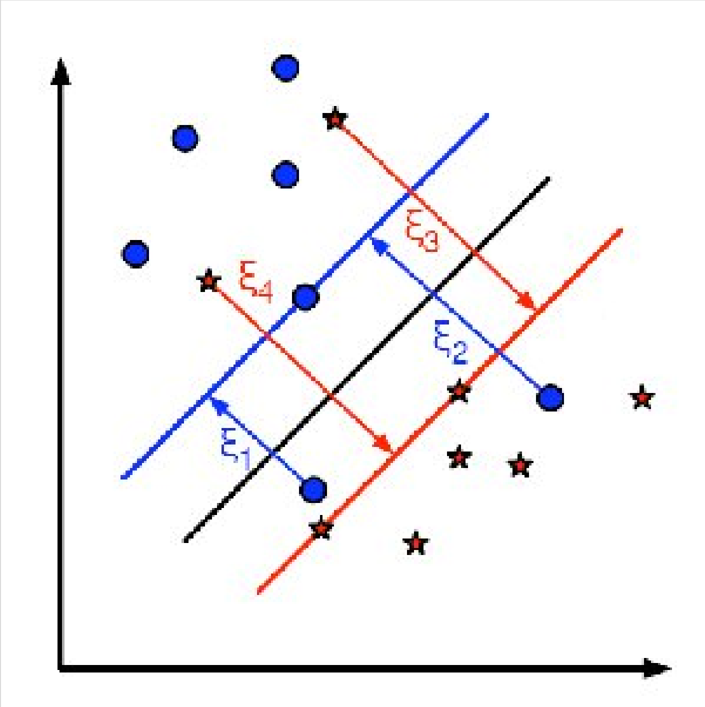
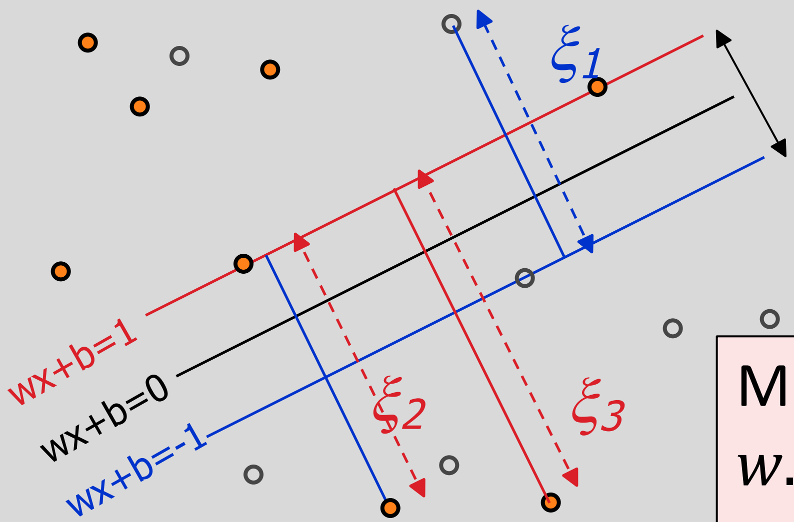


Figure from http://docs.opencv.org/2.4/doc/tutorials/ml/non_linear_svms/

Maximum Margin with Noise



$$\mathrm{M} = \frac{2}{\sqrt{w.w}}$$

Minimize

$$w.w + C\sum_{k=1}^{m} \xi_k$$

m constraints

$$w.x_k + b \geq 1 - \xi_k \text{ if } y_k = 1$$

$$w.x_k + b \leq -1 + \xi_k \text{ if } y_k = -1$$

$$\equiv$$

$$\left\{ \boldsymbol{y_k(w.x_k + b) \geq 1 - \xi_k}, \quad \text{k=1,...,m} \right\}$$

$$\boldsymbol{\xi_k \geq 0,} \qquad \text{k=1,...,m}$$

C controls the relative importance of maximizing the margin and fitting the training data.
Controls overfitting.

Lagrangian

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2} w.w + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i [y_i(x.w + b) - 1 + \xi_i] - \sum_{i=1}^{m} \beta_i \xi_i$$

α_i's and β_i's are Lagrange multipliers (≥ 0).

Dual Formulation

Find $\alpha_1, \alpha_2, \ldots, \alpha_m$ s.t.

$$\max_\alpha \, \mathrm{J}(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j)$$

Linear SVM

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, m$$

$$\sum_{i=1}^{m} \alpha_i y_i = 0.$$

Noise Accounted

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, m$$

$$\sum_{i=1}^{m} \alpha_i y_i = 0.$$

Solution to Soft Margin Classification

- x_i with non-zero α_i will be support vectors.
- Solution to the dual problem is:

$w = \sum_{i=1}^{m} \alpha_i y_i x_i$
$b = y_k(1 - \xi_k) - \sum_{i=1}^{m} \alpha_i y_i x_i x_k$
for any k s.t. $\alpha_k > 0$
For classification,
$f(x) = \sum_{i=1}^{m} \alpha_i y_i x_i . x + b$
(no need to compute w explicitly)

Thank You