Solving SVM by Decomposition

Marcin Orchel

AGH University of Science and Technology
in Poland

Support Vector Classification

- SVM hard margin classifier – classifies data with the hyperplane that has the largest distance to the closest training vectors
- SVM soft margin classifier – extended version able to classify nonseparable data

Support Vector Classification

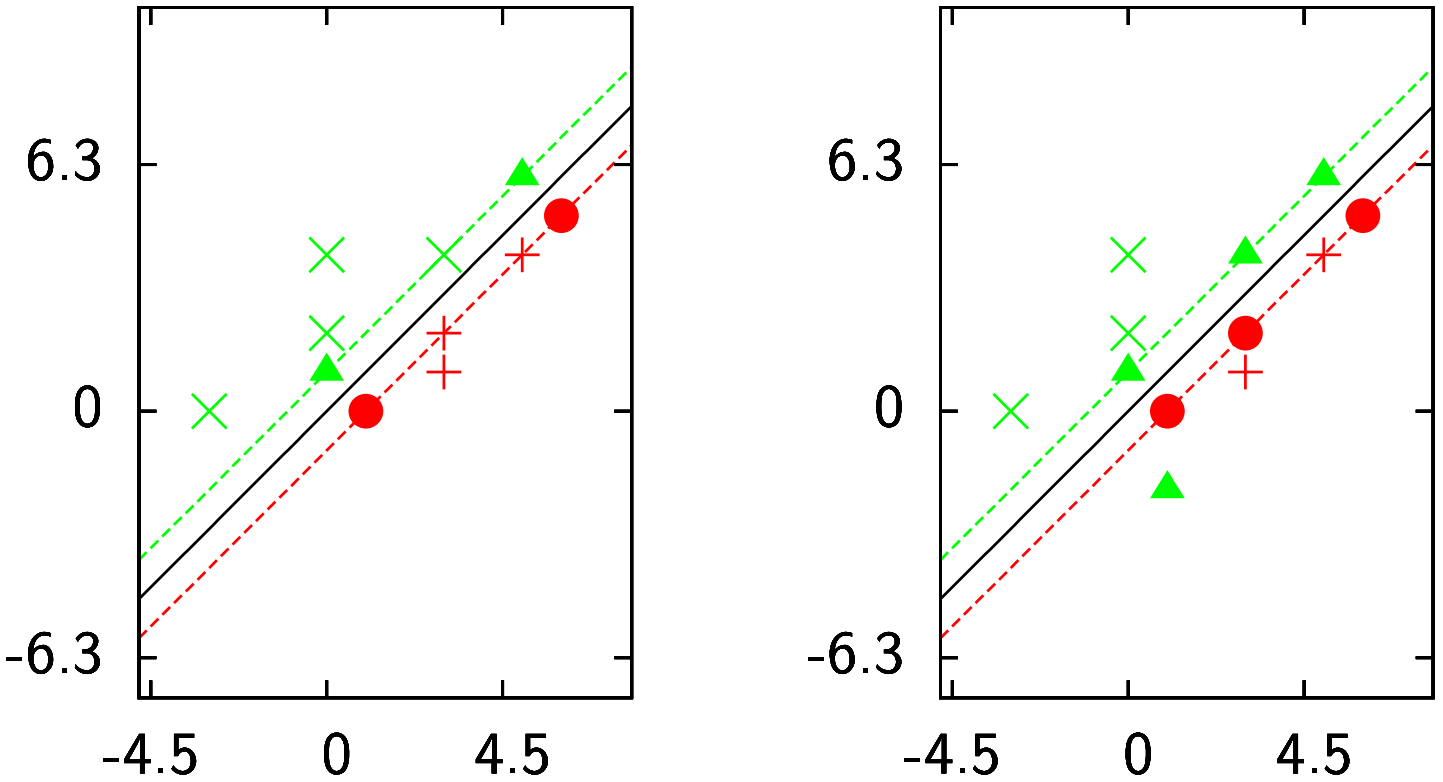

Figure: Two types of margin classifiers: hard on the left, and soft on the right. In the figures, there are example points, support vectors (triangles and circles), solutions (solid lines), margin lines (dashed lines). In the right figure, we can see a misclassified point (1, -2)

Support Vector Classification Formulation

- SVC primal problem:

$$\min f\left(\vec{w}, b, \vec{\xi}\right) = \frac{1}{2}\|\vec{w}\|^2 + \vec{C} \cdot \vec{\xi}$$

subject to $y_i h\left(\vec{x_i}\right) \geq 1 - \xi_i,\ \vec{\xi} \geq 0$ for $i \in \{1, \ldots, n\}$,
where $\vec{C} \gg 0,\ h\left(\vec{x_i}\right) = \vec{w} \cdot \vec{x_i} + b$.

- SVC dual problem:

$$\max W\left(\vec{\alpha}\right) = 1 \cdot \vec{\alpha} - \frac{1}{2}\vec{\alpha^T} Q \vec{\alpha}$$

subject to

$$\vec{\alpha} \cdot \vec{y} = 0,\ 0 \leq \vec{\alpha} \leq C$$

where $Q_{ij} = y_i y_j \left(\vec{x_i} \cdot \vec{x_j}\right)$ for all $i, j \in \{1, \ldots, n\}$.

Support Vector Classification Solution

- decision boundary

$$h^*(\vec{x}) = \sum_{i=1}^{I} y_c^i \alpha_i^* K(\vec{x_i}, \vec{x}) + b_c^* = 0 \qquad (1)$$

- non support vectors have $\alpha_i^* = 0$
- they are ignored in the solution

SVM Optimization Problem Solvers

- Quadratic optimization problem solvers are slow for a big number of data vectors
- Decomposition technique: in every iteration optimize only a few chosen parameters (the active set), the other parameters are frozen, geometrically optimize only chosen subspace
- Sequential Minimal Optimization (SMO): choose only two parameters to the active set
- Solution for 2 parameters can be easily found analytically
- One of the parts of decomposition methods is heuristic: choosing parameters to the working set.
- Strategy of heuristics is to approach to the solution as much as possible in every iteration

AGH

SMO Heuristic

- Optimization conditions: Choose in every iteration the worst two parameters to the active set, which violate Karush-Kuhn-Tucker (KKT) conditions
- More precisely, choose a pair of parameters, that can satisfy the linear and nonlinear conditions after optimization, and optimization is optimal based on partial derivatives of the objective function value with the linear constraint included

AGH

Heuristic of Alternatives

- Heuristic of alternatives chooses those two parameters, for which objective function value growth would be maximal
- In practice, we check only pairs that violate KKT conditions the most or near the most
- In a set of checked pairs there is always a pair, that would be chosen by SMO heuristic
- Time complexity: SMO heuristic: $O(kl)$, heuristic of alternatives: $O(kl + km)$

AGH

Results

- Heuristic of alternatives with 16 alternatives is faster than default heuristic in more than 76% tests and is faster by 32%

Solving SVM Optimization Problem

1. Solve SVM optimization problem by heuristic algorithm with two parameter subproblems solved analytically (SMO algorithm)
2. Solve SVM optimization problem by heuristic algorithm with more than two parameter subproblems solved by quadratic programming solvers
3. Proposed: solve SVM problem by heuristic algorithm with more than two parameter subproblems solved by sequentially running SMO algorithm

Solving multivariable subproblems with SMO

- Comparison of the second solver with the first one. The multivariable solver is faster than SMO.
- Comparison of SMS with the second solver. Both solvers have computation time depended on the subproblem length ($O(p)$), but the SMS is easier to implement
- Comparison of SMS with the first solver. The first solver computes subproblems with two parameters, whilst the SMS computes subproblems with more than two parameters.

AGH

Conclusions

- Heuristic of Alternatives can speed up SMO method
- We can get rid of complex quadratic optimization solvers by using SMS