Machine Learning, Tutorial 6
Universität Bern

Simon Jenni (simon.jenni@inf.unibe.ch)

Support Vector Machines

Let $\{x^{(i)}, y^{(i)}\}_{i=1}^m$ be a set of m training examples of feature vectors $x^{(i)}$ and corresponding labels $y^{(i)} \in \{-1, +1\}$. Bob wants to use a Support Vector Machine (SVM) to learn a binary classifier. The primal formulation is given by:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i \\ \text{s.t.} \quad & y^{(i)}(w^\top x^{(i)} + b) \geq 1 - \xi_i,\ i = 1, \dots, m, \\ & \xi_i \geq 0,\ i = 1, \dots, m \end{aligned}$$

1. Give the decision function Bob should use for a new unseen data-point $x^{(k)}$ in terms of the optimal parameters w^*, b^*, ξ^*.

   Solution:

   The decision should be $+1$ if ${w^*}^\top x^{(k)} + b^* > 0$ and -1 otherwise.

2. Bob knows that the dataset is linearly separable. He argues that in this case it is best to remove the ξ_i. What is your response?

   Solution:

   Bob is wrong. The ξ_i can still lead to better generalization, especially in the presence of outliers.

3. Bob wants to find the best value for the hyper-parameter C using cross-validation. He suggests to perform an initial search with C taking values in $\{-10, -1, -0.1, 0.1, 1.0, 10\}$. What is your advice?

   Solution:

   Negative values for C would encourage violations of the margin. Only positive values for C should be considered.

4. After the initial tuning Bob observes zero training error but high test error. He thinks the problem is high variance and asks you whether he should increase or decrease C. How do you respond?

   Solution:

   The problem is indeed high variance or overfitting. This could be reduced by lowering C.

5. Write the complete dual formulation of the optimization problem.
   Hint: Recall that the dual formulation involves the Lagrangian of the primal problem, and that the Lagrangian is some linear combination of the objective function and of the constraints.

   Solution:

   The Lagrangian is given by

   $$L(w, b, \xi, \alpha, r) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i(y_i(w^T x_i + b) - 1 + \xi_i) - \sum_{i=1}^{m} r_i \xi_i, \tag{1}$$

   where $\alpha_i \geq 0$ and $r_i \geq 0$ are the Lagrange multipliers. The dual formulation is then given by

   $$\max_{\alpha, r} \min_{w, b, \xi} L(w, b, \xi, \alpha, r). \tag{2}$$