20CS6037 Fall 2017: Machine Learning
Instructor: Anca Ralescu

SMO Algorithm

Assigned 10/24/2017
Due 11/03/2016 (11:59PM on BB)

You are to write in MATLAB your own implementation of SVM SMO algorithm along the following lines:

- Given $S = \{(\mathbf{x}_i, y_i); y_i = \pm 1; i = 1, \ldots, l\}$ linearly separable training set, $\epsilon > 0$, find the optimal separating hyperplane using the steps 1-10 shown below.
- Let K denote a kernel on $S \times S$. Here, since the data is linearly separable, you can use $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$, where $\cdot$ denotes the dot product.
- Denote by E_i the following expression:

$$E(i) = \left(\sum_{j=1}^{l} \alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) + b \right) - y_i, i = 1, 2$$

1. Initialize $\alpha = \{\alpha_1, \ldots, \alpha_l\}$ randomly subject to constraint $\sum_{i=1}^{l} y_i \alpha_i = 0$. Set $b = 0$.
2. Calculate the weight vector $\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i$;
3. Calculate KKT conditions:
$$KKT(i) = \alpha_i \{y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1\},$$
where $\cdot$ denotes the *dot product of the two vectors,* $\mathbf{w}$*, and* $\mathbf{x}_i$*.*
4. Pick $\mathbf{x}_1, \mathbf{x}_2$:
   (a) Let $i_1 = \arg\max\{KKT(i) \,|\, \mathbf{x}_i, i = 1, \ldots, l\}$.
   (b) Pick $\mathbf{x}_1 = \mathbf{x}_{i_1}$.
   (c) Calculate $e(i) = E(1) - E(i) = \sum_{j=1}^{l} \alpha_j y_j (K_{j1} - K_{ji}) + y_i - y_1 K_{ij}$
   (d) Let $i_2 = \arg\max\{|e(i)| \,|\, \mathbf{x}_i, i = 1, \ldots, l\}$.
   (e) Pick $\mathbf{x}_2 = \mathbf{x}_{i_2}$.
   (f) Calculate $k = K_{11} + K_{22} - 2 * K_{12}$
5. Update α_2:
$$\alpha_2^{new} = \alpha_2^{old} + \frac{y_2 E(2)}{k}$$
6. Update α_1:
$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$$
7. For $i = 1, \ldots, l$, if $\alpha_i < \epsilon$, $\alpha_i \leftarrow 0$;
8. Select $\alpha_i > 0$, calculate b (from KKT conditions)
9. Test for classification;
10. Repeat from Step 2. until classified.

Please document your program. I will supply you with a data set on which to run your final result. Please upload your program along with results on Blackboard.