20CS6037, ML, Fall 2021
Instructor: Anca Ralescu

Assignment # 6: Implementation of the SMO algorithm
Assigned: November 23, 2021
Due on Canvas: December 1st, 2021, 11:59PM
50 points

**Problem 1.** (15 points) It is easy to compute the change in the dimensionality of the feature space, when a kernel is defined by a polynomial of the type $(p+1)^2$ where $p$ is the dot product in the original feature space. That is, $\mathbf{K}(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z} + 1)^2$.

Consider the more general case: given vectors $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$, and the constant $d$, define $\mathbf{K}(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z} + 1)^d$. This kernel corresponds to an *implicit mapping* $\phi : \mathbb{R}^n \mapsto \mathbb{R}^m$, where $m >> n$ (that is, $m$ is much greater than $n$).

**Find an equation/expression relating $m$, $n$ and $d$.**

**Hint:** Consider successively different values of $n$, for example $n = 2, 3, 4$ and $d$, for example $d = 2, 3$.

**Problem 2.** (35 points) Write your own Matlab/python implementation of the SMO algorithm for hard margin SVM, according to its description shown below for the hard margin SVM.

**Hard Margin SMO**

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 as follows:

1. Initialize $\alpha = \{\alpha_1, \ldots, \alpha_l\}$ <u>randomly, subject to the constraint</u>

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

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:

$$KKT(i) = \alpha_i \{y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1\}; \quad \text{for } i = 1, \ldots, l$$

$$E(i) = \sum_{j=1}^{l} \alpha_j y_j K_{ji} - y_i; \quad \text{for } i = 1, \ldots, l$$

4. Pick $\mathbf{x}_1, \mathbf{x}_2$:

(a) Let $i_1 = \arg\max_{i=1,\ldots,l} KKT(i)$.

(b) Pick $\mathbf{x}_1 = \mathbf{x}_{i_1}$.

(c) Calculate $e(i) = E(1) - E(i) = \sum_{j=1}^{l} \alpha_j y_j (K_{j1} - K_{ji}) + y_i - y_1$

- Note: $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$ when the data set is linearly separable. Here $\cdot$ means the *dot product.*

(d) Let $i_2 = \arg\max_i e(i)$.

(e) Pick $\mathbf{x}_2 = \mathbf{x}_{i_2}$.

(f) Calculate $k = K_{11} + K_{22} - 2K_{12}$

5. Update $\alpha_2$:

$$\alpha_2^{new} = \alpha_2^{old} + \frac{y_2 e(2)}{k}$$

6. Update $\alpha_1$:

$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$$

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.

## NOTE

In the above, you may find it easier not to rename $i_1$ and $i_2$ as 1 and 2 respectively. In that case, you will use everywhere $i_1$ and $i_2$. That means

$$y_1 \to y_{i_1}; y_2 \to y_{i_2};$$

$$\alpha_1 \to \alpha_{i_1}; \alpha_2 \to \alpha_{i_2};$$

$$K_{12} \to K_{i_1 i_2}; K_{11} \to K_{i_1 i_1}; K_{22} \to K_{i_2 i_2};$$

$$E(1) \to E(i_1)$$

$$e(i) = E(i_1) - E(i); i = 1, \ldots, l$$