Assignment 4

Instructions:

1. The problems can be coded in either in R or python.
2. In Q1 and Q2, use built-in commands for computing the SVM classifier. Use *SVC* from *sklearn.svm* in python, and *svm* from the library *e1071* in R. Do not use built-in commands for any other functions (e.g: cross-validation).
3. In Q1, the value of C and the parameters can be computed using trial and error. The intention is to let you know what happens for different values of the parameters.
4. In Q3, the initial learning rate γ_0 needs to be "guessed" so as to converge to the optimal coefficients.

Problems:

1. Consider the data set where $X_i \in \mathbb{R}$ and the true response is such that

$$Y_i = \begin{cases} +1 & \text{if } |X_i| > 3, \\ -1 & \text{otherwise.} \end{cases}$$

Fix the feature set in the training data set to be all integers in the interval $[-100, 100]$, except (-3) and $(+3)$ (i.e., a total of 199 data points). Fit an SVM classifier on the training data set. The fit requires you to find the Kernel (i.e., linear, polynomial, radial, or neural network), the value of C in the optimization problem, and the parameters required in the Kernel. **Print** the support vectors in the training data, and the dual coefficients α_i corresponding to the support vectors. **Plot** the decision function of the classifier along with the scatter plot of the generated test data. (Note that the right choice of parameters would give only four support vectors: $\{-4, -2, 2, 4\}$.)

2. (a) Generate 200 data points $X_1, X_2, \ldots, X_{200}$ i.i.d. $\sim \mathcal{N}(0, 1)$. Decide Y_i as follows:

$$Y_i = \begin{cases} +1 & \text{if } X_i^{(1)} + X_i^{(2)} > 0, \\ -1 & \text{otherwise.} \end{cases}$$

Plot the generated data using scatterplot. Use different colors to represent features with different labels.

(b) Among $C = \{1, 10, 100, 1000, 10^4\}$, use a five-fold cross-validation to determine which value of C gives the least cross-validation error for an SVM classifier with linear kernel. **Print** the value of C. Compute the SVM classifier using this value of C. **Plot** the classifier along with the scatterplot of the features.

(c) Now construct the test data $X_{201}, \ldots, X_{400}$ exactly the same way as in Q2a. Predict the response using the SVM classifier constructed in Q2b. **Plot** the classifier along with the scatterplot of the features of the test data. **Print** the test error.

3. Consider minimizing the following objective function with respect to β_1, β_2:

$$\sum_{i=1}^{200} \left(1 - Y_i(\beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)})\right)_+ + \frac{1}{2C}(\beta_1^2 + \beta_2^2),$$

where (X_i, Y_i) refer to the data generated in Q2a.

(a) Let $\beta^{(0)} = [\beta_0, 0, 0]$, and $g(x) = \sum_{i \in S} \nabla f_i(x)$, where β_0 is the value of β_0 computed in Q2b, and

$$\nabla f_i(x) = \begin{cases} -Y_i[X_i^{(1)}, X_i^{(2)}] + \frac{1}{C}[\beta_1, \beta_2] & \text{if } (Y_i(\beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)})) \leq 1, \\ \frac{1}{C}[\beta_1, \beta_2] & \text{otherwise.} \end{cases}$$

Fix $|S| = 1$. Apply the stochastic gradient descent algorithm with inverse scaling. Choose an appropriate value of the initial learning rate γ_0, with $t = 0.9$. **Print** the number of iterations needed to get to $||\beta^{(k)} - \beta^*|| \leq 10^{-4}$. **Plot** the trajectory of β vector.

(b) Repeat Q3a with $|S| = 5$. **Print** the number of iterations needed to get to $||\beta^{(k)} - \beta^*|| \leq 10^{-4}$. **Plot** the trajectory of β vector.