REI602M **Quiz 29.1.2021**

Hand in on Gradescope before 22:00 on Jan. 30 (Saturday). Each question will be given 1, 0.5 or 0 points as follows. If the question is more or less correct it gets 1 point. If it is partly correct it gets 0.5, and if it is missing or completely wrong it gets 0 points.

1)

An algorithm for *fault detection* takes an unlabeled data set as input, $\{x^{(1)}, x^{(2)}, \dots x^{(n)}\}$ and provides a decision rule for classifying examples x as faulty or not. A simple fault detection algorithm solves the following optimization problem

$$\min_{R,c} R^2$$

$$\text{subject to } ||x^{(i)} - c||^2 \leq R^2,\ i = 1, \dots n.$$

This corresponds to finding the radius R of the smallest sphere with center c that contains all the points in the training set. A point x is classified as not faulty if it lies inside the sphere, and faulty otherwise.

The above algorithm is sensitive to outliers in the sense that a single point can increase the radius significantly. Modify the above algorithm so that it becomes more resistant to outliers. *Hint*: Recall soft vs hard-margin SVM.

I would use the SVM with a Radial Basis Kernel.

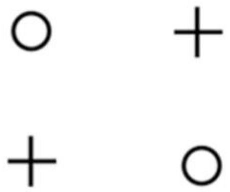
$$\min_{R,c,\gamma} R^2$$

subject to $e^{-\gamma||x^{(i)}-c||^2} <= R^2, i = 1, \dots . n.$

where γ controls the decision boundary.

2)

Consider the following data set



Circle *all* the classifiers that can achieve zero training error on this data set.

i) Logistic regression

ii) SVM with a RBF kernel.

iii) 3-NN classifier.

only SVM with a RBF kernel.