1.

Given a collection of labeled 2-dimensional points in Euclidean space as the training data, each of them has a label of either 0 or 1: $(1, 9)$, $(2, 2)$, and $(8, 1)$ have labels of 0; $(8, 9)$, $(9, 6)$, and $(10, 10)$ have labels of 1. Consider using the K-nearest neighbors (KNN) algorithm in a **semi-supervised** way to predict the labels of unlabeled points: Use the training data as the KNN model to predict the labels with confidence scores for all unlabeled points in the test data; take the test point with its predicted label that has the highest confidence score, add it with its predicted label to the training data, and remove it from the test data; repeat the process mentioned above iteratively until all the given test points are labeled. Here, the distance d between two points (x_1, y_1) and (x_2, y_2) is the square of the L_2 distance between them: $(x_1 - x_2)^2 + (y_1 - y_2)^2$. For simplicity, use $K = 1$ for KNN and, when predicting the label for a test point, use $1/d^*$ as the confidence score, where d^* is the distance between that test point and its nearest point in the training data. Given a collection of unlabeled points as the test data: $(0, 0)$, $(4, 7)$, and $(6, 8)$, use the semi-supervised method mentioned above to predict the labels for the test data. Write the prediction process step by step.

2.

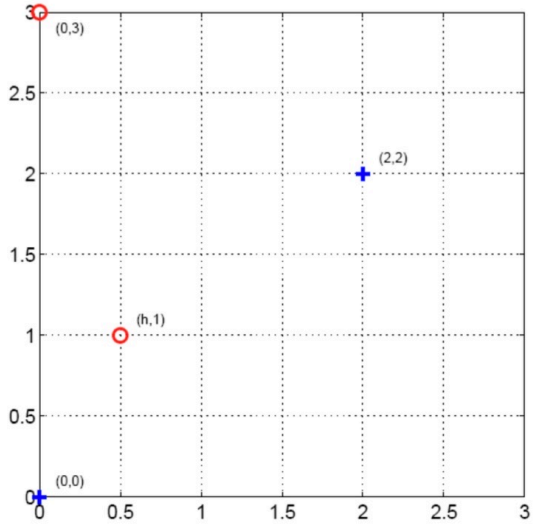
We have a dataset whose points have one features and are in two classes: $\{-4, 2, 3\}$ are in the negative class and $\{-3, -2, 1\}$ in the positive class.

(a) Disregard the labels and cluster the data in two groups using K-means. Label all data in each cluster using the label of the center of the cluster. What is the misclassification error rate (%) of the classifier you built?

(b) Assume that we have a simple threshold classifier that classifies a data point x to the positive class if $x \geq -1.5$ and in the negative class otherwise. Assume that we are implementing *active learning* with uncertainty sampling, so we ask the "oracle" to give us the label of the most uncertain data point and update the threshold of our classifier only if it misclassifies the new point. We only use the threshold among $\{-3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$ that is the closest to the newly labeled data point and fixes its misclassification. Run the algorithm three iterations (i.e. by asking the labels of three data points from the oracle) and provide the final classifier.

3.

3. Suppose we only have four training observations in two dimensions:



positive examples are $\mathbf{x}_1 = [0\ 0]^T, \mathbf{x}_2 = [2\ 2]^T$ and negative examples are $\mathbf{x}_3 = [h\ 1]^T, \mathbf{x}_4 = [0\ 3]^T$. h is a parameter.

(a) What is the largest value of h for which the training data are still linearly separable?

(b) Determine the support vectors when $h = 0.5$.

(c) When the training points are separable, does the slope of the maximum margin classifier change? Why?

(d) Assume that $h = .5$ and we have unlabeled data $\mathbf{x}_5 = [3\ 3]^T, \mathbf{x}_6 = [2\ 0.5]^T, \mathbf{x}_7 = [1\ 1.5]^T, \mathbf{x}_8 = [2.5\ 1.5]^T$. Which one will be labeled first, if we are performing self-training? Which one will be labeled first, if we are performing active learning?