CSE 151A Winter 2021 - Take Home Exam 2

1 hour **Name:** **PID:**

---

(1) This is an open book, open notes exam. You are free to consult any text book or notes. **You are not allowed to consult with any other person.**
(2) If you need any clarification, please post a private message to the instructors on Piazza.
(3) Remember that your work is graded on the *clarity* of your writing and explanation as well as the validity of what you write.
(4) This is a one-hour exam.

(1) Draw the decision boundary for the nearest neighbor classifier on the following data points in 2 dimensions.

$$((0,0),0),((4,1),1),((-2,3),0)$$

For full credit, write down the equation for each segment of the decision boundary, and label each region with the label assigned to it by the classifier.

Solution. We first find the equations of each segment of the decision boundary. There are at most three segments which are each part of the three lines separating any pair of the points. The equations of these lines are given by:

(a) Line separating $(0,0)$ and $(4,1)$:

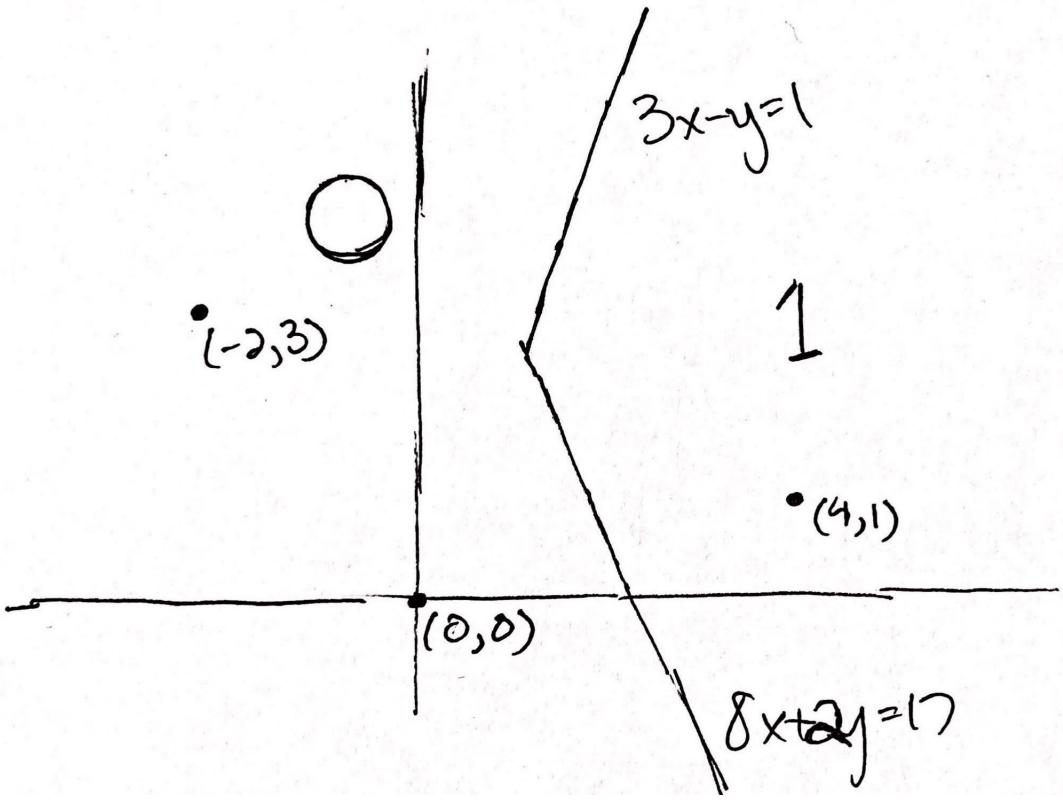
$$\begin{aligned}(x-0)^2+(y-0)^2 &= (x-4)^2+(y-1)^2\\ x^2+y^2 &= x^2-8x+16+y^2-2y+1\\ 8x+2y &= 17\end{aligned}$$

(b) The line separating $(0,0)$ and $(-2,3)$ does not affect the decision boundary since both points are classified as 0.

(c) Line separating $(4,1)$ and $(-2,3)$:

$$\begin{aligned}(x-4)^2+(y-1)^2 &= (x+2)^2+(y-3)^2\\ x^2-8x+16+y^2-2y+1 &= x^2+4x+4+y^2-6y+9\\ -12x+4y &= -4\\ 3x-y &= 1\end{aligned}$$

The segments, when plotted, look like this:


3x-y=1
(-2,3)
1
(4,1)
(0,0)
8x+2y=17

(2) Remember that any classifier is basically a function that takes in a feature vector in $\mathbb{R}^d$ and outputs a label. We say that two classifiers C and C' are equal if they output the same label for every feature vector x in $\mathbb{R}^d$. Formally, for all $x \in \mathbb{R}^d$, $C(x) = C'(x)$.

Suppose Alice and Bob are both building a 1-nearest neighbor classifier on the same training dataset S. To build the classifier Alice computes the nearest neighbosr using the Euclidean distance to get the classifier C.

(a) (5 points) Suppose Bob computes the nearest neighbors using the square of the Euclidean distance and gets the classifier C'. Is C' equal to C for all training dataset? If yes, provide a short proof, and if no, provide a counterexample.

Solution. The answer is yes, with the following reasoning. To classify a point x using k-nearest neighbors, one picks the closest k points to x. Whether using Euclidean distance or squared Euclidean distance, the order of the closest points to x doesn't change. Thus, Alice's classifier and Bob's classifier both pick the same k points.

(b) (5 points) Now, instead of the square of the Euclidean distance, Bob computes the nearest neighbors using the following distance function:

$$d(x, x') = \sum_{i=1}^{d} (x^i - x'^i)^3$$

where x^i denotes coordinate i of the vector x. This gives him a classifier C''. Is C'' equal to C for all training datasets? Justify your answer.

Solution. The answer is no. The Euclidean distance and Bob's distance do not preserve the ordering of the points. To see this, consider the following training set $\{((a, a), 1), ((b, 0), 0)\}$, and suppose we classify the point $(0, 0)$. Alice's classifier calculates a distance of $\sqrt{2}a$ and b, respectively. Bob's classifier calculates a distance of $2^{1/3}a$ and b, respectively. Setting a, b such that $2^{1/3}a < b < \sqrt{2}a$, which has a solution, forces the classifiers to predict different labels.

Thus, the k nearest points selected by Alice's classifier need not be the k nearest points selected by Bob's, possibly resulting in different classification.