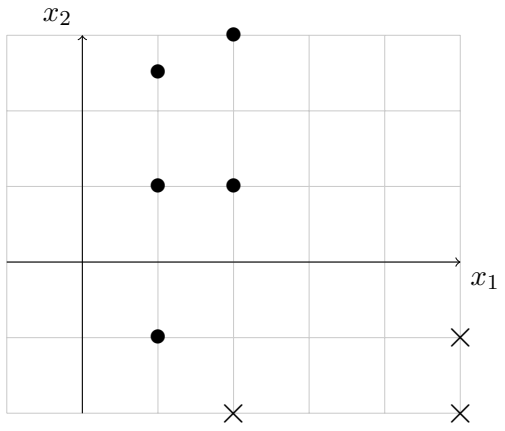
Name: _____

Nearest Neighbors

1. (8 points) Consider the following 2D dataset:

x	y
$(1, -1)$	$+1$
$(1, 1)$	$+1$
$(1, 2.5)$	$+1$
$(2, -2)$	-1
$(2, 1)$	$+1$
$(2, 3)$	$+1$
$(5, -1)$	-1
$(5, -2)$	-1

The dataset is plotted below, with positively labeled points as solid points (•) and negatively labeled points as X marks (×):



Break ties in distance by choosing the point with smaller x_1 coordinate, and if still tied, by smaller x_2 coordinate.

(a) Compute the leave-one-out cross validation accuracy (i.e., average 8-fold cross validation accuracy) of the 1-nearest-neighbor learning algorithm on this dataset.

> **Solution:** 6/8. When left out of the training set, the point at (1,-1) will be misclassified during testing; similarly for the point at (2,-2).

(b) Compute the leave-one-out cross validation accuracy of the 3-nearest-neighbor learning algorithm on this dataset.

Solution: 7/8. Now only the point at (2,-2) will be misclassified during testing, when left out of the training set.

(c) In the case of the 1-nearest-neighbor learning algorithm, is it possible to strictly increase the leave-one-out cross validation accuracy on this dataset by changing the label of a single point in the original dataset? If so, give such a point.

Solution: Yes. Change either point at $(2, -2)$ to $+1$, or point at $(1, -1)$ to -1.

(d) How about in the case of the 3-nearest neighbor algorithm? If so, give such a point.

Solution: No, not possible. If we try to change the point at (2, -2) to $+1$, then that point will be correctly predicted during cross-validation as $+1$. Unfortunately, with that change the two points at (5,-1) and (5,-2) will now be misclassified, making our cross-validation accuracy worse.