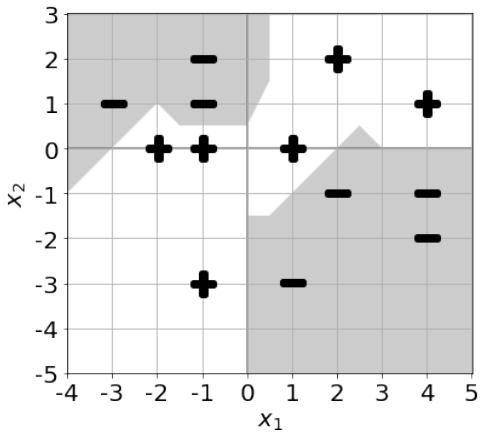
Name: _____

Nearest neighbor classifiers

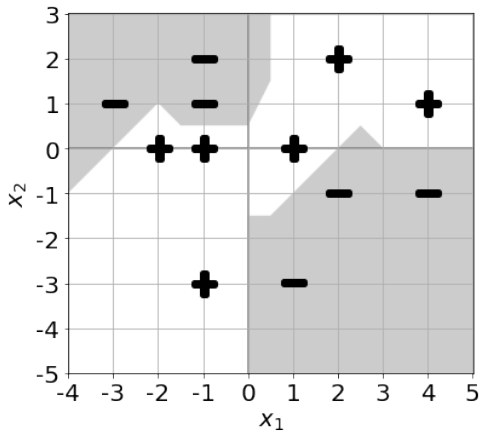
4. (12 points) This question asks about learning nearest neighbor (NN) classifiers. Assume that we are using Euclidean distance squared as the distance metric, i.e. $d(x, x') = ||x - x'||^2$.

   (a) Draw on the below figure the decision boundary for a 1-NN classifier on this data set. In each region, denote whether the classification of any point (*any* point, not just the training data) in that region would be $+1$ or -1. (Note, all data points are assumed to be on integer coordinates.)



Solution:

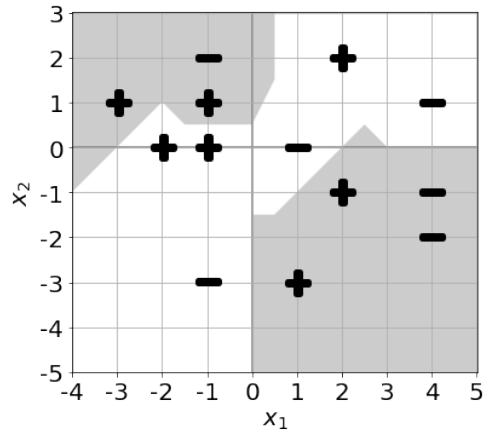
   (b) Which training data points, if any, could you remove and keep the decision boundary identical? Answer using their (x_1, x_2) coordinates.



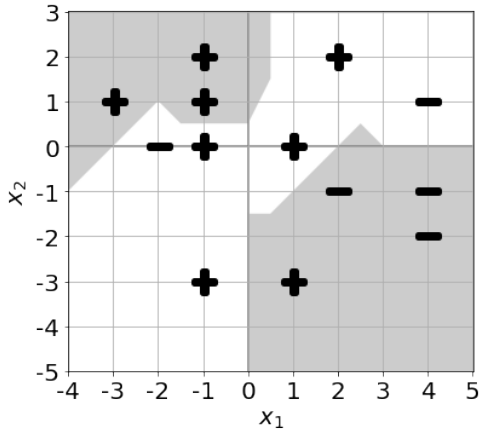
Solution: (4, -2)

   (c) You perform leave-one-out cross-validation of the 1-NN and 3-NN classifiers on this data set, i.e. you use use cross-validation with a chunk size of 1 data point. Assume ties go to the $+1$ region. What cross-validation errors do you obtain?

Name: _____



Solution: 1-NN: 7 / 13



3-NN: 6 / 13

(d) Suppose we now use the following feature transformation, $\phi(x_1, x_2) = x_1 x_2$, and seek to learn a nearest neighbor classifier in the transformed space. This is equivalent to using a different distance metric, $d(x, x') = \|\phi(x) - \phi(x')\|^2$. What is the average leave-one-out cross-validation error of a 3-NN classifier using this new distance metric? Which points would be misclassified (specified using their (x_1, x_2) coordinates)?

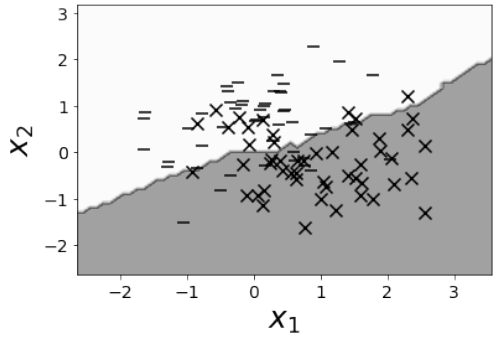
Solution:
3-NN:

1 / 13 Misclassified points:

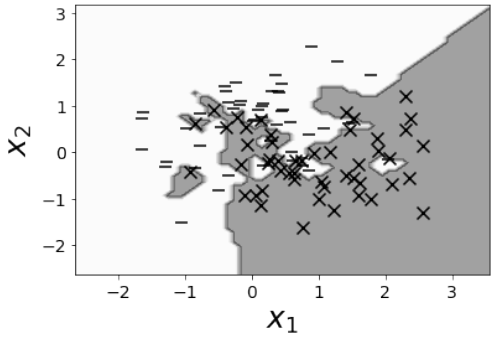
(-1, 1)

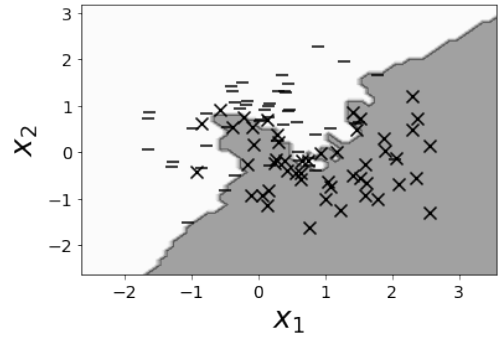
(e) The plots below show the decision boundaries as predicted by a k-NN classifier for four different values of k: 1, 5, 20, 40. Map each plot to the corresponding value of k.

(I)

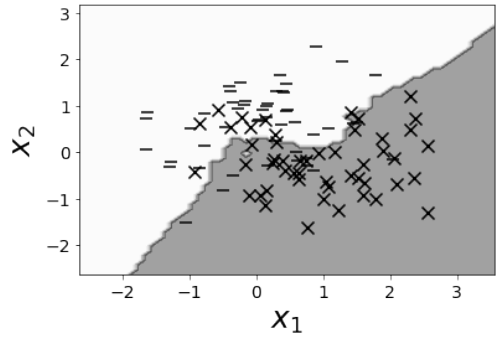


(II)





(III)



(IV)

$k = 1$:	◯ (I)	√ **(II)**	◯ (III)	◯ (IV)
$k = 5$:	◯ (I)	◯ (II)	√ **(III)**	◯ (IV)
$k = 20$:	◯ (I)	◯ (II)	◯ (III)	√ **(IV)**
$k = 40$:	√ **(I)**	◯ (II)	◯ (III)	◯ (IV)