Name: _____

Trial separation

Let's look at linear separability and linear classification.

3. (10 points) Linear separability.

   (a) Consider the following $n = 4$ data set with 4-dimensional data points (recall that each column represents one data point):

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} +1 & +1 & -1 & -1 \end{bmatrix}$$

   Is the data linearly separable? If yes, please provide a classifier θ, θ_0 that correctly classifies the data. If no, please explain why not.

   For each of the following True/False questions, please provide a brief explanation following your answer.

   (b) If we take any linearly separable data set and *add* a new feature, it is still guaranteed to be linearly separable.

   ◯ True ◯ False

   (c) If we take any linearly separable data set and *remove* a feature, it is still guaranteed to be linearly separable.

   ◯ True ◯ False

(d) If we take any data set that is not linearly separable and *remove* a feature, it is still guaranteed to not be linearly separable.

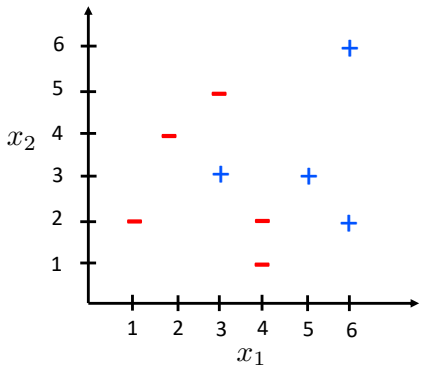
◯ True ◯ False

(e) If we take any data set that is not linearly separable and remove a *data point*, it is still guaranteed to not be linearly separable.

◯ True ◯ False

4. (6 points) Consider the data set shown in the box below.

(a) Draw a hyperplane that obtains the smallest training error (i.e., highest accuracy). Be sure to also draw the normal vector.



(b) Suppose we remove data points ($x_1 = 3$, $x_2 = 3$) and ($x_1 = 4$, $x_2 = 2$). And let us say that two hypotheses are considered different if there exists a test point (i.e., not necessarily from the data set shown) that they would classify differently.

How many different hypotheses are there that obtain zero training error? Explain your answer.