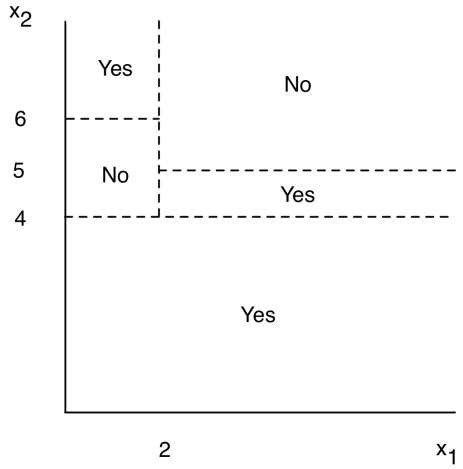
Name: _____

1. **Decision Trees** Consider the following figure, which shows the decision surface learned by a decision tree over two numerical features X_1 and X_2. The tree predicts a target attribute that takes on the values "Yes" and "No."



Draw a decision tree that is consistent with the above classification decisions. Try to make your tree as "compact" as possible (i.e., as few decision nodes as possible).

2. **Perceptrons** Suppose you use a perceptron to learn a linear discriminant of a target attribute that takes on two values "Yes" and "No." You give it labeled data that has two features, X_1 and X_2 and it returns the following set of three weights: $w_0 = 12$, $w_1 = 2$, $w_2 = -4$.

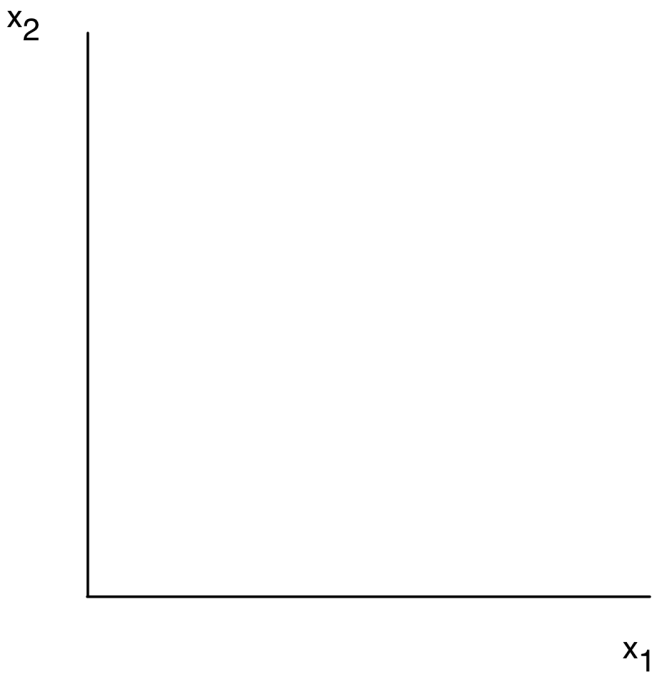
   Recall that a perceptron makes a prediction by calculating

$$h(x) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right)$$

   Here we'll assume that $h(x) = 1$ will be interpreted as a prediction of "Yes" and $h(x) = -1$ will be interpreted as a prediction of "No."

   (a) What label would the input $x_1 = 2$, $x_2 = 5$ be assigned?

   (b) Draw the decision boundary in the feature space below. In other words, show which points will be labeled "Yes" and which ones labeled "No."



3. **Feature Engineering** You are a Hollywood producer. You have a script in your hand and you want to make a movie. Before starting, however, you want to predict if the film you want to make will rake in huge profits, or utterly fail at the box office. You hire two critics A and B to read the script and rate it on a scale of 1 to 5 (assume only integer scores). Each critic reads it independently and announces their verdict. Of course, the critics might be biased and/or not perfect. For instance, for the past five movies you made, these are the critics scores and the performance of the movie:

Movie Name	A	B	Profit?
Mad Max	2	2	No
Harry Potter I	4	5	No
Harry Potter II	3	2	Yes
Mystic Pizza	3	4	Yes
Citizen Kane	2	3	Yes

You decide to use machine learning to learn a classifier that will predict the profitability of the movie based on the reviewers' ratings. Suppose you directly use the scores given above as features. That is, x_A = score given by A and x_B = score given by B.

Would the perceptron converge on this data? (Hint: you might try plotting the data as points in x_A, x_B space.)

4. Your co-executive producer, Mark Wahlberg, thinks your whole approach is ridiculous. He claims that a perceptron isn't "powerful" enough to make predictions because it's too simplistic. He wants some evidence that a perceptron could get the right answer under various scenarios.

   For each scenario below, circle it if a perceptron using the features above could indeed perfectly classify the data. You can assume that for each scenario, you will have available a large amount of training data that follows the pattern described in the scenario.

   (a) One or more good reviews leads to success. If the total of their scores is more than 8, then the movie will definitely be a success and otherwise it will fail.

   (b) Mediocre reviews breeds success. Your movie will succeed if and only if each reviewer gives either a score of 2 or a score of 3.

   (c) Your reviewers have complementary tastes. Your movie will succeed if and only if both reviewers agree.

5. At Mark Wahlberg's insistence, you try using a *different* set of features. Consider the following feature space:

$$\begin{aligned}
x_0 &= 1 \quad \text{(The bias feature, always equal to 1)} \\
x_{1A} &= 1 \quad \text{if score given by A is 1, 0 otherwise} \\
x_{1B} &= 1 \quad \text{if score given by B is 1, 0 otherwise} \\
x_{2A} &= 1 \quad \text{if score given by A is 2, 0 otherwise} \\
x_{2B} &= 1 \quad \text{if score given by B is 2, 0 otherwise} \\
&\dots \quad \dots \\
x_{5A} &= 1 \quad \text{if score given by A is 5, 0 otherwise} \\
x_{5B} &= 1 \quad \text{if score given by B is 5, 0 otherwise}
\end{aligned}$$

Consider again the three scenarios in part (4). Using a perceptron with the new features, which of the three scenarios can be perfectly classified? If it can be correctly classified, **describe a setting of the weight vectors that would achieve perfect classification**. (There are 11 weights in total, one per feature. The weights are w_0 and then weights w_{iA} and w_{iB} for each $i = 1 \dots 5$.)

(a) One or more good reviews leads to success. If the total of their scores is more than 8, then the movie will definitely be a success and otherwise it will fail.

(b) Mediocre reviews breeds success. Your movie will succeed if and only if each reviewer gives either a score of 2 or a score of 3.

(c) Your reviewers have complementary tastes. Your movie will succeed if and only if both reviewers agree.

6. For any above scenario above where the perceptron would fail, suggest a feature or features that could be added to ensure perfect classification.

Last exercise adapted from materials by Dan Klein and Pieter Abbeel.