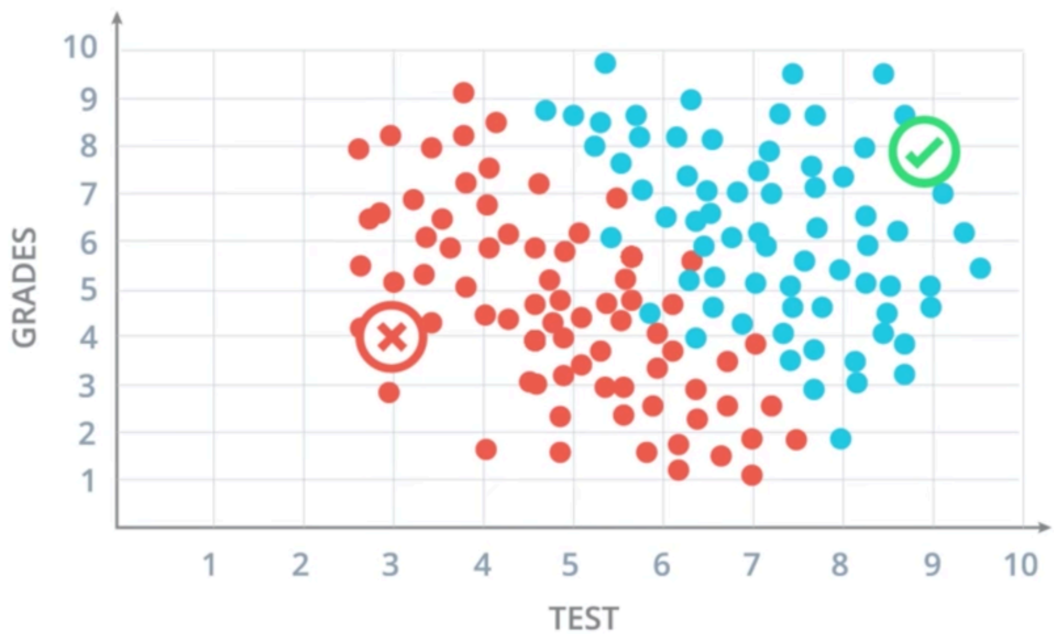
Quiz Question

Which of the following describe a classification problem?

a. A predictive modeling problem where a numeric value is predicted based on training for a given dataset.

b. A predictive modeling problem where a predefined class label is predicted based on training for a given dataset.



Quiz Question

The chart above demonstrates that students with higher test scores and higher grades are more likely to be accepted into our hypothetical university.

If the student has a test score of 7 out of 10 and grades of 6 out of 10, does the student get accepted?

a. Yes

b. No

Linear Boundaries

First, we label the horizontal axis corresponding to the test by the variable x_1, and the vertical axis corresponding to the grades by the variable x_2. So this boundary line that separates the blue and the red points is going to have a linear equation. The one drawn has equation $2x_1 + x_2 - 18 = 0$.

What does this mean?

Our method for accepting or rejecting students simply says the following: take this equation as our score, the score is $2 * Test + Grades - 18$. When the student comes in, we check their score. If their score is a positive number, then we accept the student and if the score is a negative number then we reject the student.

This is called a prediction.

- Score > 0: Accept
- Score < 0: Reject

That linear equation is our model. In the more general case, our boundary will be an equation of the following $w_1x_1 + w_2x_2 + b = 0$. We'll abbreviate this as $Wx + b = 0$, where $W = (w_1, w_2)$ and $x = (x_1, x_2)$ and we take the product of the two vectors.

- x is the input
- W is weight
- b is the bias

For a student at coordinates (x_1, x_2), we'll denote a label as Y and the label is what we're trying to predict. If the student gets accepted (the point is blue), then the label is `Y=1`. If the student gets rejected (the point is red), then the label is `Y=0`.

Our prediction is going to be called $\hat{y}$ and it will be what the algorithm predicts that the label will be. In this case, $\hat{y} = 1$ if the algorithm predicts that the student gets accepted, which means the point lies over the line. And, $\hat{y} = 0$ if the algorithm predicts that the student gets rejected, which means the point is under the line.

> The goal of the algorithm is to have $\hat{y}$ resembling Y as closely as possible, which is exactly equivalent to finding the boundary line that keeps most of the blue points above it and most of the red points below it.

Quiz Question

Now that you know the equation for the line ($2x_1 + x_2 - 18=0$), and similarly the "score" ($2x_1 + x_2 - 18$), what is the score of the student who got 7 in the test and 6 for grades?

Answer: 2*7+6-18 = 2

More columns of data = higher dimensions

What happens if we have more data columns, such as the ranking of the student in the class. How do we fit three columns of data?

We have 3 axes: x_1 for the test, x_2 for the grades, and x_3 for the class ranking. Our equation will now be a plane in three dimensions with the equation $w_1x_1 + w_2x_2 + w_3x_3 + b = 0$ which will separate this space into two regions. This equation can still be abbreviated by $Wx + b = 0$, except our vectors will now have three entries instead of two.

Our prediction will still be

- $\hat{y} = 1$ if $Wx + b \geq 0$
- $\hat{y} = 0$if $Wx + b < 0$

What if we have many columns? Our data just leaps in n-dimensional space. If we can imagine that the points are just things with n coordinates called x_1, x_2, x_3 all the way up to x_n with our labels being y, then our boundary is just an $(n-1)$-dimensional hyperplane, which is a high dimensional equivalent of a line in 2D or a plane in 3D. And the equation of this $(n-1)$-dimensional hyperplane is going to be $w_1x_1 + w_2x_2 + ... + w_nx_n + b = 0$ which we can still abbreviate as $Wx + b = 0$, where our vectors now have n entries.

Our prediction is still the same as before. It is

- $\hat{y} = 1$ if $Wx + b \geq 0$
- $\hat{y} = 0$if $Wx + b < 0$

Quiz Question

Simply put, extra higher dimensions are just added columns of data in a dataset. In an example for a dataset dealing with cars, what might be some higher dimensions

a. Engine size
b. Price
c. Safety rating
d. Feedback comments on social media about the car
e. Color

Quiz Question

Which of the following are true statements about 'Perceptrons'? (You can pick multiple answers)

a. Perceptrons can have any number of inputs

b. Perceptrons have two nodes, one that solves a linear equation and one that applies a step function.

c. The nodes of a perceptron can appear in any order

Quiz Question

Which of the following is an accurate summary of the 'step function'?

a. The step function returns 1 if the input is positive or zero, and a zero if the input is negative.

b. The step function returns zero if the input is positive or zero, and a 1 if the input is negative.

c. The step function returns one if the input is positive and a zero if the input is negative or zero.

Looking back to the video, Consider this equation about a student

$$Score = (2 Test) + (1 Grade) - 18$$

Suppose w_1 was 1.5 instead of 2.

Quiz Question

Would the student who got 7 on the test and 6 on the grades be accepted or rejected?

a. Accepted

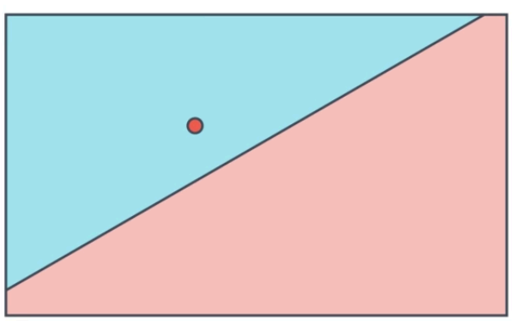
b. Rejected

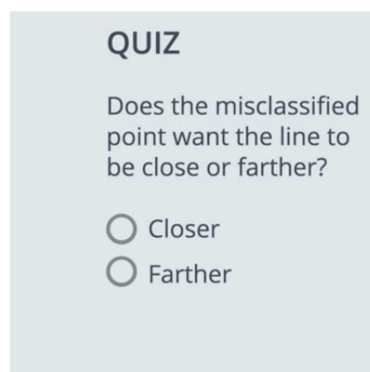
Finding Line Boundaries

Let's answer this question by first looking at a small example with three blue points and three red points. We will start to explore an algorithm that will find the line that splits these points properly.

- So the computer doesn't know where to start. It might as well start at a random place by picking a random linear equation. This equation will define a line and a positive and negative area given in blue and red respectively.
- Now we look at how badly this line is doing and then move it around to try to get better. How do we find how badly this line is doing? So let's ask all the points.
    - First, look at the **correctly classified dots**, (four points are correctly classified, two blue points in the blue area and two red points in the red area.) These points are correctly classified, so they say, "I'm good."
    - Then we have two points that are **incorrectly classified**. (the red point in the blue area and this blue point in the red area.) We want to get as much information from them so we want them to tell us something so that we can improve this line. So what is it that they can tell us?

Quiz Question





The image above, shows a misclassified point, a red point in the blue area.
If you were the point, what would you tell the line to do?

Would you like it to come closer to you or farther from you?

a. Closer

b. Farther

Quiz Question

What is the **learning rate** and what is its purpose?

a. A hyperparameter value that sets a maximum number for the number of times the line will move during training. It is used to limit the line moving closer towards the points.

b. A hyperparameter value that is used in the calculation that controls how fast/far the model adjusts during training. It is used to move the line closer towards the points.

c. A hyperparameter value that sets a minimum number for the number of times the line will move during training. It is used to limit the line moving closer towards the points.

Quiz Question

Which of the following highlights the important aspects of the Perceptron Trick?

a. Use the coordinates of misclassified points to adjust the parameters in the line equation. To keep the adjustments small, the adjustment calculations are multiplied by the Learning Rate.

b. Use the coordinates of correctly classified points to adjust the parameters in the line equation. So that the adjustments are significantly noticeable, the calculations are multiplied by the Learning Rate.

For the second example, where the line is described by $3x_1+4x_2-10=0$, if the learning rate was set to 0.1, how many times would you have to apply the perceptron trick to move the line to a position where the blue point, at (1, 1), is correctly classified?

Answer: 10 times

e = 3*1 + 4*1 -10 = -3
w1 = w +x*LR = 3+1*0.1 = 3.1
w2 = 4.1
b = -10+0.1 = -9.9
e1 = -2.7
w1 = 3.2
w2 = 4.2
b = -9.8
e2 = -2.4
e3 = -2.1
e4 =-1.8
e5 =-1.5
e6 = -1.2
e7 = -0.9
e8 = -0.6
e9 = -0.3
e10 = 0