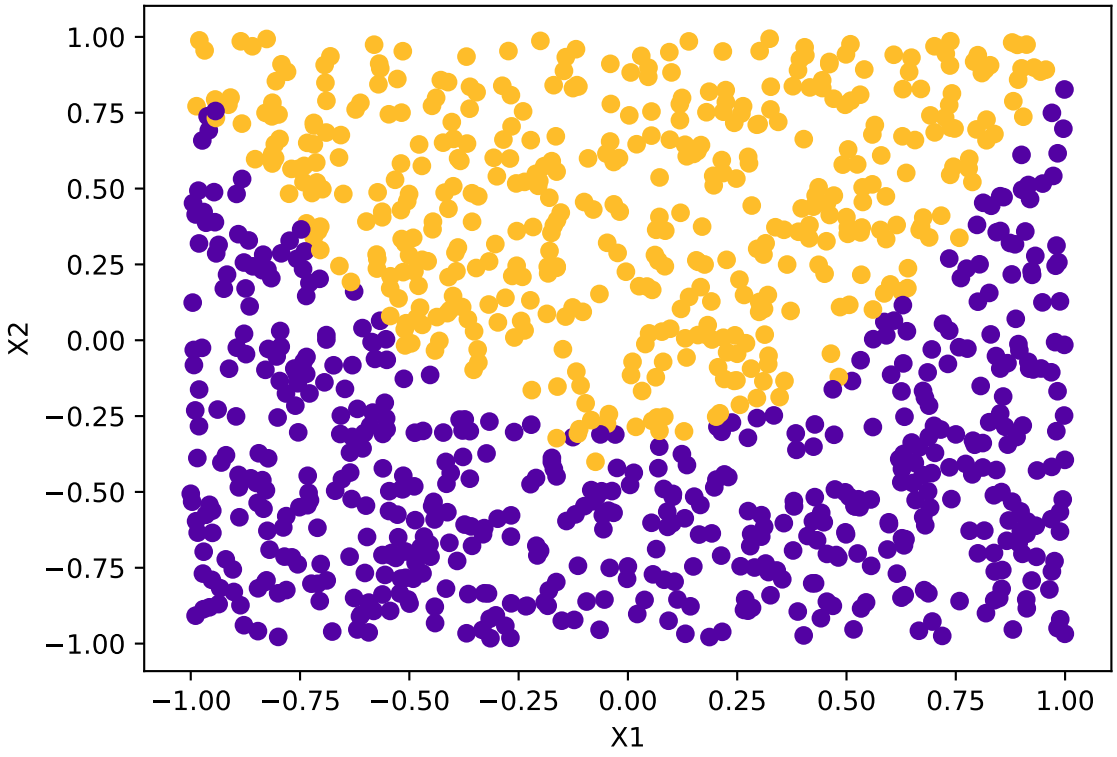
HW1 Written Part

Due 5pm Wednesday Jan. 29, 2020

What is your name?

Problem 1

Here's a fake data set with 1000 observations of two quantitative features X_1 and X_2 and one categorical response Y:



(a) On the plot, draw your guess at a good decision boundary: you will predict that y = 0 (plotted in purple) on one side of the decision boundary and y = 1 (plotted in orange) on the other side of the decision boundary.

(b) Write down a reasonable equation for the decision boundary in terms of $x_1^{(i)}$, $x_2^{(i)}$, and parameters b and w. You don't need to pick numbers for the parameters b and w, you're just looking to get a reasonable functional form. Your equation can involve as many w's as you need (w_1, w_2, ...). It may be conceptually easiest to start with $x_1^{(i)}$ and $x_2^{(i)}$ on different sides of the equals sign, and then rearrange to get an expression that is equal to 0.

(c) Write down a complete specification of a logistic regression model you might use to predict $y^{(i)}$ as a function of $x_1^{(i)}$ and $x_2^{(i)}$. This should include a probability distribution for $Y^{(i)}$ and any equations needed to calculate the probability that $Y^{(i)} = 1$ in terms of $x_1^{(i)}$ and $x_2^{(i)}$

(d) Suppose the training data set has the three observations in the table below. Write down the likelihood function. Your answer should involve only b, w's, and numbers from the table. You don't need to simplify it.

i	$x_1^{(i)}$	$x_2^{(i)}$	$y^{(i)}$
1	-0.5	0.25	1
2	0	0.75	1
3	0.75	-0.5	0