DS 100: Principles and Techniques of Data Science **Date: July 31, 2019**

Discussion #11

Name:

Logistic Regression

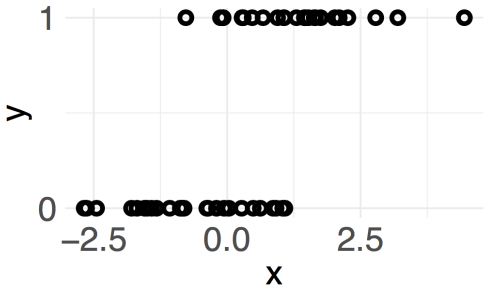
1. State whether the following claims are true or false. If false, provide a reason or correction.
   (a) A binary or multi-class classification technique should be used whenever there are categorical features.

   (b) A classifier that always predicts 0 has a test accuracy of 50% on all binary prediction tasks.

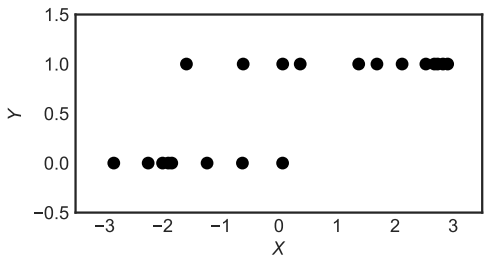
   (c) For a logistic regression model, all features are continuous, with values from 0 to 1.

   (d) In a setting with extreme class imbalance in which 95% of the training data have the same label, it is always possible to get at least 95% testing accuracy.

2. The next question refers to a binary classification problem with a single feature x. Based on the scatter plot of the data below, draw a reasonable approximation of the logistic regression probability estimates for $\mathbb{P}(Y = 1|x)$.



3. Suppose you are given the following dataset $\{(x_i, y_i)\}_{i=1}^n$ consisting of x and y pairs where the covariate $x_i \in \mathbb{R}$ and the response $y_i \in \{0, 1\}$.



Given this data, the value $\mathbb{P}(Y = 1|x = -1)$ is likely closest to:

☐ 0.95 ☐ 0.50 ☐ 0.05 ☐ -0.95

4. You have a classification data set, where x is some value and y is the label for that value:

x	y
2	1
3	0
0	1
1	0

Suppose that we're using a logistic regression model to predict the probability that $Y = 1$ given x:

$$\mathbb{P}(Y = 1|x) = \sigma(\mathbf{X}\theta)$$

(a) Suppose that $\mathbf{X} = \begin{bmatrix} 1 & x & x^2 \end{bmatrix}^T$ and our model parameters are $\theta^* = \begin{bmatrix} 1 & 0 & -2 \end{bmatrix}^T$. For the following parts, leave your answer as an expression (do not numerically evaluate $\log$, e, π, etc).

i. Compute $\hat{\mathbb{P}}(y = 1|x = 0)$.

ii. What is the loss for this single prediction $\hat{\mathbb{P}}(y = 1|x = 0)$, assuming we are using Cross Entropy as our loss function (or equivalently that we are using the cross entropy as our loss function)?

5. Suppose we train a binary classifier on some dataset. Suppose y is the set of true labels, and $\hat{y}$ is the set of predicted labels.

y	0	0	0	0	0	1	1	1	1	1
$\hat{y}$	0	1	1	1	1	1	1	0	0	0

Determine each of the following quantities.

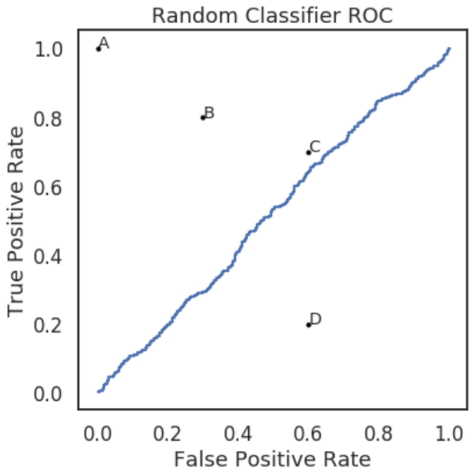
(a) The number of true positives

(b) The number of false negatives

(c) The precision of our classifier. Write your answer as a simplified fraction.

ROC Curves

6. State whether the following claims are true or false. If false, provide a reason or correction.



(a) Point A (0, 1) represents our ideal classifier.

(b) Point C performs well compared to a classifier that guesses each class randomly.

(c) The classifier at Point B performs better than the one at Point D.