Michigan State University **Feb 1st, 2023**
CMSE381 - Data Science **Dr. Xie**

CMSE381 - Quiz 3

I will adhere to the Spartan Code of Honor in completing this assignment.

Signed: _____ Print Name: _____

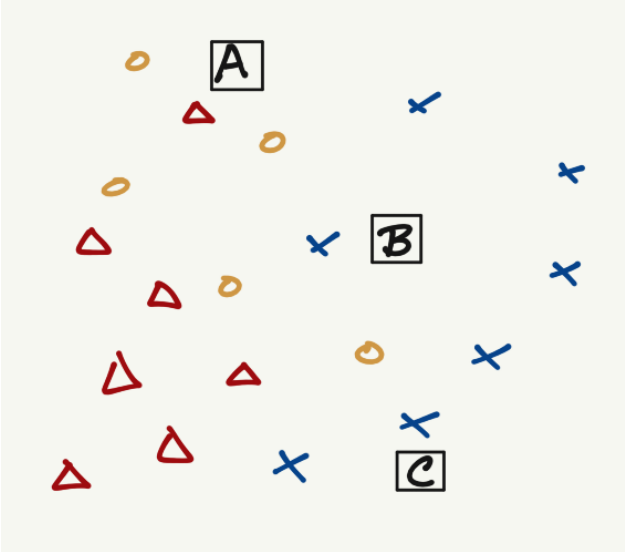
1. I'm building a model to predict whether a student can publish a journal article before graduation using his/her GPA of the first three years. I collected 100 MSU alumnus' records and fitted a logistic regression where $Y = 1$ means the student published a paper and $Y = 0$ otherwise. The outcome is $\beta_0 = 0.5$ and $\beta_1 = 0.003$. Now given a student with average first three year GPA of 3.6, what is the probability he/she will publish a journal article before graduation (you don't need to calculate the exact number. Show the formula is enough)?

$$P(Y=1) = \frac{e^{0.5+0.003\times 3.6}}{e^{0.5+0.003\times 3.6}+1}$$

2. For Q1, what is the best classifier (oracle)? You need to write down the exact form.

$C(x) = j$ s.t. $P(Y=j)$ is the maximum among $\{P(Y=1), P(Y=0)\}$

3. I am running a classification problem where data points as drawn below have coordinates X_1 and X_2 drawn in the plane, and we are predicting labels consisting of circle, X, or triangle. If I use KNN with $K = 1$, what will the predictions be for the new data points A, B, and C?



A: △

B: X

C: X