**Michigan State University** **Feb 1, 2024**
**CMSE381 - Data Science**

---

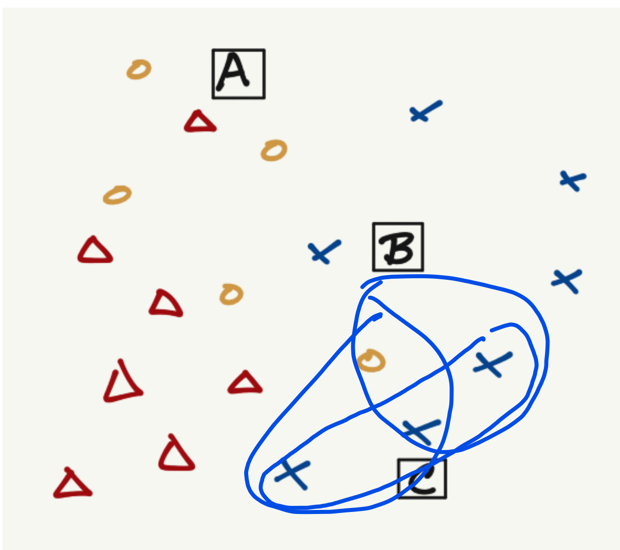
# CMSE381 - Quiz 3

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

Signed: ______________________________ Print Name: ______________________________

1. (3pts) On what type of problems is Logistic regression more suitable than linear regression ?

2. (4pts) 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 $K = 3$, what will the predictions be for the new data point $C$? Note the predictions should be one of {circle, cross, or triangle}.




$$\hat{\beta}_0, \hat{\beta}_1 = \underset{\beta_0, \beta_1}{\arg\max} \frac{e^{\beta_0+2000\beta_1}}{1+e^{\beta_0+2000\beta_1}} \cdot \left(1 - \frac{e^{\beta_0+300\beta_1}}{1+}\right) \cdot \left(1 - \frac{e^{\beta_0+100\beta_1}}{1+e}\right)$$

3. (3pts) I want to use simple logistic regression to predict the default status from credit card balance. The first step is to find $\beta_0$ and $\beta_1$ in the model $p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$ using the training data shown in the table. What is the objective function you want to maximize in order to find $\hat{\beta}_0$ and $\hat{\beta}_1$?

| Balance | Prediction |
|---|---|
| 100 | No |
| 300 | No |
| 2000 | Yes |

$$\prod_{i \mid y_i=1} p(x_i) \prod_{i' \mid y_{i'}=0} (1 - p(x_{i'}))$$