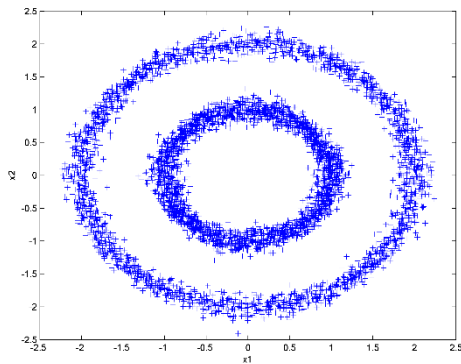Name:

DATA SCIENCE – FINAL EXAM

1. Perform clustering on the two-dimensional data illustrated below.
   a. Explain how the following four algorithms would split the data into **two clusters**: **K-means**, **hierarchical clustering** (separately **single linkage** and **complete linkage**), **DBSCAN**. Draw the clusters and give brief explanations to your answers!
   b. Let us assume that the goal is to find the two annular natural clusters. Give an $R^2 \rightarrow R^2$ **coordinate transformation** that assists the bad-performing clustering algorithms to find the natural clusters. Plot the coordinate system transformation!

   (20%)

2. kNN regression
   a. Determine the target value of the record marked by a star using **kNN regression** with the choice of $k = 4$ (without using distance-weights)!
   b. Name at least **three further algorithms** that can solve **regression** problems!

   (10%)
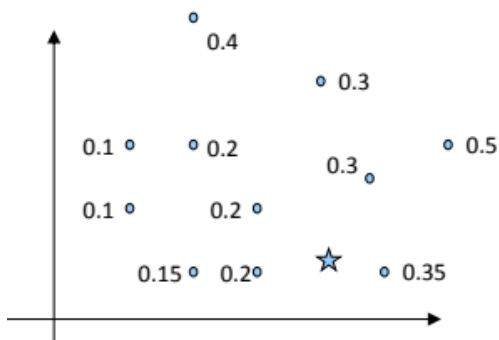
3. Are the following statements **true or false**? **Explain** your answer!
   (15%)

   a. Using AdaBoost algorithm, if the jth classifier correctly classifies the ith record then the weight of the ith record will be certainly reduced in the $(j+1)$th step.

   b. Perceptron algorithm is a universal function approximator.

   c. One-vs-one strategy is computationally more expensive than the one-vs-rest strategy.

   d. SMOTE is an efficient undersampling method.

   e. Logistic regression is a binary classification algorithm.

4. Represent the "(**NOT** A) **AND** B" Boolean function with a perceptron or show that it is not possible to do so. In the latter case, construct a neural network with one hidden layer. Use the usual activation function!

   (10%)

5. Hierarchical clustering
    a. Using the following **distance** matrix draw the dendrograms corresponding to the single linkage (MIN) and complete linkage (MAX) clustering algorithms.
    b. What advantages does hierarchical clustering have in general compared to K-means algorithm?
    c. What advantages and disadvantages do single and complete linkage techniques have compared to one another?

    (20%)

Item	A	B	C	D	E
A	0	1	2	2	3
B	1	0	2	4	3
C	2	2	0	1	5
D	2	4	1	0	3
E	3	3	5	3	0

6. We aim to minimize the squared error with regularization term regarding a linear regression model:

$$((w_0 + \mathbf{w}^T \mathbf{x}) - y)^2 + \lambda ||\mathbf{w}||_2^2,$$

where $\mathbf{w} = (w_1, w_2, ..., w_p)$: vector of weights (parameters, coefficients), $\mathbf{x}$: feature vector, λ: regularization parameter, $||\mathbf{w}||_2$ denotes the Euclidean length of the vector $\mathbf{w}$. Note that the constant term (w_0) is not regularized.

    a. What is the role of regularization?

    b. Give the formula for one update step of stochastic gradient descent method for w_0 and for w_i $(i \neq 0)$, with η learning rate for a training record $(x_1, x_2, ..., x_p)$ with target variable y.

    c. What does the learning rate refer to? What are the advantages and disadvantages of using a small/large learning rate?

    d. What is the difference between gradient descent method and stochastic gradient descent method?

(25%)