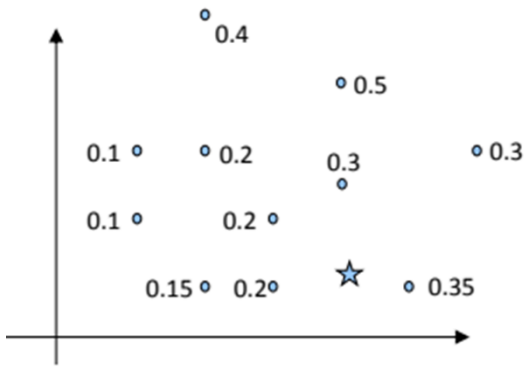
DATA SCIENCE 2. – SAMPLE FINAL EXAM

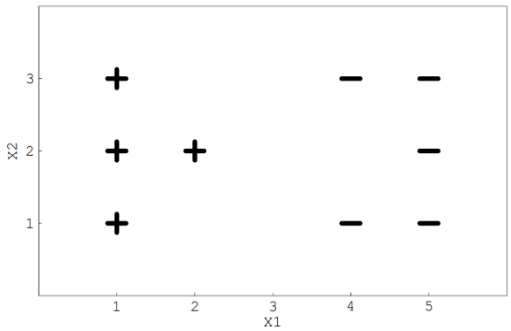
1. Determine the target value of the record marked by a star using…
   a. ***k*NN regression** with the choice of $k = 4$ (without using distance-weights).
   b. **a decision tree** with maximum number of leaves set to 3. (Also sketch the splits!)

   (10%)



2. Suppose we are using a (linear) **SVM**, with some large C value, and are given the following data set.
   a. Draw and give the equation of the **decision boundary** of linear SVM. Give a brief explanation.
   b. Which records are **support vectors** and why? Circle them!

   (10%)



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

   a. The AdaBoost algorithm can be easily parallelized since its weak models are independent parallel models.

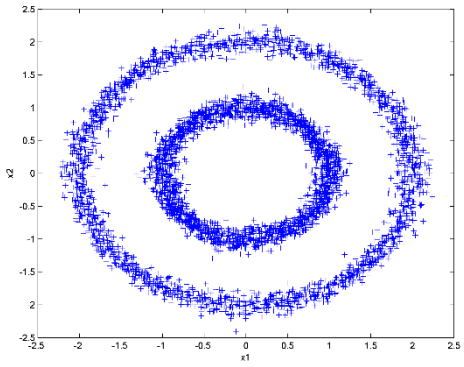
   b. One-vs-one strategy is computationally more expensive than the one-vs-rest strategy.

   c. In bagging, we choose random subsamples of the input points with replacement.

   d. Logistic regression models always output values between 0 and 1.

   e. Convolutional neural networks have more parameters than fully connected networks with the same number of layers and the same numbers of neurons in each layer.

4. 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!



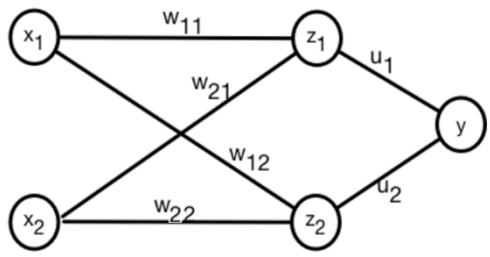
(20%)

5. Based on the following **distance** matrix draw the dendrograms corresponding to the **hierarchical clustering** algorithms using…
   a. the **single linkage (MIN)** method!
   b. the **complete linkage (MAX)** method!

(10%)

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. Consider the **neural network** with one hidden layer shown in the figure below. We aim to solve a **regression** problem on a two-dimensional input (represented by two features x_1, x_2). The **activation function** applied at the two nodes in the hidden layer is $tanh(\theta)$, the activation function at the output node is the identity. The output y is compared against the target output t to minimize the **squared loss** $E^2 = (y - t)^2$.



   a. Express the output y in terms of $x_1, x_2, w_{11}, w_{12}, w_{21}, w_{22}, u_1, u_2$.
   b. Write $\frac{\partial E^2}{\partial w_{11}}$ using the chain rule and the following hint: $tanh'(\theta) = 1 - tanh^2(\theta)$.
   c. Derive the **stochastic gradient descent update rule** for w_{11}. Let λ be the learning rate.
   d. What is the role of **learning rate** in general? What are the advantages and disadvantages of using a small/large learning rate?
   e. What is the **difference** between gradient descent method and stochastic gradient descent method in general?
   f. State at least two possible approaches that you can use to reduce the **overfitting** in a neural network.

(35%)