# Compsci 571 HW6
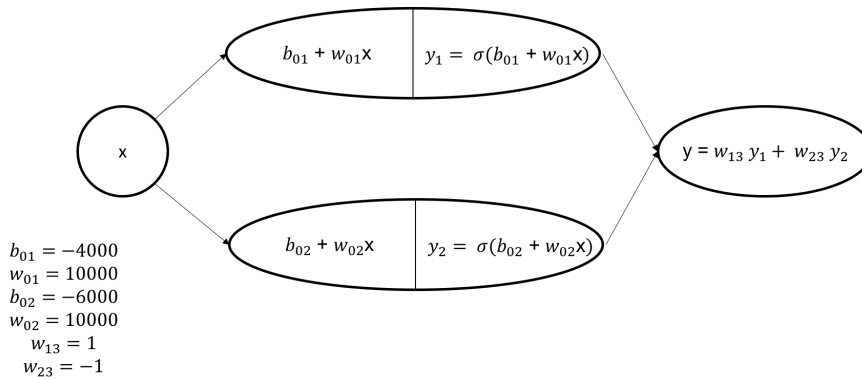
Yilin Gao (yg95)

April 5, 2018
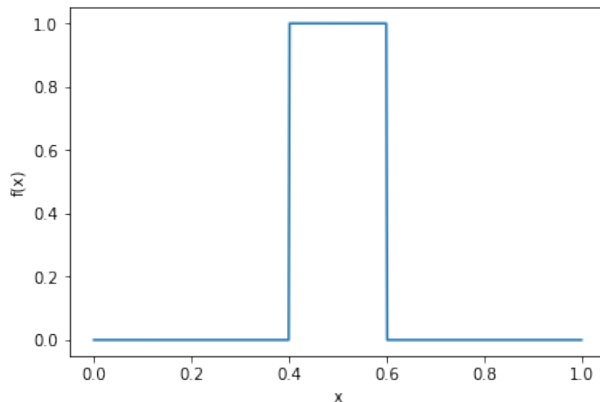
## 1 Neural Networks and Universal Approximation Theorem

### 1.1

(a) The NN architecture is like following:



$b_{01} = -4000$
$w_{01} = 10000$
$b_{02} = -6000$
$w_{02} = 10000$
$w_{13} = 1$
$w_{23} = -1$

The implementation is in `q1.ipynb`. The approximated function is as following:



The minimal number of hidden neurons is 2, because the bump is a combination of 2 step functions, and each neuron (with the sigmoid activation function) is able to approximate one step function with any given step direction, location and height.

(b) In the NN, $w_{01}$ determines the steepness of the step-up part of the bump, $w_{02}$ determines the steepness of the step-down part of the bump. $-\frac{b_{01}}{w_{01}}$ determines the step-up location, $-\frac{b_{02}}{w_{02}}$ determines the step-down location. And $w_{13}$ and $w_{23}$ determine the height of the bump.
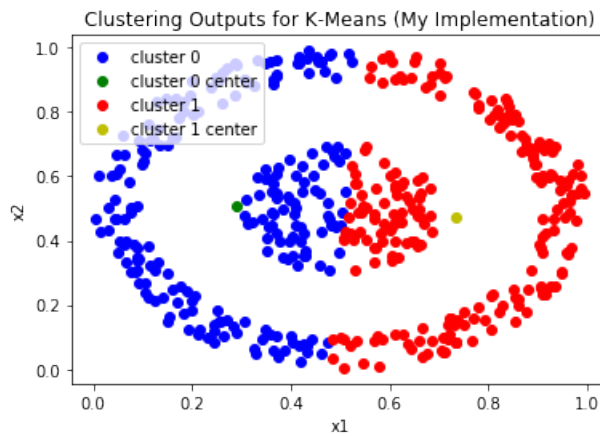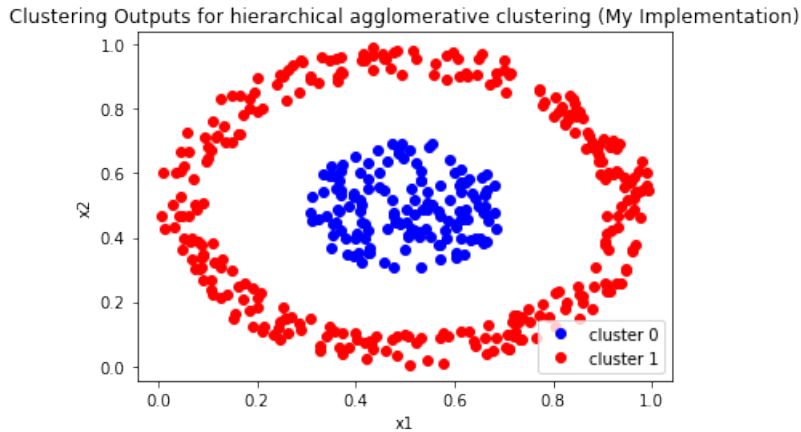
### 1.2

(a)

(b)

(c)

# 2  EM

# 3  Clustering

(a) See the implementation in `q3.ipynb`.

(b) See the implementation in `q3.ipynb`.

(c) The empirical clustering results on the dataset with both algorithms are as following:

K-Means:



Hierarchical Agglomerative Clustering:



From the visualization, we can tell the hierarchical agglomerative clustering algorithm performs better.

On this specific dataset, the possible reason for the discrepancy is that K-Means assumes the variance of each X variable is spherical. However, the given data doesn't satisfy this assumption.

(d) Possible preprocessing on data to make K-Means perform better: normalize each X variable to mean = 0 and variance = 1.