

# Introduction to Machine Learning

## Assignment 4

Group 31

Stijn Kammer (s4986296) & Ramon Kits (s5440769)

October 12, 2022

### 1. INTRODUCTION

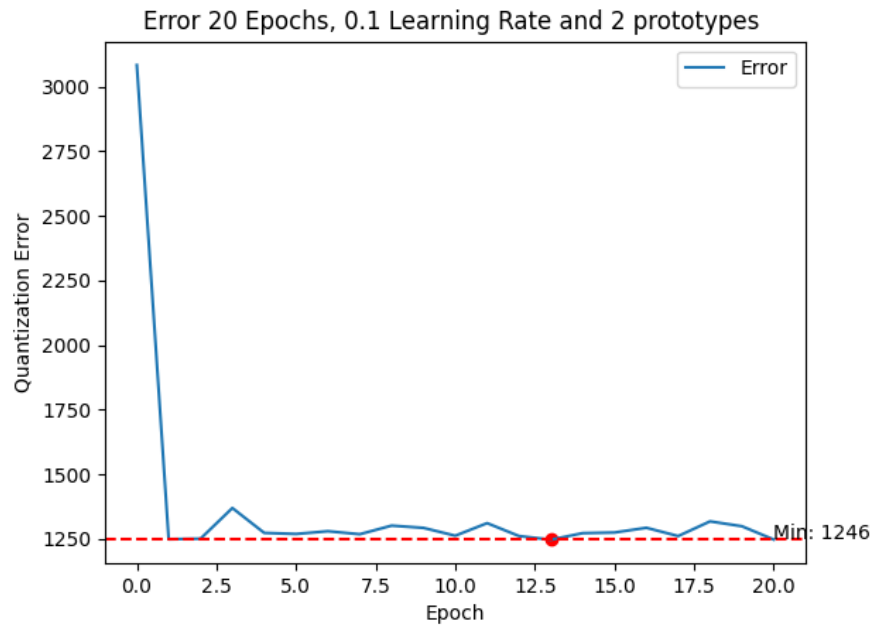
The method used in this assignment is Winner-Takes-All unsupervised competitive learning Vector Quantization. The use of this method is to classify the data into clusters and to find the center of each cluster. These centers are used to classify the data into the clusters making it possible to classify new data. The Winner-Takes-All refers that one prototype is chosen to represent the cluster and the other prototypes are pushed away from the cluster into the other clusters. In this assignment we use the `simplevqdata` dataset to test the method. The dataset contains 1000 data points with 2 features.

### 2. METHODS

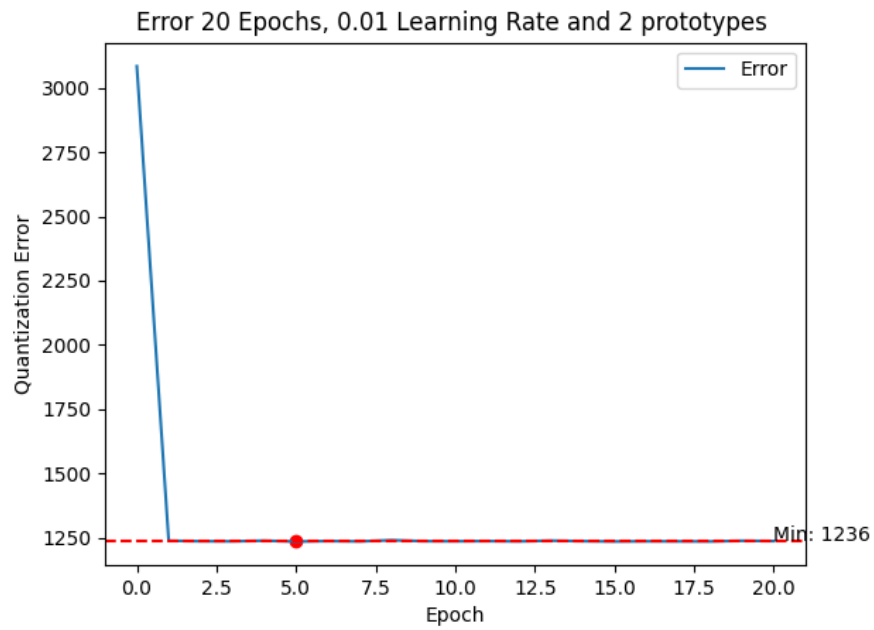
Vector Quantization makes use of prototype vectors to represent the data and is often used for identification and grouping of clusters in similar data. To use Vector Quantization, we first need to define a set of prototype vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ . The prototype vectors are chosen by selecting  $k$  data points from the data set. After which present a single example  $\mathbf{x}$ , we can find the closest prototype vector  $\mathbf{v}_i$  by minimizing the distance between  $\mathbf{x}$  and  $\mathbf{v}_i$ . The distance between  $\mathbf{x}$  and  $\mathbf{v}_i$  is defined as the Euclidean distance between the two vectors. When the closest prototype vector  $\mathbf{v}_i$  is found, we move the prototype vector  $\mathbf{v}_i$  towards  $\mathbf{x}$  by a fraction  $\eta$  of the distance between  $\mathbf{x}$  and  $\mathbf{v}_i$ . Repeat this process for all the examples in the training set and the prototype vectors will converge to a set of vectors that represent the data well. The prototype vectors are then used to classify new examples.

Quantization error is a number that is calculated when using the prototype vectors to classify new examples. To be able to estimate a good value for  $k$ , we can use the quantization error to find the optimal value for  $k$ , the error should be as low as possible. The quantization error is defined as the sum of distances between the prototype vectors and the examples in the training set. The quantization error is minimized by moving the prototype vectors towards the examples in the training set. The quantization error is also minimized by choosing a good set of prototype vectors. The prototype vectors should be chosen in a way that they are spread out over the data set. This is done by choosing the prototype vectors to be the  $k$  data points that are furthest apart from each other. The prototype vectors should also be chosen in a way that they are representative of the data set. This is done by choosing the prototype vectors to be the  $k$  data points that are closest to the mean of the data set.

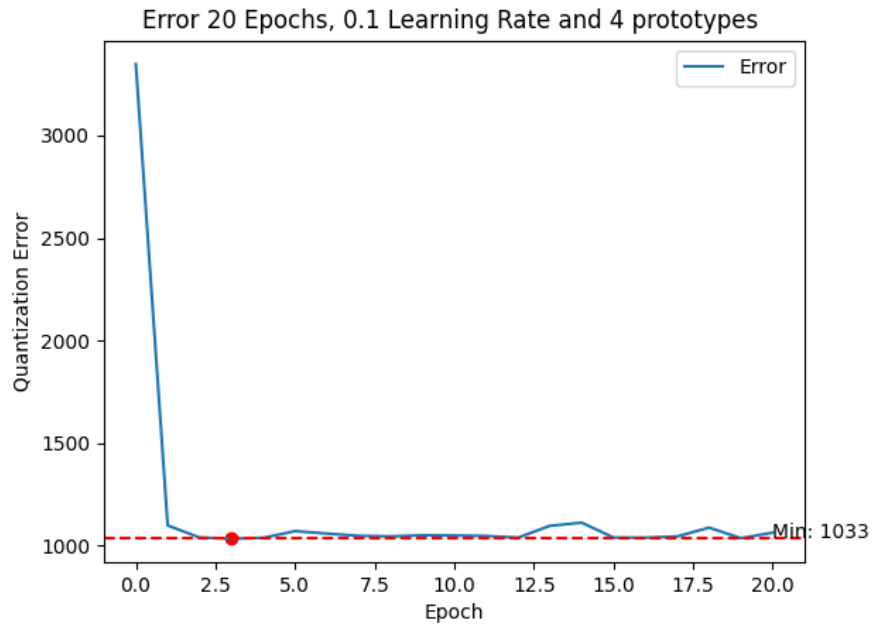
### 3. LEARNING CURVES



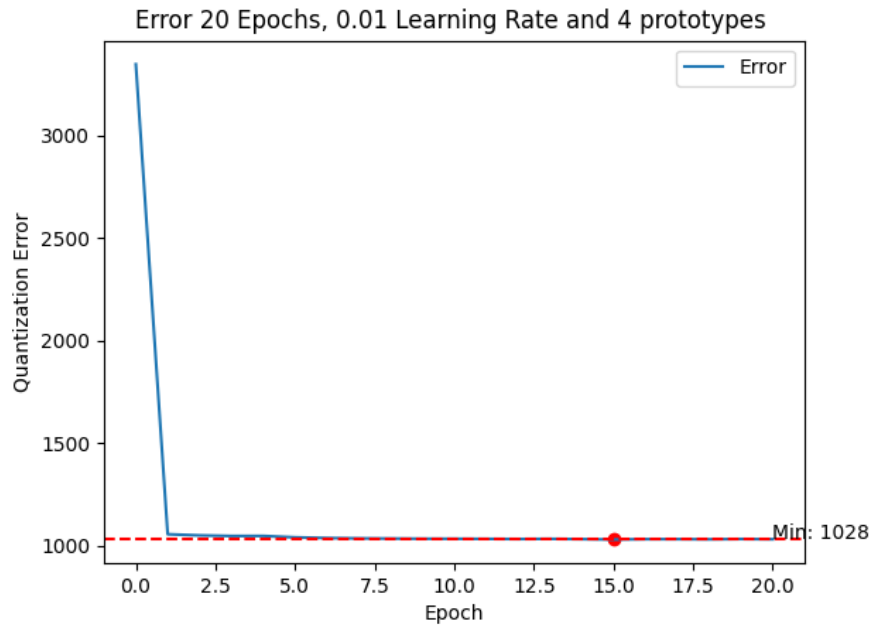
**Figure 1:** Error for  $t_{max} = 20$ ,  $\eta = 0.1$  and  $K = 2$



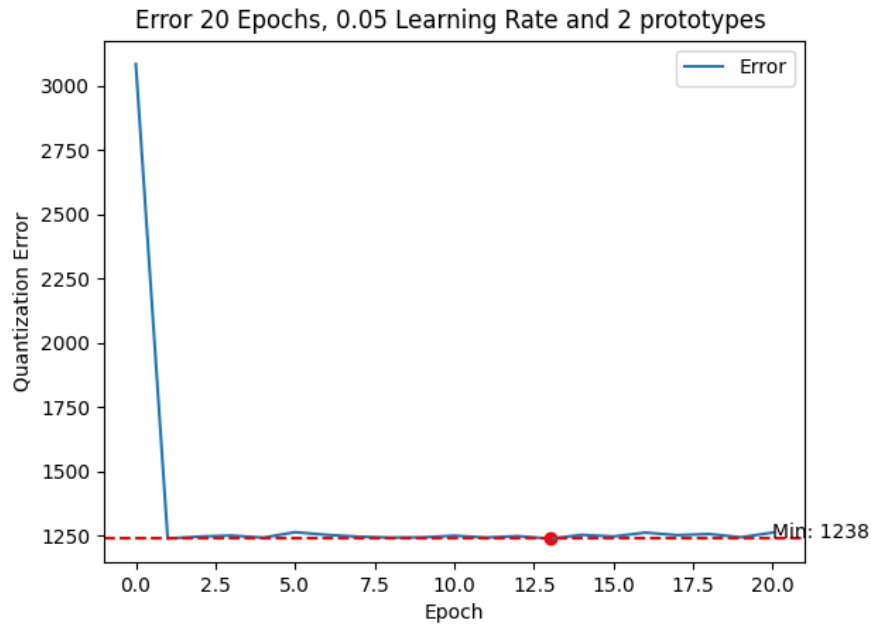
**Figure 2:** Error for  $t_{max} = 20$ ,  $\eta = 0.01$  and  $K = 2$



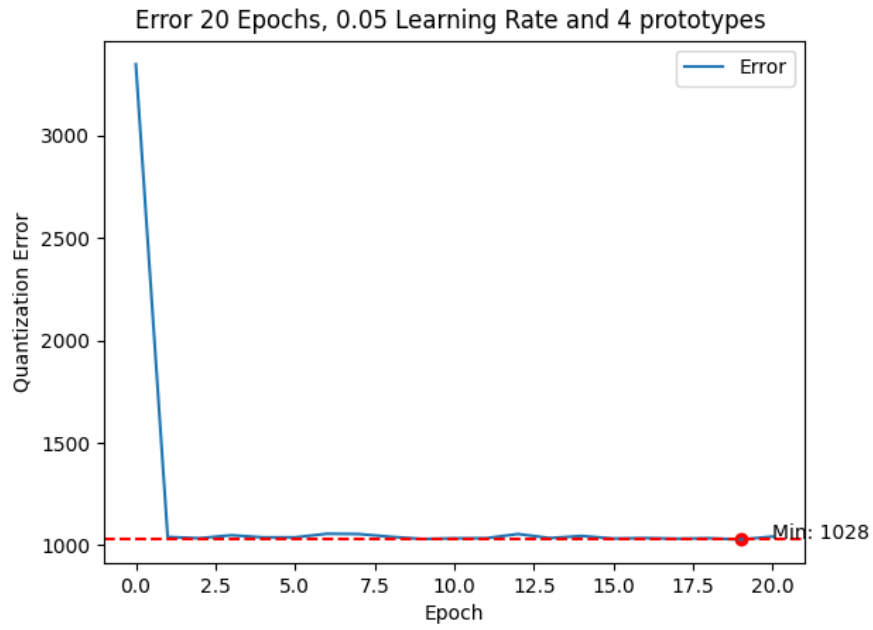
**Figure 3:** Error for  $t_{max} = 20$ ,  $\eta = 0.1$  and  $K = 4$



**Figure 4:** Error for  $t_{max} = 20$ ,  $\eta = 0.01$  and  $K = 4$

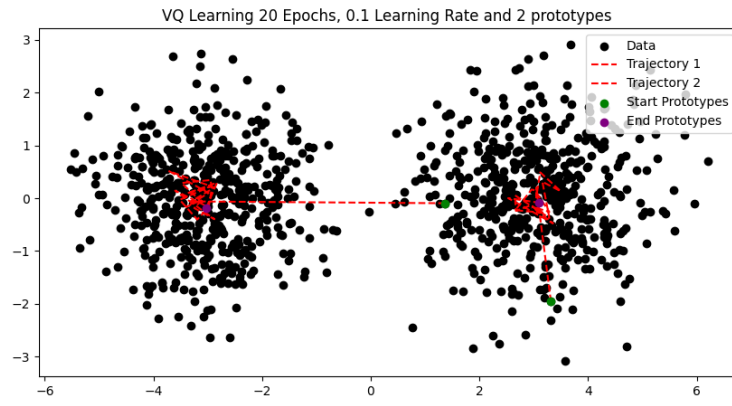


**Figure 5:** Error for  $t_{max} = 20$ ,  $\eta = 0.05$  and  $K = 2$

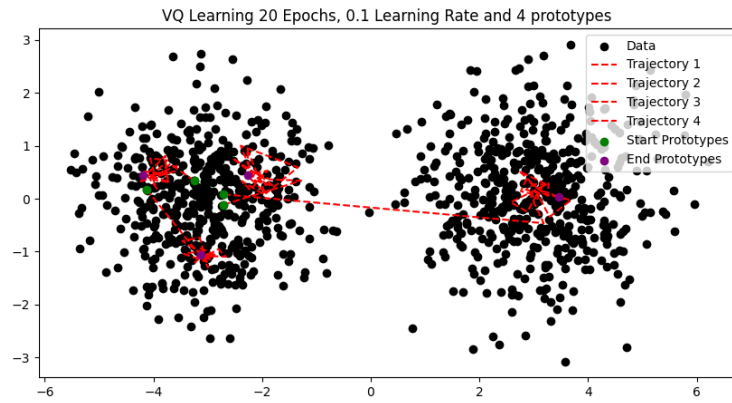


**Figure 6:** Error for  $t_{max} = 20$ ,  $\eta = 0.05$  and  $K = 4$

#### 4. TRAJECTORIES OF PROTOTYPES

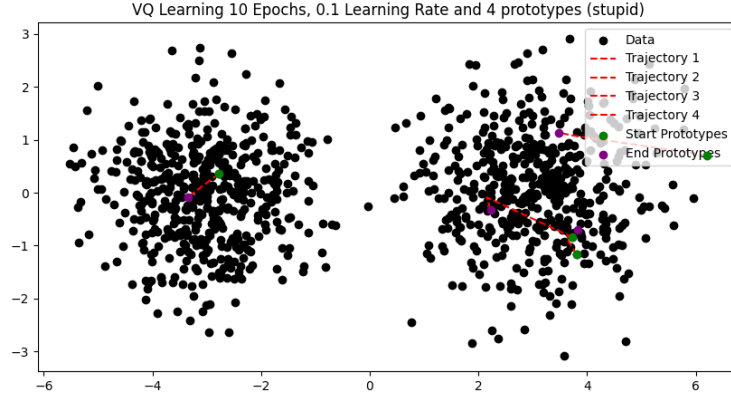


**Figure 7:** Learning curve for Vector Quantization with  $t_{max} = 20$ ,  $\eta = 0.1$  and  $K = 2$



**Figure 8:** Learning curve for Vector Quantization with  $t_{max} = 20$ ,  $\eta = 0.1$  and  $K = 4$

## 5. "STUPID" TRAJECTORIES OF PROTOTYPES



**Figure 9:** Learning curve for Vector Quantization with  $t_{max} = 10$ ,  $\eta = 0.1$  and  $K = 4$

## 6. DISCUSSION

In this report we have implemented Vector Quantization and used it to classify data. We have also used learning curves to find the optimal values for  $\eta$  and  $K$ . We have also used learning curves to find the optimal value for  $t_{max}$ .

In Figure 7 we can see that the trajectory of the prototypes move towards the data points. This is because the prototypes are moved towards the data points by a fraction  $\eta$  of the distance between the prototype and the data point. The prototypes are moved towards the data points because the quantization error is minimized by moving the prototypes towards the data points.

In Figure 8 we can see that the trajectory of the prototypes move towards the data points. This is because the prototypes are moved towards the data points by a fraction  $\eta$  of the distance between the prototype and the data point. The prototypes are moved towards the data points because the quantization error is minimized by moving the prototypes towards the data points. One thing that is different is that three of the prototypes stay close to each other while the other prototype moves towards the data points.

## 7. BONUS