Advanced Techniques in Machine Learning
Ex 1

Eyal Orbach - ID 015369317
Daniel Juravski - ID 206082323

Below is a table showing the 'correct prediction' percentage on the validation data (20% of the train data that was not trained on)

distance function\comparison type	One vs All	All Pairs	Our own (exhaustive)
Hamming	92.99%	96.90%	95.11%
Loss	95.94%	96.72%	96.12%

Our own matrix:
We Implemented an exhaustive code matrix as follows:

$$\begin{matrix} class0 = \\ class1 = \\ class2 = \\ class3 = \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

classifiers fx1 to fx7 were trained to predict the values correlating to this matrix.

One vs All:
Here we have noticed the most significant difference between Hamming and loss distance. This can be explained because the loss distance outperforms the Hamming distance in these 2 use cases:
-) The relevant vector did not classify correctly the item, but the other vectors have strongly classified it as not belonging to their class. In this case Hamming will have a tie of all values equal to 1, resulting in predicting class 0 due to our tie breaking rule.
Where as the loss distance will recognize the significance of the other vectors negative values.
-) More than one vector classified with a positive value but not with a large degree of confidence (a value smaller than 1). the Hamming distance again will tie between those classes where the loss distance can determine by the level of confidence (values above 1 will still tie though)

All pairs:
Here the results were not conclusive. Hamming beat the loss in this documented run but vice versa was also noticed on similar runs.
This could be explained due to the larger granularity of comparison, meaning the indiction from more classifiers, makes the Hamming distance more reliable than it was in the OA case.

Role of distance in rows
The bigger the distance between the rows in the m matrix, the bigger the distinction the prediction logic can make. The m matrix determines the relation between the different classifiers, and a large distance between the rows means that the combination of the classifiers is more distinctive towards each class. Therefore using such matrices can help us reach a better separation of the data into the various classes.

conclusions:

- The All Pairs produced the best results, with either Hamming or Loss distance
- The biggest difference is in 'One vs All' where loss distance gives much better results than Hamming distance.
- Loss distance dominates Hamming on most experiments.
- Although the experiment here reached the best prediction percentage with the AP method, a different data set might have a better classification with the exhaustive matrix method.