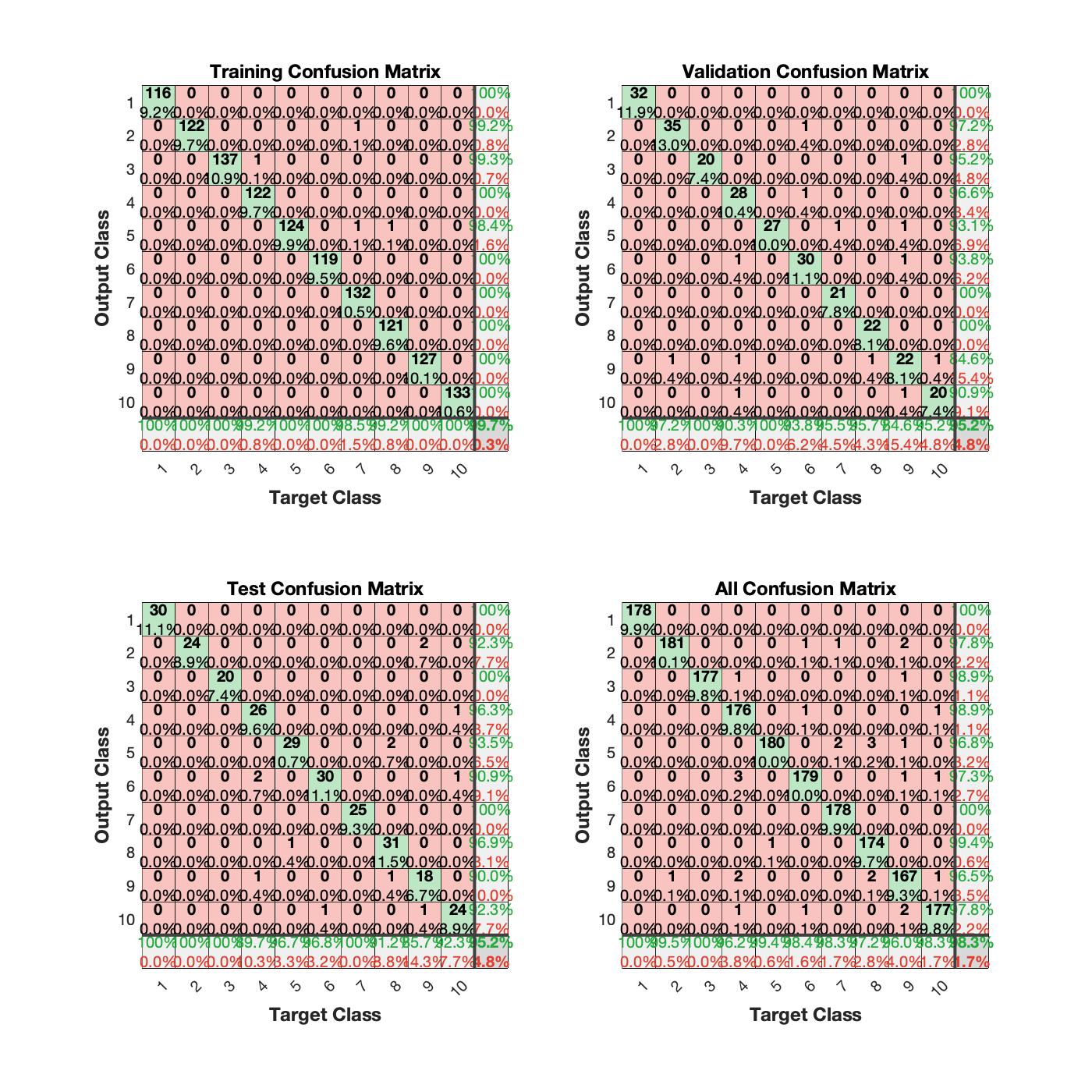
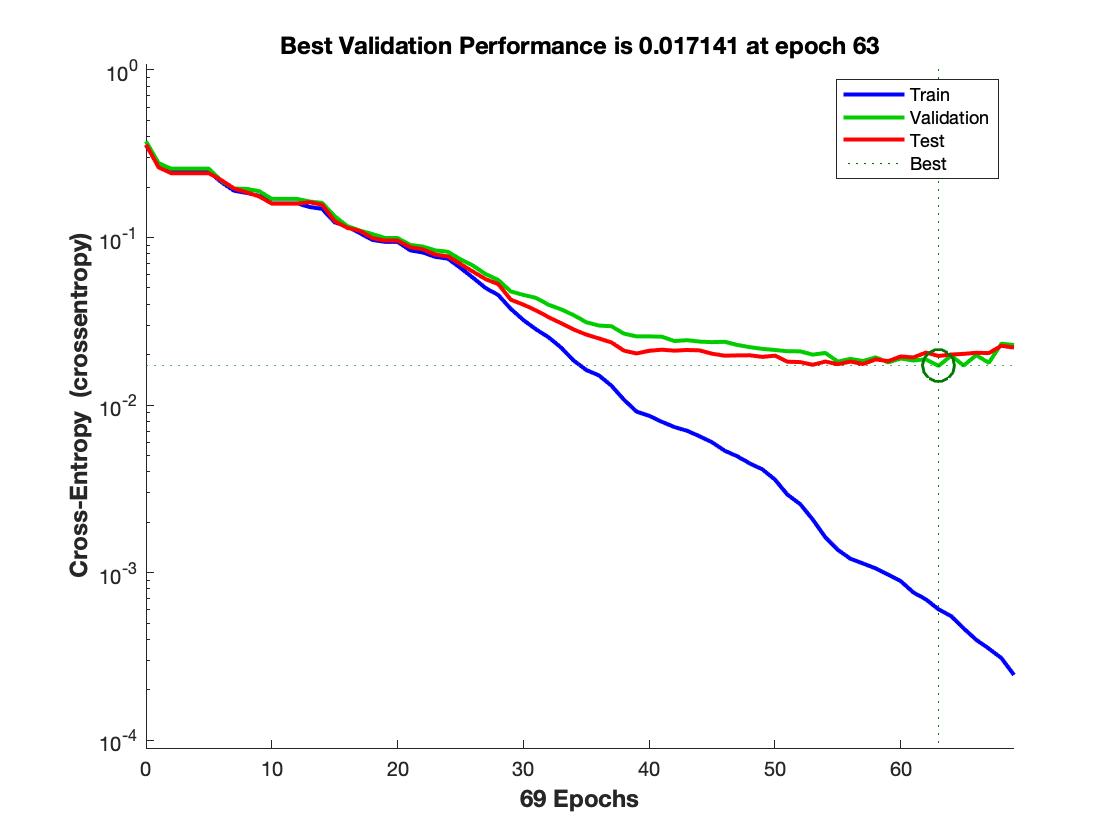
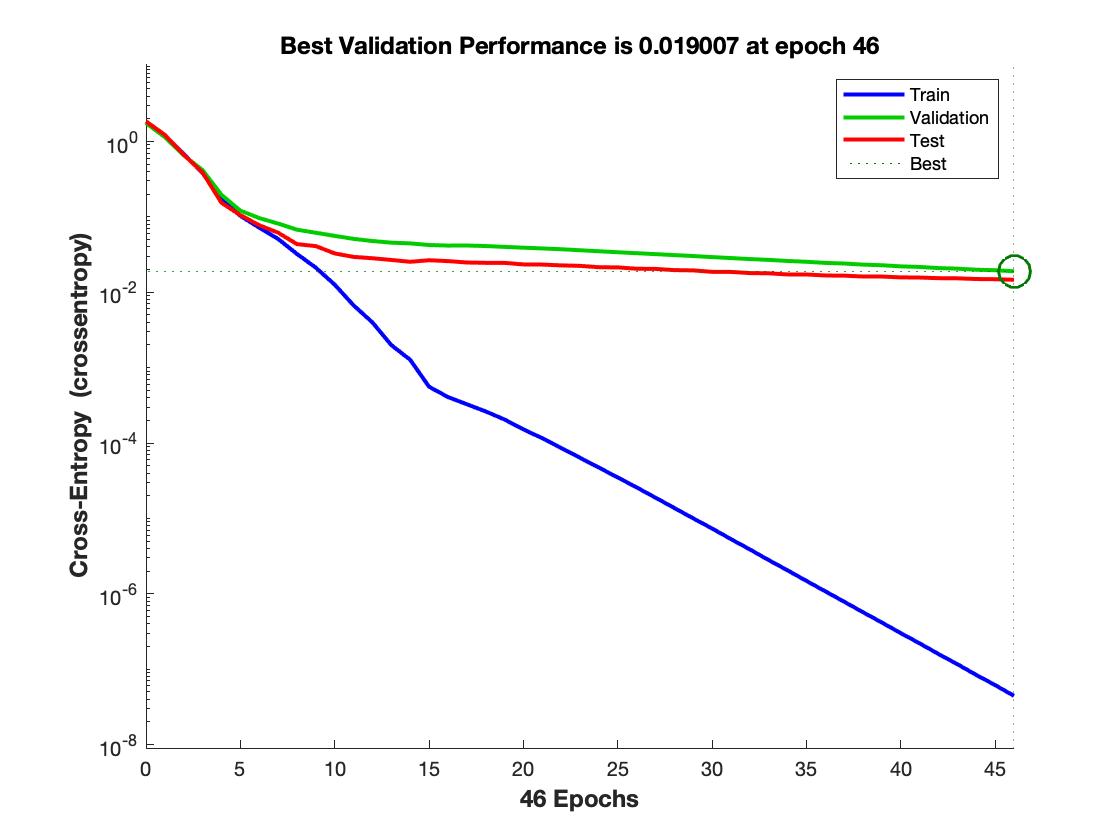
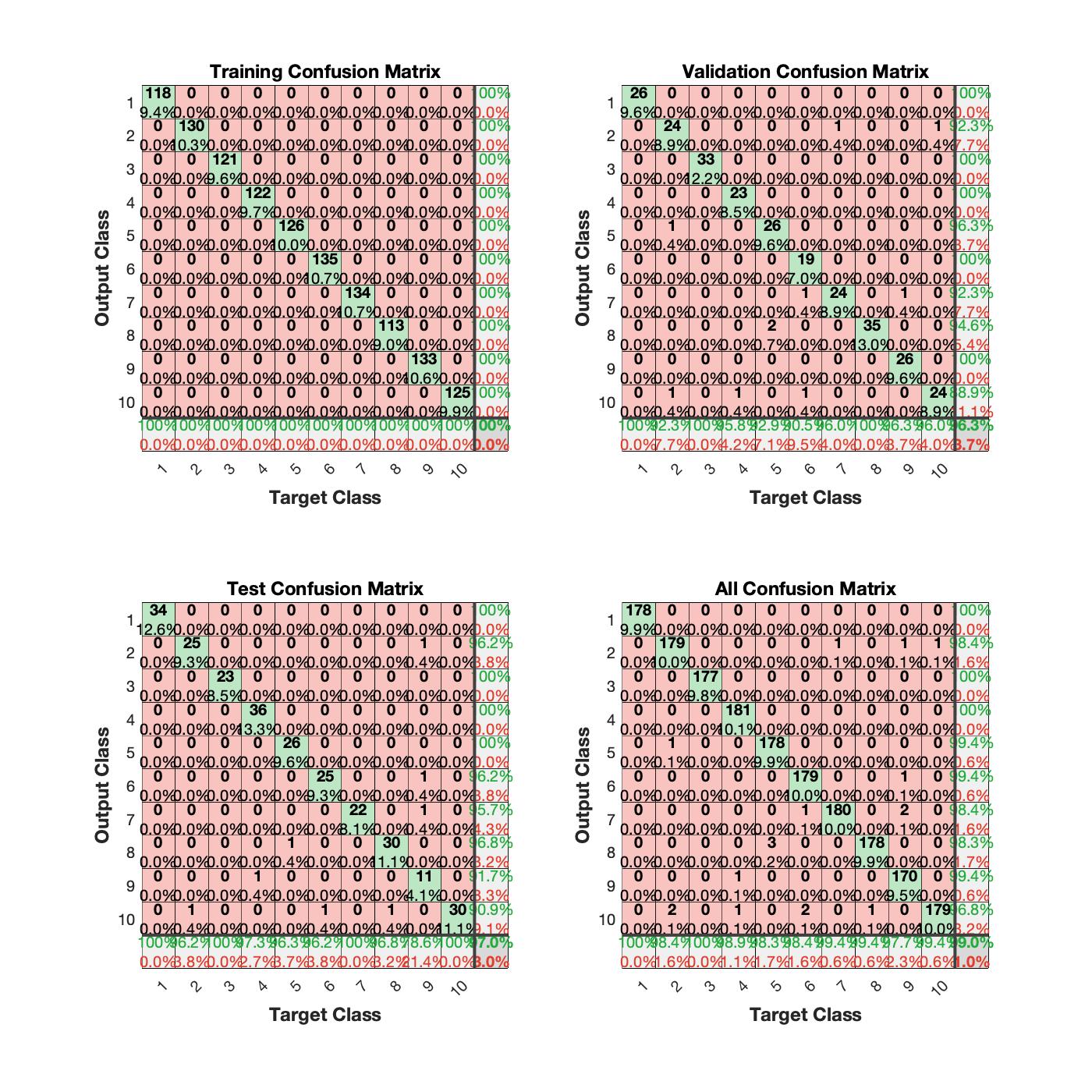
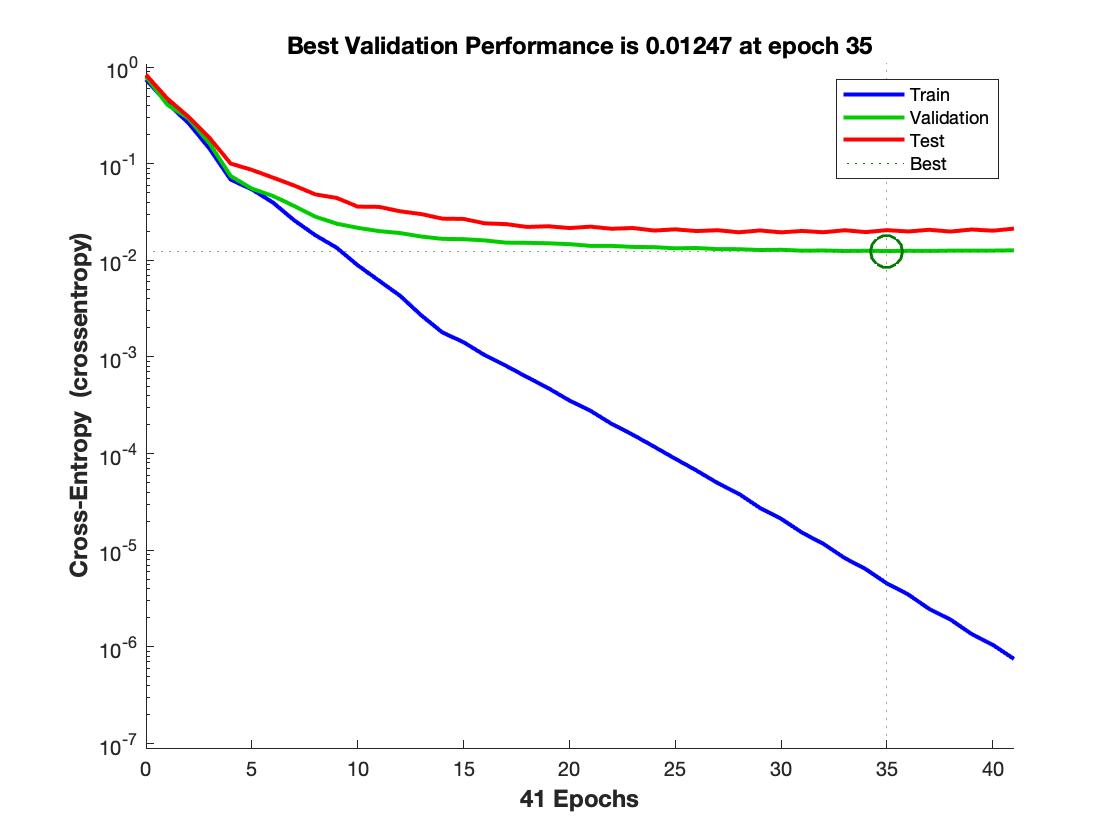
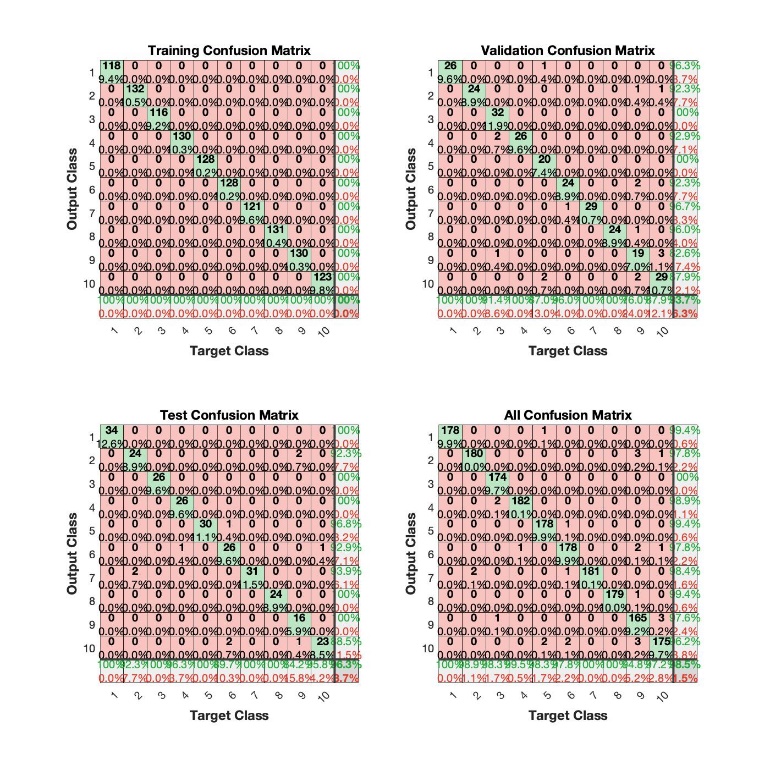
# For 10 Nodes



# For 100 Nodes



# For 500 Nodes



Bonus: (10 pts)

Using 1024 features for neural network is very cumbersome. Is there a way to reduce the number of features? If you have an idea, show me how effective it is.

I will use Principal component analysis (PCA) to reduce features.

* Principal Component Analysis (PCA) is an unsupervised linear transformation technique that is widely used across different fields, most prominently for feature extraction and dimensionality reduction. Other popular applications of PCA include exploratory data analyses and de-noising of signals in stock market trading, and the analysis of genome data and gene expression levels in the field of bioinformatics.
* PCA helps us to identify patterns in data based on the correlation between features. In a nutshell, PCA aims to find the directions of maximum variance in high-dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original one.
* The orthogonal axes (principal components) of the new subspace can be interpreted as the directions of maximum variance given the constraint that the new feature axes are orthogonal to each other.
* If we use PCA for dimensionality reduction, we construct a d x k–dimensional transformation matrix W that allows us to map a sample vector x onto a new k–dimensional feature subspace that has fewer dimensions than the original d–dimensional feature space.
* As a result of transforming the original d-dimensional data onto this new k-dimensional subspace (typically k ≪ d), the first principal component will have the largest possible variance, and all consequent principal components will have the largest variance given the constraint that these components are uncorrelated (orthogonal) to the other principal components — even if the input features are correlated, the resulting principal components will be mutually orthogonal (uncorrelated).

Note that the PCA directions are highly sensitive to data scaling, and we need to standardize the features prior to PCA if the features were measured on different scales and we want to assign equal importance to all features.

References:

<https://towardsdatascience.com/principal-component-analysis-for-dimensionality-reduction-115a3d157bad>

<https://stats.stackexchange.com/questions/27300/using-principal-component-analysis-pca-for-feature-selection>