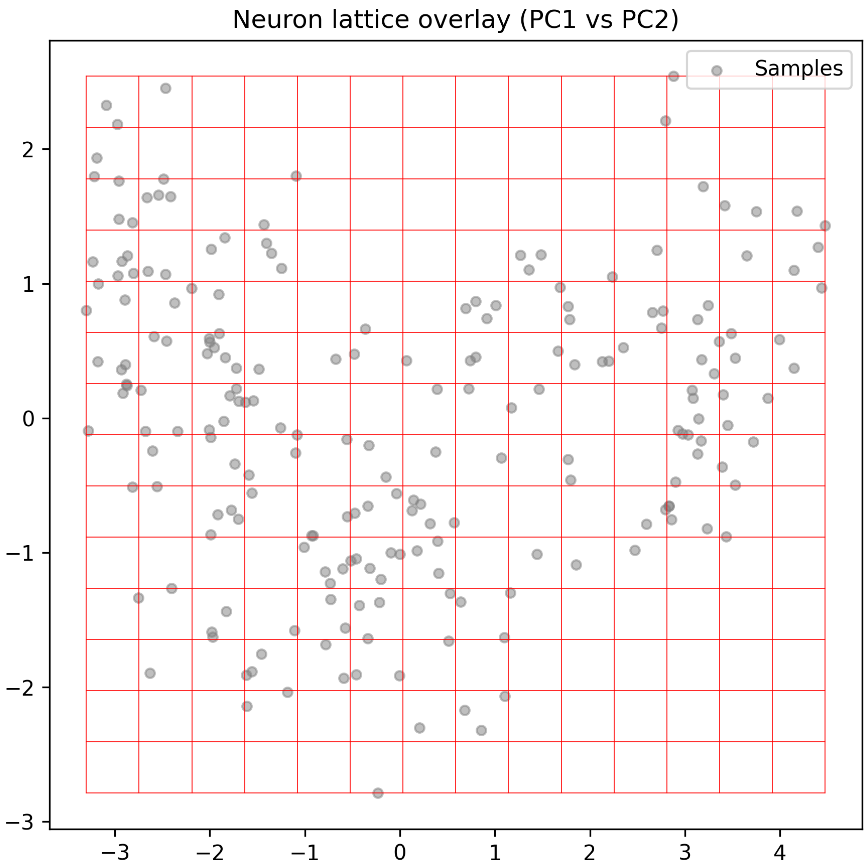
Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics
Department of Applied Informatics

Neural Networks

Project 2b – Visualizing Data

Zulfiya Usmonova



Self-organising map of the Seeds dataset: the trained 15×15 *neuron lattice (red) projected onto the first two principal components*

Overview

In this project, we try to implement a self-organizing map with rectangular topology and use it to visualize the Seeds data set from the UCI Machine Learning Repository in two dimensions. The dataset consists of 210 samples with 7 features and a single class label column.

This project aims to implement (design and fine-tune) a SOM that minimises quantisation error on the provided dataset, comparing several grid sizes, distance metrics, neighbourhood functions, and learning-rate schedules; and to produce a set of figures illustrating training dynamics and the final internal representations.

The report consists of:

- Pre-processing and hyper-parameter choices;
- Resulting maps and training curves;

1 Pre-processing and Hyper-Parameter Choices

The raw file seeds.txt was read into a pandas DataFrame; the first seven columns form the feature matrix X, and the eighth column is the class label y $\in \{1, 2, 3\}$.

All coordinates were centred and rescaled feature-wise:

$$\tilde{x} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}}$$

using the mean μ_{train} and standard deviation σ_{train} computed on the 210 set only.

A PCA on $\mathbf{X}_z$ yielded eigen-values $\lambda_1 = 5.06$ and $\lambda_2 = 1.20$ (ratio ≈ 4.2). Consequently, we chose elongated SOM grids whose longer side is roughly 1.5-2 times the shorter side, instead of a square layout.

The hyperparameter grid values are shown in the following table:

Table 1: Hyper–parameter search space explored in this study.

Parameter	**Values explored**
Candidate grid sizes	$\mathbf{18 \times 12},\ \mathbf{21 \times 9},\ \mathbf{15 \times 15}\}$
Grid distance metrics	$\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{L}_\infty$
Neighbourhood shape	Gaussian *vs.* Discrete
Initial radius λ_0	$0.5\times$ map diagonal
Final radius λ_f	Gaussian: $\{0.5, 0.3, 10^{-6}\}$; Discrete: $\{1, 2\}$
Learning–rate $\alpha(t)$	Pairs $\alpha_0 \to \alpha_\text{f}$ in $\{(0.5, 0.02), (0.5, 0.01)\}$
Epochs	150

Here are the hyper-parameters of the best model:

Table 2: Chosen hyper-parameters of the best-performing SOM configuration.

Hyper-parameter	**Value**
Grid size	$\mathbf{15 \times 15}$
Distance metric	L_1
Neighbourhood function	Gaussian
Learning-rate schedule	$\alpha_0 = 0.50 \;\rightarrow\; \alpha_\mathrm{f} = 0.02$
Radius schedule	$\lambda_0 = 0.5\times$ diagonal $\rightarrow\; \lambda_\mathrm{f} = 10^{-6}$
Epochs	150
Final quantisation error	$\mathbf{QE} = 0.0963$

2 Resulting maps and training curves

The square 15×15 map trained with an L_1 (Manhattan) grid metric and a Gaussian neighbourhood has the lowest quantisation error, QE ≈ 0.096. It outperformed Euclidean (L_2) and L_∞ variants, and also the elongated grids that PCA had originally suggested (which is an interesting observation).

Learning curves support convergence, quantisation error drops sharply during the first 30 epochs and becomes stable afterwards, while average weight adjustment falls below 0.01 (Fig. 1). The near-zero final neighbourhood radius, $\lambda_\mathrm{f} = 10^{-6}$, enabled fine-grain tuning without numerical instability, although tests showed no additional QE benefit for $\lambda_\mathrm{f} < 10^{-4}$.

You can observe the plots on the next page.

The codes, figures, and additional files are within their respective folders. Additionally, the grid metrics can also be found.

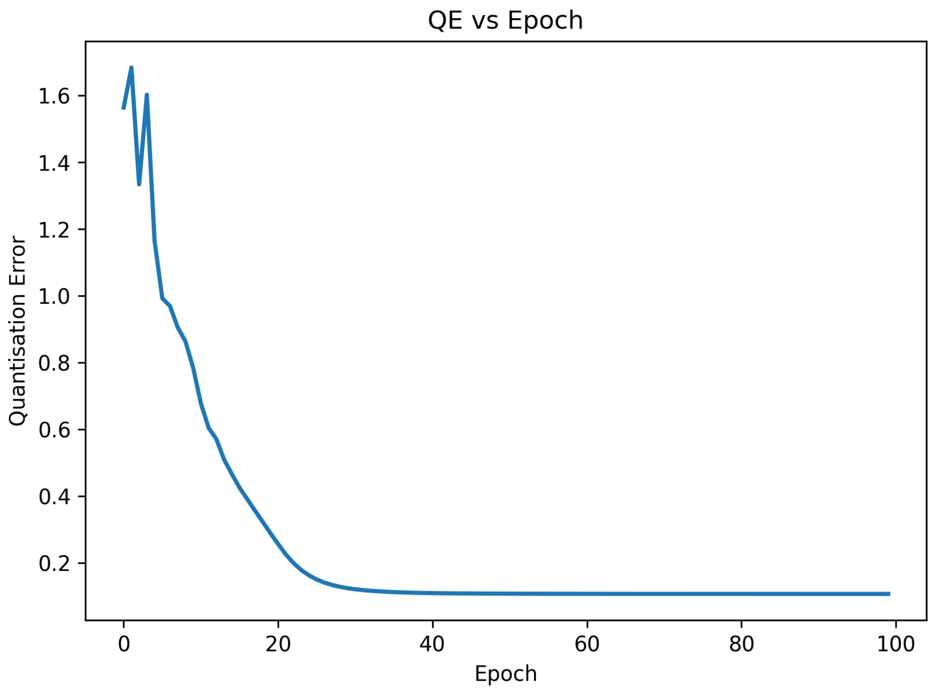


Figure 0: Quantisation error as a function of epoch. The error falls steeply during the ordering phase (epochs 0–20) and reaches a stable plateau of QE ≈ 0.10 after $\sim$40 epochs.

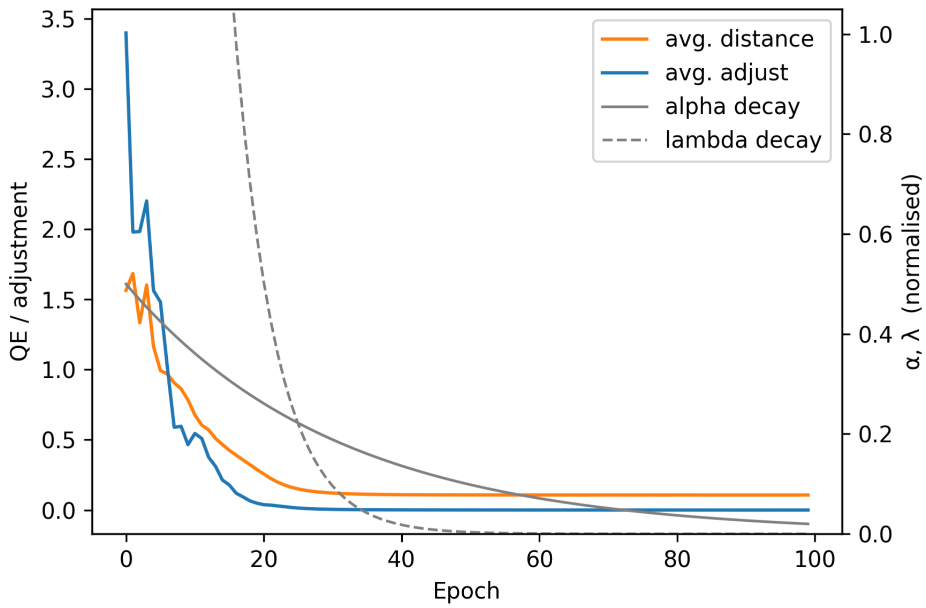


Figure 1: Combined view of training curves: quantisation error (orange), average adjustment (blue), and the analytic decay schedules of the learning-rate $\alpha(t)$ (solid grey) and neighbourhood radius $\lambda(t)$ (dashed grey). All values are normalised to $[0, 1]$ on the right y-axis.

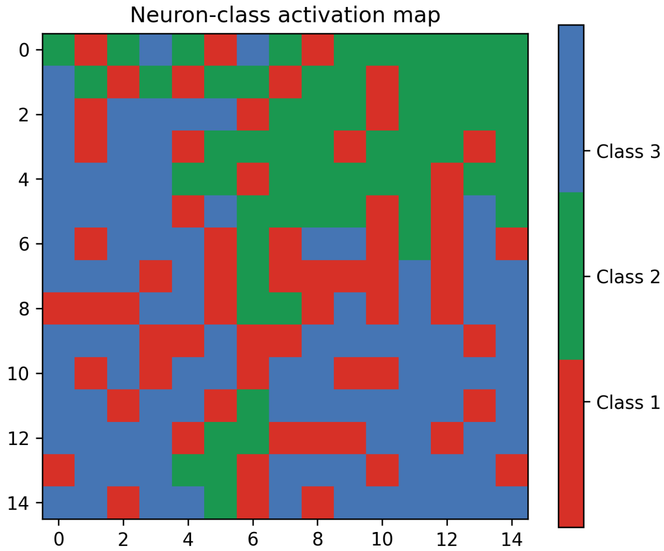


Figure 2: Neuron-class activation map. Each cell is coloured by the most frequent wheat variety (class 1–3) among the samples for which it is the BMU. Only minor class collisions occur, mostly along the central boundary, suggesting good topological ordering.

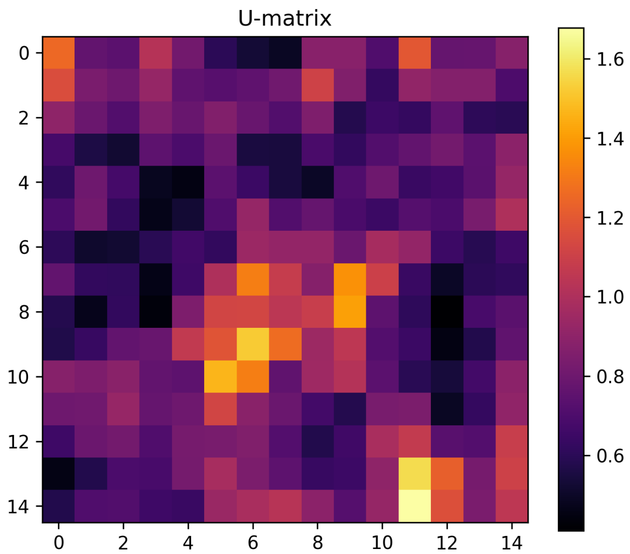
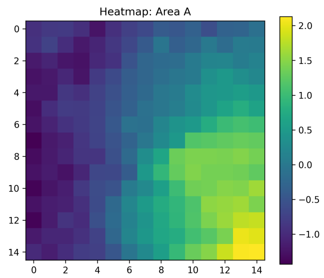
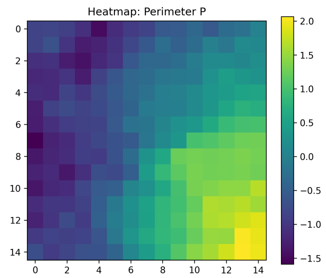


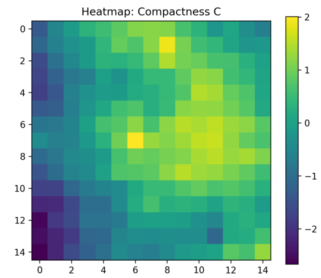
Figure 3: U-matrix of average distances between adjacent neurons. Dark low-distance regions correspond to dense areas within a class, whereas bright parts are the boundaries between clusters.



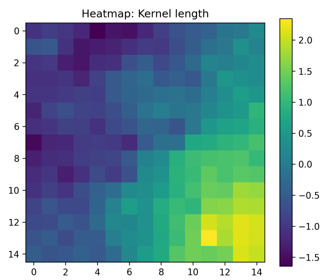
(a) Area A



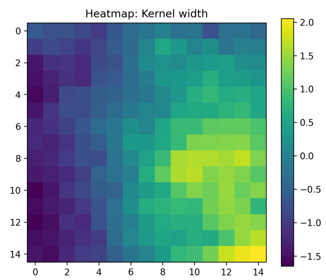
(b) Perimeter P



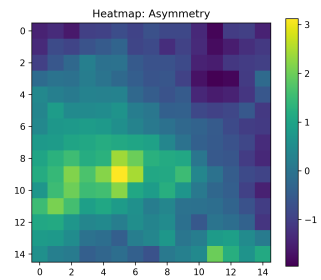
(c) Compactness C



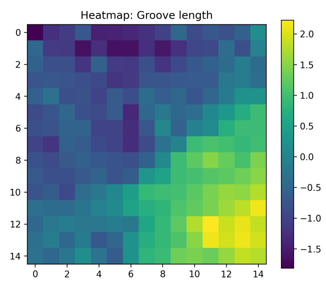
(d) Kernel length



(e) Kernel width



(f) Asymmetry



(g) Groove length

Figure 4: Distribution of each of the seven standardised attributes across the neuron lattice. Smooth gradients within classes and sharp transitions along cluster borders correspond well to the U-matrix structure.