Authors:

- Mhd Jawad Al Rahwanji - 7038980 - mhal00002@stud.uni-saarland.de
- Christian Singer - 7039059 - chsi00002@stud.uni-saarland.de

Exercise 4.2 - Supervised and Unsupervised Learning

a) The disadvantages of using MSE:

- The fact that MSE is squared has two implications: 1) being hard to interpret compared to R^2 2) squaring a large error will overestimate the true error 3) in contrast, squaring errors between -1 and 1 will underestimate the true error.
- An inherent issue with averaging is sensitivity to outliers, as they greatly affect MSE.

Using MSE for binary classification is okay. In case of a misclassification, the SE is 1 and MSE is (1 - accuracy) where accuracy is a common metric in balanced binary classification.

b) Some of the unsupervised learning tasks:

- Dimensionality reduction, reducing the number of input features. Some common methods are PCA, manifold learning and autoencoders.
- Clustering, grouping similar data points together using methods like k-means and hierarchical clustering.

PCA leverages SVD to project data in a lower-dimensional space while maximally describing residual (distance from the mean) variance. Whereas, manifold learning reduces dimensionality while preserving distances between inputs. Autoencoders are NNs that try to encode inputs with fewer and fewer neurons in each layer while still being able to restore/decode them.

c) $T > 1 \Rightarrow$ a more even (softer) predicted class distribution while maintaining relative ranks, while $T < 1 \Rightarrow$ (harder) confidence in one class overshadowing the rest. This is useful when sampling from the distribution rather than picking the highest ranking prediction. In text generation, this way we get the model to use different vocabulary instead of always repeating itself. Another use case is knowledge distillation. It is the task of training a small model to achieve a “softer” version of the large model’s predicted class distribution on a given input *x*.