Lab 2 Exercise – PyTorch AutoGrad

Justin Ugwudike (jknu1g19@soton.ac.uk)

Feb 27, 2023

**1.1 Question 1\_1**

**Text

Description automatically generated**

Figure 1 – The function to optimise.

Using the code shown in Appendix A, the rank-2 factorisation of the matrix is achieved by minimising the function shown in Figure 1. The results are shown in Figure 2.

Graphical user interface, text, application, chat or text message

Description automatically generated

Figure 2 – The reconstruction of the target matrix using U and V (obtained via gradient descent utilising automatic differentiation). The mean squared error loss is shown.

The code produces and approximation of the target matrix with the MSE loss being 0.1224, this value differs by +0.0004 to the loss achieved using stochastic gradient descent with non-automatic gradients in Lab 1.

**1.2 Question 1\_2**

With the code from Appendix A the rank-2 factorisation of the data from a dataset of 150 instances and 4 features was estimated using gradient descent. The function in Appendix B allowed me to reconstruct the data matrix using Singular Value decomposition (SVD), the losses achieved using both methods is shown in Figure 3.

Graphical user interface

Description automatically generated with low confidence

Figure 3 – The reconstruction loss of the data matrix through gradient descent and the rank-2 reconstruction loss through SVD.

The reconstruction loss of the data matrix using the function in Appendix A is 15.2292. The losses achieved when using each method are very close in value (0.0004 difference).

**1.3 Question 1\_3**

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Figure 4 – The first graph represents the data in the U matrix computed by SVD and the second shows the data compute by gradient descent.

By looking a Figure 4 both methods have grouped the data into two classes. The two pairs of classes retain similar shapes and sizes when using gradient descent and SVD however, the orientations of the two pairs of data groups are different.

**1.4 Question 2\_1**

The function used to train an MLP through gradient descent is shown in Appendix C.

**1.5 Question 2\_2**

From Appendix D I can see that either increasing the learning rate and maximum iterations leads to a higher prediction accuracy in both the training and validation data. In every experiment, the accuracy when using the training data is much greater than the accuracy when using the validation data. From these results I can infer that using a iteration count of 1000 and learning rate of 0.01 (to avoid overfitting when using a learning rate of 0.1) would be the best parameters to use in this example as the accuracies for both the training and validation data reach 99% and above.

**2 Appendix**

Appendix A - Completed gd\_factorise\_ad() function.

Text

Description automatically generated

Appendix B – Singular Value Decomposition and reconstruction function.

Graphical user interface, text

Description automatically generated

Appendix C – Function to perform MLP gradient descent using automatic differentiation.

Text

Description automatically generated

Appendix D – Three sets of accuracies recorded from an MLP when using different learning rates and maximum iterations.

A screenshot of a computer

Description automatically generated with low confidenceA screenshot of a computer

Description automatically generated with low confidenceA screenshot of a computer

Description automatically generated with low confidence