Deep Learning: Lab 1

Johnny Joyce (jvjj1u19@soton.ac.uk)

1 Implement a matrix factorisation using gradient descent

We were tasked with implementing low-rank matrix factorisation through gradient descent. With default parameters, the implemented algorithm factorised the matrix $\mathbf{A}$ as $\mathbf{U}\mathbf{V}^\intercal$, where:

$$\mathbf{A} := \begin{bmatrix} 0.3374 & 0.6005 & 0.1735 \\ 3.3359 & 0.0492 & 1.8374 \\ 2.9407 & 0.5301 & 2.2620 \end{bmatrix} \quad \mathbf{U} = \begin{bmatrix} 0.6539 & -0.1247 \\ 0.3231 & 1.5237 \\ 1.0369 & 1.1576 \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} 0.7180 & 1.9812 \\ 0.7644 & -0.1685 \\ 0.7645 & 1.1339 \end{bmatrix}$$

This gave a mean square error of 0.1219.

2 Comparison to SVD

By computing the SVD of matrix $\mathbf{A}$ and changing the last singular value to 0, we achieved a mean square error of 0.1219 on the reconstructed matrix. This is the same value as in Section 1, though these values will not always be identical - rather, our new value is a lower bound for the possible values we could have found in Section 1. This is because removing the last singular value reduced the ranks of both $\mathbf{\Sigma}$ and the reconstructed matrix to 2, making the resulting reconstruction a best-case rank-2 approximation.

3 Matrix completion

An updated version of the factorisation algorithm was implemented which performed matrix completion. The algorithm was then run on a new matrix

$$\mathbf{B} := \begin{bmatrix} 0.3374 & 0.6005 & 0.1735 \\ * & 0.0492 & 1.8374 \\ 2.9407 & * & 2.2620 \end{bmatrix}$$

...where entries marked by * represent unknown values. With the resulting output $(\mathbf{U}, \mathbf{V})$, our reconstruction was as follows:

$$\mathbf{U}\mathbf{V}^\intercal = \begin{bmatrix} 0.3344 & 0.5942 & 0.1770 \\ 2.4562 & 0.0511 & 1.8366 \\ 2.9404 & -0.4539 & 2.2626 \end{bmatrix}$$

If we consider the original matrix $\mathbf{A}$ to be our target, this reconstruction had a mean square error of 1.7422. Furthermore, we have that the difference between our reconstruction and the original matrix $\mathbf{A}$ is

$$\mathbf{A} - \mathbf{U}\mathbf{V}^\intercal = \begin{bmatrix} -0.0030 & -0.0063 & -0.0035 \\ -0.8797 & -0.0019 & -0.0008 \\ -0.0003 & -0.9840 & 0.0006 \end{bmatrix}$$

...which shows that the errors were largest for the positions with missing values, as one would expect.