2IIG0 - Homework 3 - Question 5

M.F.J. Moonen (1234115), Jin Ouyang (1608541), Bas Witters (1625187)

1 Subquestion A

The function is implemented following the pseudocode mentioned in the HW3 pdf. Please take a look at the code and its respective comments that we provided on canvas. The Mean Squared Error on the Observed (MSEO) entries against the iterations of training can be seen in Figure 1.

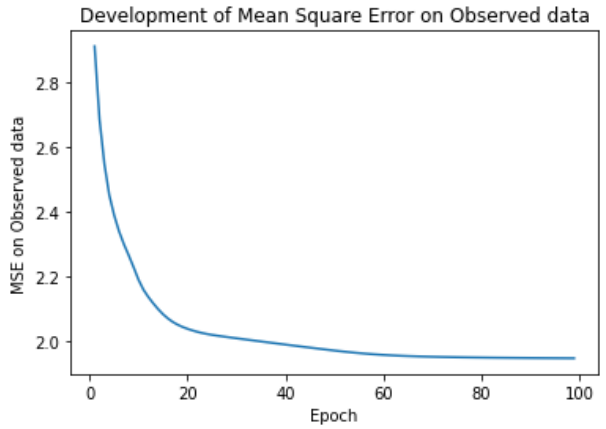

Fig. 1. MSEO of recommender system on observed data during training.

2 Subquestion B

The stationary points X_k computed in Eq.4 are the minimizers of the block-coordinate objectives. According to the slide 44 of L06, the slide 26 of L09, which are shown in Figure 2 and Figure 3, and exercise 5 itself, we know that the optimization problems of Eq.1 and Eq.3 are equivalent. The penalization terms will not influence the minimizers it finds. Thus, the stationary points which are well-defined by Eq.4 are actually the stationary points of block-coordinate objectives.

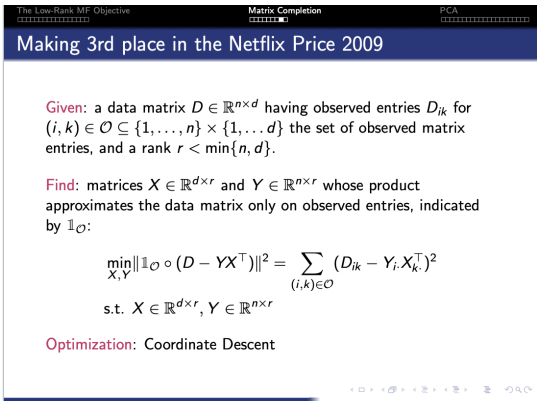

Fig. 2. slide 26 of L09

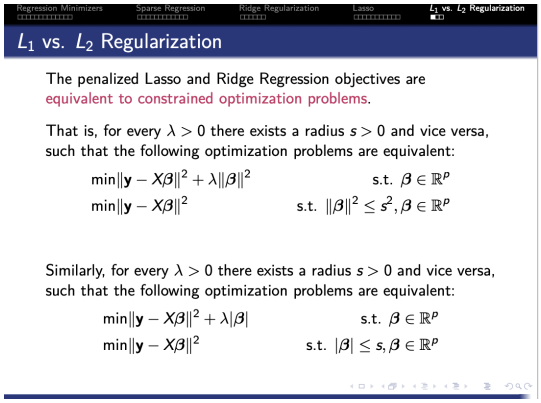

Fig. 3. slide 44 of L06

3 Subquestion C

The stopping criterion we have implemented is that a decrease in the MSEO must be achieved of at least 0.0025 within three epochs, otherwise training is stopped. The most common "train until validation error stops improving" would not work with this type of machine learning system, as it is guaranteed to always improve until the optimum is reached. The impact of this stopping criterion can be seen in Figure 4, and is detailed below.

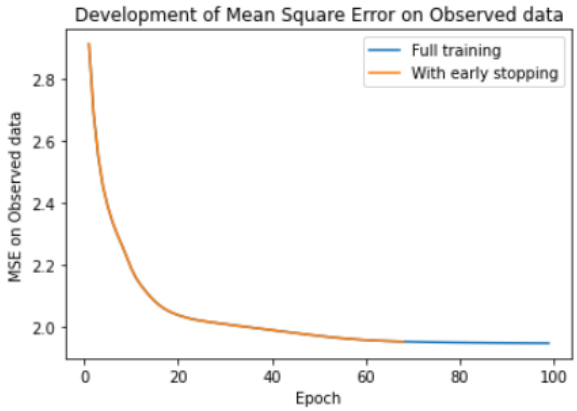

Fig. 4. Impact of early stopping criterion on training of recommender system. Where the line turns blue is where early stopping would have stopped the training.

It takes 67 epochs before convergence is reached according to our stopping criterion and the MSEO when using the stopping criterion is 1.9525, whereas the MSEO after a full run is 1.9471. Thus, the full run does indeed improve the MSEO but only by a very slight margin. In a real world scenario using the stopping criterion would likely be recommended.

Note: subquestion D can be found on the next page.

4 Subquestion D

The further the regularizing parameter is decreased, the more the values produced by the recommender system leave the range of expected values. With $\lambda = 0.0001$ we can predicted values as high as 22.25 or as low as -5.16, which become less extreme with lower λ values. It seems that the result remains more interpretable with a higher λ.

With a smaller λ value, the MSEO decreases visibly more slowly. The impact of different values on this development can be seen in Figure 5. It is also noteworthy (and expected) that the early stop is triggered later when this development happens more slowly. The MSEO always seems to end up at roughly the same value, outside of that for $\lambda = 0.0001$ but this setting had not yet converged when the cutoff of 100 epochs was reached.

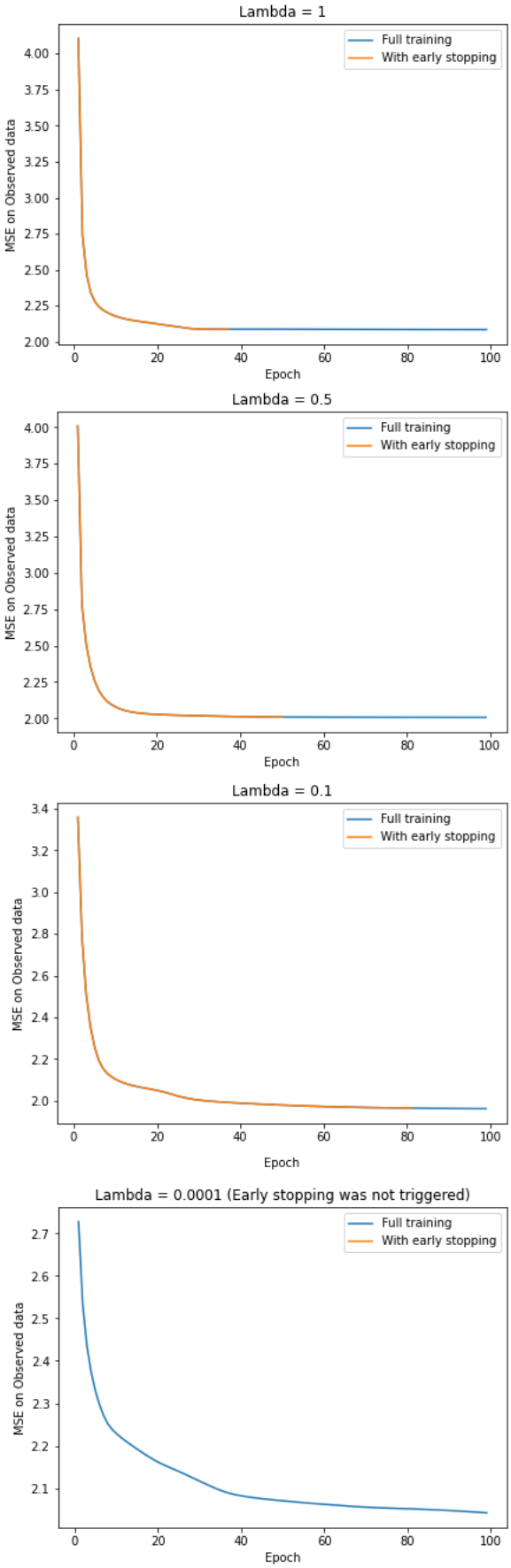

Fig. 5. MSEO development on different lambda settings.