SF2525 HW5

Wille Viman Vestberg
980329-7317
willev@kth.se

26 Maj 2023

1 TensorFlow minimization

I ran the code *tf2_mnist_lsq.py* and the accuracy was 0.8852.

We can improve it by decreasing the hyper-parameter Nbatch which corresponds to the batch size.

I decreased it to 4 and the accuracy became: 0.9150

The drawback with decreasing Nbatch is the runtime.

2 Cross Entropy

1

The accuracy of the code *tf_mnist_entrypy.py* was for me: 0.9230

2

Changing the activation function from sigmoid to Relu the accuracy increased to: 0.9531

3

Changing to Adams Optimizer the accuracy was increased to 0.9631. Doubling M gives the sama result.

4

Now we will add 2 hidden layers to the neural network. The accuracy is increased to 0.9670. Using 'he_normal' we obtain a slight decrease in accuracy.

5

Finally we will investigate the impact of increasing K. Using K = 40, we get the accuracy 0.9687. Using K = 50 the accuracy was increased further to 0.9709. And using K = 100 gives 0.9752. But the run-time became a lot slower.

Part 2

1

The code can be written:

```
function OneHot = OH(x)
    N = length(x)
```

```
    M = zeros(N,10);
    for i = 1:N
        index = x(i)+1;
        M(i, index) = 1;
    end
    OneHot = M;
end
```

2

Choosing K = 1024 J = 60000 and $\lambda = 0.01$ I obtained the percentage of correct labeled digits: 0.9395 and hence the percentage of mislabeled digits was: 0.0605.

3

Then I tried to replicate the figure from the problem sheet by plotting the Percentage of miss-labeled digits against K. Using Logarithmic scaling for both x-and y axis and scaling the y-values by 10^2.

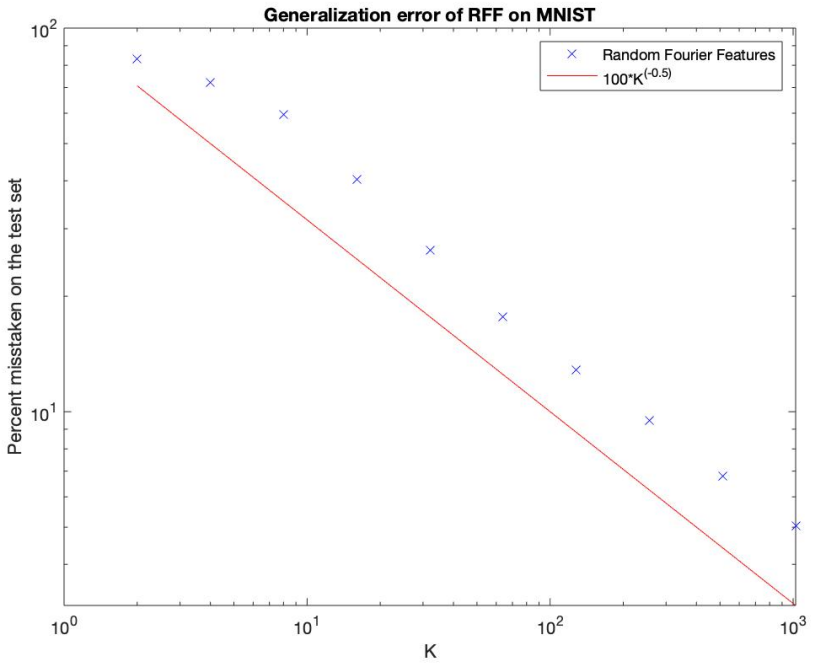

Figure 1: Recreation of the figure from the Assignment sheet.

I was able to run for K = $\begin{pmatrix} 2 & 4 & 8 & 16 & 32 & 64 & 128 & 256 & 512 & 1024 \end{pmatrix}$ Without excessive run-time. The figure was replicated pretty accurate.

4

The generalization error is given by the percentage of miss-classified labels.

I chose to use Lambda values 0.0001. 0.001, 0.01, 0.1, 1 and 10 and i fix K = 1024.

The following plot was generated,

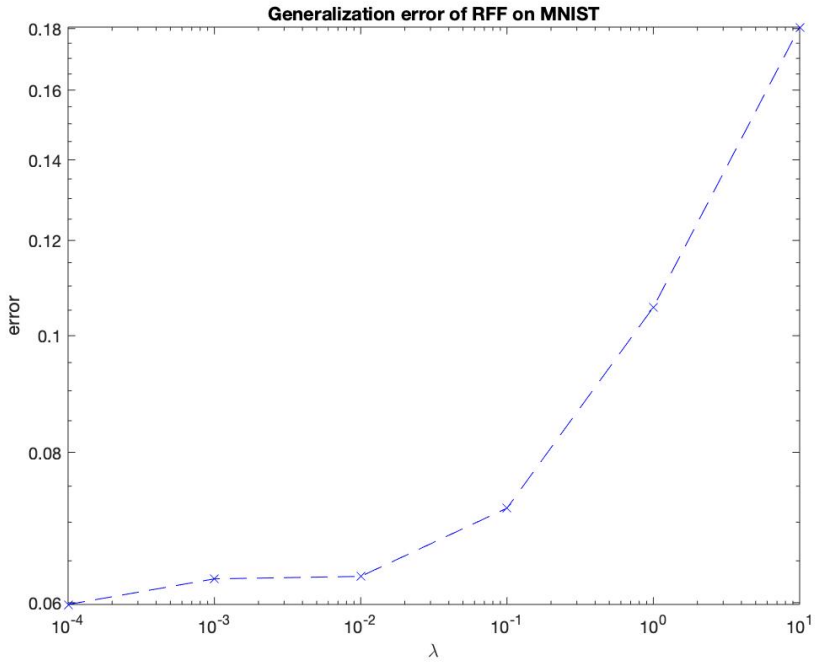

Figure 2: The effect of lambda on the generalization error

As we can see the error the error first increases when going from $\lambda = 10^{-4}$ to $\lambda = 10^{-3}$. Then it remains constant when going up to $\lambda = 10^{-2}$. then Increases in an exponential way. Hence we would be wise to choose $\lambda \leq 10^{-2}$ to keep the error low.

5

I tried using an exponential distribution but it gave similar results, no improvements. Gamma and uniform distributions both gave very bad results.