Effect of Decay Rate

COL870: Assignment-1

Kshitij Alwadhi (2019EE10577)

5 October 2022

Contents

1 Introduction

In this assignment we check the effect of decay rate (regularization parameter) while training a neural network. In the loss function, we add a term which penalizes the squared sum of the weights of the neural network. This summation is multiplied by a constant called "decay rate" of the model.

2 Experimental Setup

We pick up the best performing model from the first assignment of the course and vary the regularization parameters. We store the train, validation and test losses and accuracies and plot them with respect to the decay rates. We take the following decay rates for our experiments:

$$[]0, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1]$$

3 Results

Here are the plots we get from our experiments:

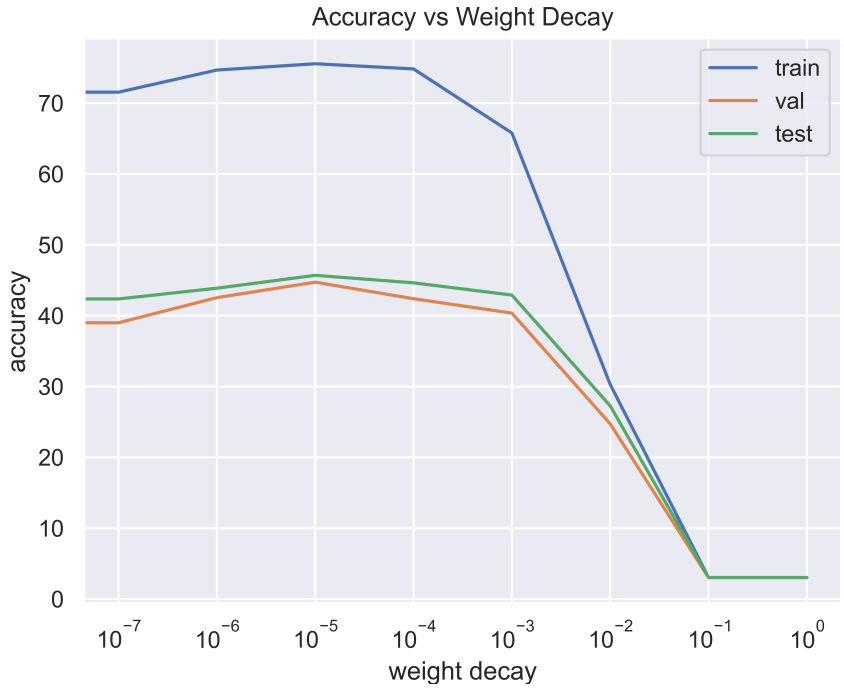

Figure 1: Accuracy vs Decay Rate

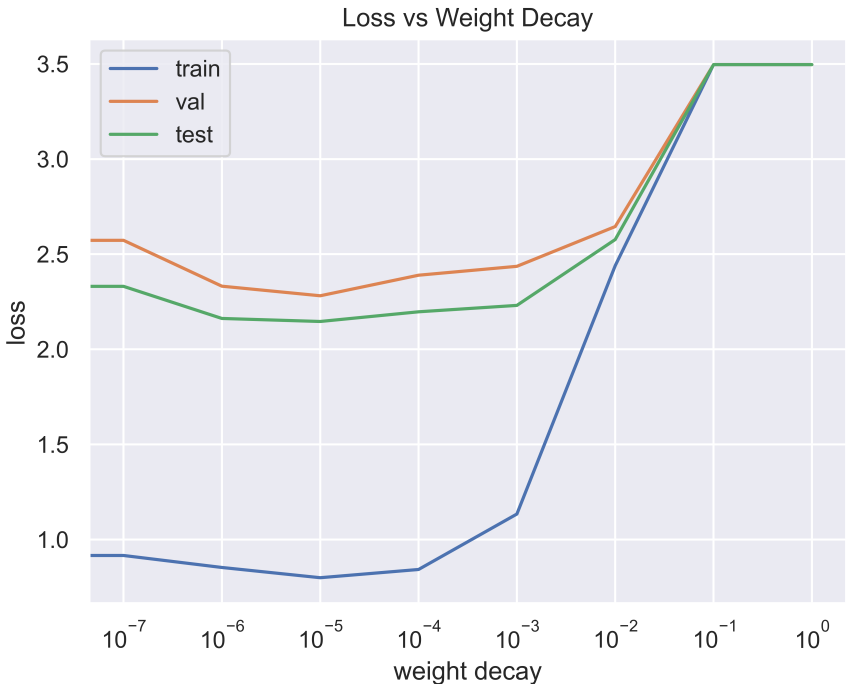

Figure 2: Loss vs Decay Rate

We observe that when we have a very large value of regularization parameter, the model fails to

learn anything since there is a high penalty on the weights. In this setting, we have the situation of under-fitting. Also, we observe the peak performance when the value of weight decay is somewhere between 10^{-5} and 10^{-3}. This is the case of best fitting. Before that, i.e. for smaller values of λ we have the case of overfitting, i.e. the model fails to generalize to unseen data.