336546 Machine Learning in Healthcare - HW4

Adi Waisman 316407279, Efrat Ben-Sadeh 204232722

Part 1-

- Task 2: Activation function is used to determine the output of the NN. It transforms the weighted sum of the inputs into the activation of the node. The activation maps the resulting values to a fixed interval according to the used function:
  Sigmoid is a non-linear function that puts the values between 0 and 1, therefore, it is mostly used for models where we have to predict the probability as an output.
  Tanh function is similar to the sigmoid function, but it has the effect of centering the data at zero, which makes learning easier for the next layer. Tanh puts the output values between -1 and 1.
  Leaky ReLU is an improved version of the ReLU function, where we remove the zero gradient. Leaky ReLu puts the values between -infinity and infinity.

- Task 3: The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset. By increasing the epoch's number, we are minimizing the model's error. But if the number of epochs is increased beyond the optimal/sweet spot the model can start to over-fit the data.

- Task 4: Mini batches:
  The advantages of mini-batch over SGD is that it provide a more computationally efficient process than SGD. It can be performed in a distributed manner, meaning each mini-batch can be computed in parallel across multiple servers to achieve significant improvements in training speeds. Another advantage is the more stable converge towards the global minimum, since it calculates an average gradient over more samples that results in less noise.

- Task 4: Batch normalization: Batch normalization accelerates the learning process; it stabilizes the learning process and reduces the number of training epochs required to train the NN. It also provides regularization reducing the generalization error. In our work, adding the batch normalization layers did not improve the model's performance, we even got higher loss and slightly lower accuracy. A possible explanation could be that the quality of the input/data is bad, therefore, even batch normalization could not help to improve the learning.

Part 2-

- Task 1:
  *The CNN has 8 layers: 7 hidden layers and 1 output layer.
  *In the first layer there is 64 filters, in the second and third layer there is 128 filters, and in the fourth and fifth layers there is 256 filters.
  *The number of parameters in a CNN is not similar to a fully connected NN, CNN enables to reduce the number of parameters by orders of magnitudes.
  *Yes, this CNN does perform regularization

- Task 2: The model with the reduced number of filters had a lower loss and a slightly lower accuracy.