Machine Learning Assignment 5 - Image Recognition with Convolutional Neural Networks (CNN)

In this assignment you are going to use convolutional neural networks to perform some basic image recognition tasks. The image dataset we have provided you contains images (in jpeg format) of dimensions 120x128, with the training set consisting of 315 images, a validation set consisting of 45 images and a test set consisting of 45 images.

In each image, the subject has the following characteristics:

Name – name of the subject

Direction Faced – left, right, straight, up

Emotion – happy, sad, neutral, angry

Eyewear – open, sunglasses

Each image follows the naming convention "*name_directionFaced_emotion_eyewear.jpg"*

Even though the original dataset contains images of size 120x128, you will use the same image compression technique you applied in *assignment 4* to convert the 120x128 pixel images into images of size 30x32, and for most of this assignment you will be working with the compressed version of the original images.

We will be using the convolutional neural model described here http://neuralnetworksanddeeplearning.com/chap6.html so you may need to study the network3.py code before starting your assignment.

(a) **Training and Model Evaluation – Direction Faced**

Your first task is to design a convolutional neural network (CNN) in the task of predicting the direction a given subject is facing. Your CNN should have a local receptive field of size 5 x 5, one input feature map and 20 filters. In addition to that, your CNN should also have a pooling layer of size 2 x 2, a fully connected hidden layer with 100 neurons and a softmax layer with 4 output units. You should use a learning rate of 0.01, a mini_batch_size of 10 and train your model for 60 iterations.

In the script *train_and_eval_direction_faced.py* you should plot both the validation and test accuracies per epoch (on the same graph) and print the final accuracy of your model on the test set. Do this using the ReLU activation function and the sigmoid activation function. Which one performs better? Comment. – **50 points**

(b) **Training and Model Evaluation – Emotion Felt**

Your next task is to design a convolutional neural network (CNN) in the task of predicting the emotion a given subject is feeling. Your CNN should have a local receptive field of size 5 x 5, one input feature map and 20 filters. In addition to that, your CNN should also have a pooling layer of size 2 x 2, a fully connected hidden layer with 100 neurons and a softmax

layer with 4 output units. You should use a learning rate of 0.01, a mini batch of size 10 and train your model for 60 iterations.

In the script *train_and_eval_emotion_felt.py* the model_evaluation function takes as input a **test** image of dimensions 30 x 32 and using your network, it should return a string specifying the emotion that the test subject is feeling ("happy", "sad", "neutral", "angry") – **20 points**

(c) **Convolutional Neural Net Optimization**

For this section you are free to use any CNN architecture that would give you the highest possible accuracy in predicting the emotion that a subject is feeling. You may use multiple convolutional, pooling and fully connected layers. You should also optimize the hyperparameters of your network. In the script *cnn_model.py* the model_evaluation function takes as input a **test** image of dimensions 120x128 and using your optimized network, it should return a string specifying the emotion that the test subject is feeling ("happy", "sad", "neutral", "angry"). **– 30 points**