TBMI26 – Computer Assignment Reports Deep Learning

Deadline – March 14 2021

Author/-s:Wuhao Wang wuhwa469 Hong Zhang honzh073

In order to pass the assignment you will need to answer the following questions and upload the document to LISAM. Please upload the document in PDF format. **You will also need to upload the Jupyter notebook as an HTML-file (using the notebook menu: File -> Export Notebook As…)**. We will correct the reports continuously so feel free to send them as soon as possible. If you meet the deadline you will have the lab part of the course reported in LADOK together with the exam. If not, you'll get the lab part reported during the re-exam period.

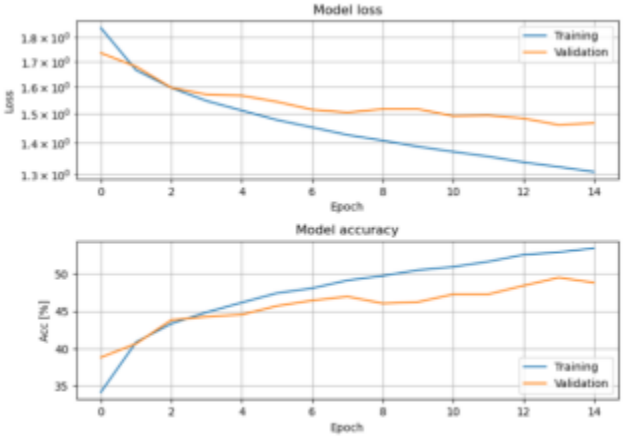
1. **The shape of X_train and X_test has 4 values. What do each of these represent?**

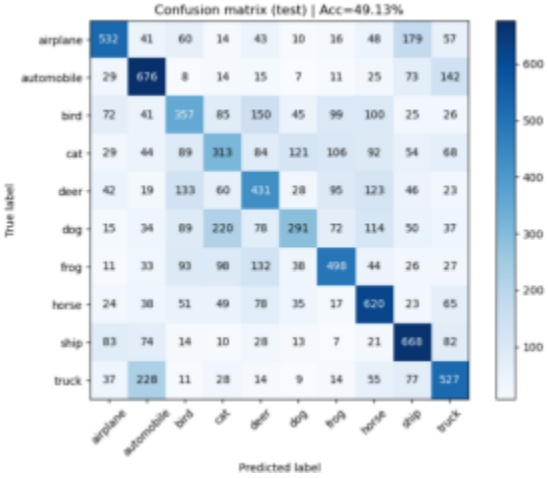
   The meaning of these 4 dimensions is :

   (Number of data, row_pixels, col_pixels, color_channel)

2. **Train a Fully Connected model that achieves above 45% accuracy on the test data. Provide a short description of your model and show the evaluation image.**

   Before the output layer, we add two Dense layers using tanh as activation function, both of them have 64 neurons. The out put layer uses SoftMax as activation function, it should be 10 neurons since there are 10 classes.





3. **Compare the model from Q2 to the one you used for the MNIST dataset in the first assignment, in terms of size and test accuracy. Why do you think this dataset is much harder to classify than the MNIST handwritten digits?**

   1: The model will take all the pixels into count. And the backgrounds in the first dataset are pure black while they are colorful in the second dataset. Meanwhile, we can see that some pictures from different classes will have very similar backgrounds. In this case, the background in the second dataset would be an interference for prediction which make it harder.

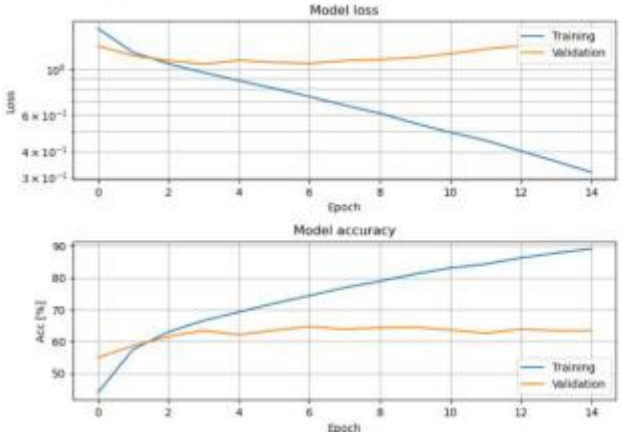
2: In the cv, we need details(contours, corner, etc.) to describe an object. And some classes in the second data share similar details(e.g., truck vs. automobile, cat vs. dog which We can find these from the confusion matrix). In other words, we think hand-writting digital number is less complicated than images in CIFAR10, so we need deeper features to help explore the difference between different classes. However, when the model become deeper and deeper, the overfitting risk also increase, so, it's more difficult to find a good solution in the CIFAR10 dataset.

4. **Train a CNN model that achieves at least 62% test accuracy. Provide a short description of your model and show the evaluation image.**

We add 2 stacks of [converlution-activation-pooling] at the beginning. After pooling we flatten the data and go through 2 hidden layers and one output layer.

These 2 converlution parts share same parameters: 64 channels, (3,3) kernel size.

These 2 poolings are maxpooling with same parameters : (2,2) pool size and (2,2) strides.



5. **Compare the CNN model with the previous Fully Connected model. You should find that the CNN is much more efficient, i.e. achieves higher accuracy with fewer parameters. Explain in your own words how this is possible.**

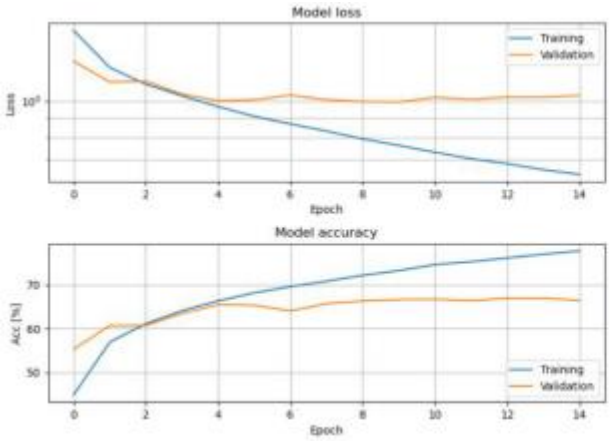
1 In the CNN each neuron does not connect to all the neurons in the last layer but only a part of them which can help reduce the parameters.

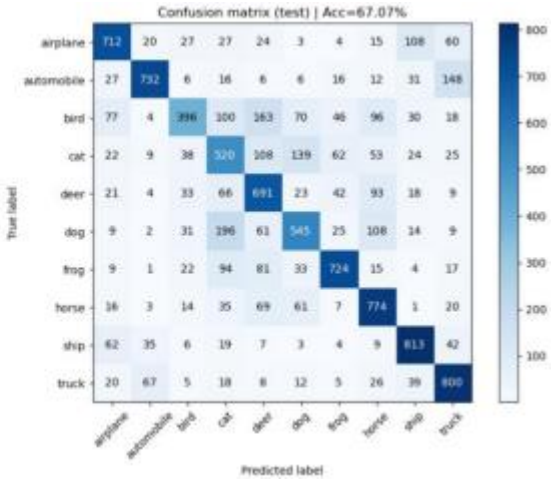
2 In the CNN weights are shared (by using the window) which also help reduce the number of parameters.

3 Pooling is used in CNN which can help reduce the samples in each layer and ,at the same time, reduce the number of parameters. Meanwhile, pooling can also filter the less important features so that making the model more robust and avoiding overfitting.

6. **Train the CNN-model with added Dropout layers. Describe your changes and show the evaluation image.**

We add dropout layer between pooling first and second dense layer and the dropout rate is 0.2. A small improvement can be seen.





7. Compare the models from Q4 and Q6 in terms of the training accuracy, validation accuracy, and test accuracy. Explain the similarities and differences (remember that the only difference between the models should be the addition of Dropout layers).

Hint: what does the dropout layer do at test time?

	Training accuracy	Validation accuracy	Test accuracy
With dropout	**0.7849**	**0.6648**	**0.671**
Without drop out	**0.9046**	**0.6339**	**0.636**

Dropout layer:

During the test stage, dropout layer will do nothing.

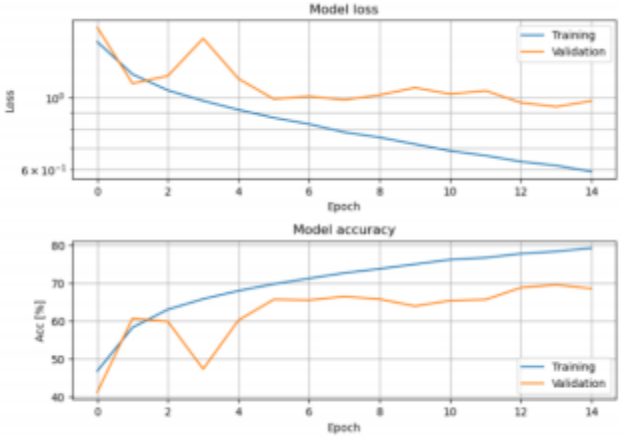
In the training stage, dropout layer will force every neuron(in this layer) to be deactivated with certain probability. So during the training, at each iteration, only a part of neurons will be updated. This make the neurons become more independent to eachother, thus reducing the risk of overfitting.

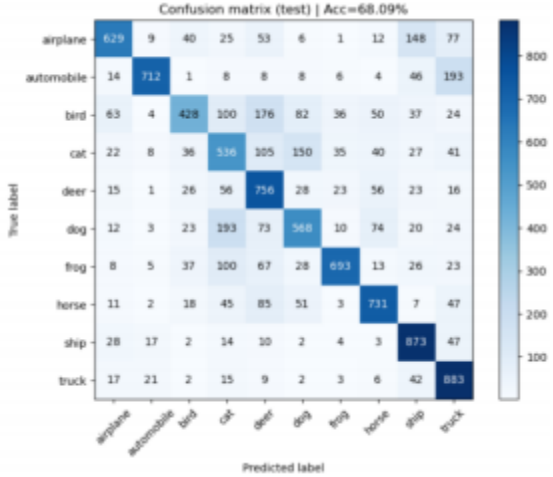
However, for the training data, applying dropout layer will make model 'smaller' . So, the training accuracy will be lower than the model without dropout layer.

Similarities: From the structure level :Both these two models have classic CNN procedure (convolution - pooling). Both of them are fully connected nets.

Difference: The model in the Q6 applies dropout layers

8. Train the CNN model with added BatchNorm layers and show the evaluation image.





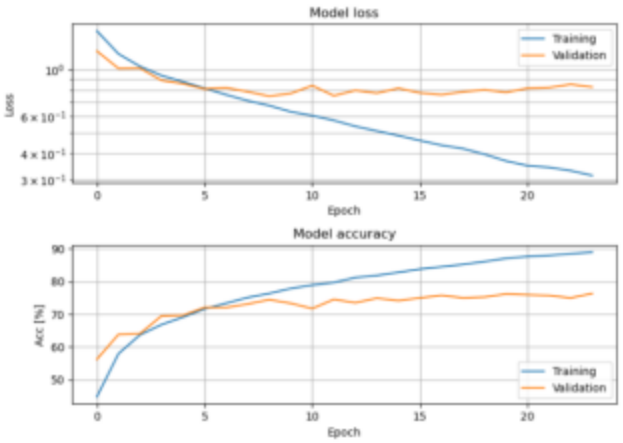
9. **When using BatchNorm one must take care to select a good minibatch size. Describe what problems might arise if the wrong minibatch size is used.**

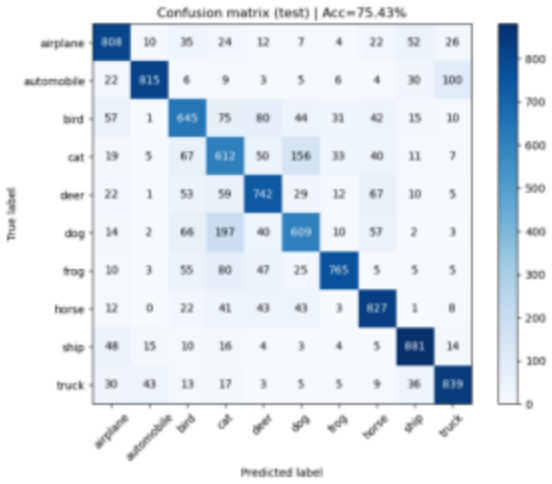
You can reason about this given the description of BatchNorm in the Notebook, or you can search for the information in other sources. Do not forget to provide links to the sources if you do!

The parameter of bachNormlization is learned during the training process. So, if the batch is too small then this group may have different distribution of the whole data which means those parameters are not suitable for the whole data. As a result, it may cause slower convergence. But it also provide the possibility of jumping out of local minimum.

Meanwhile, the first point of too big batch size maybe the insufficient memory resource (especially when images are quite huge). Secondly, bigger batch size will make the direction of gradient in each batch become more stable (since the distribution is closer to whole data), so it would accelerate the convergence but also have higher local minimum risk.

10. **Design and train a model that achieves at least 75% test accuracy in at most 25 epochs. Explain your model and motivate the design choices you have made and show the evaluation image.**





Like in the previous tasks, We use 2 stacks of [convolution-pooling] and three dense layers. The number of neurons goes down from 512 to 128 each layer contains half than previous layer. Meanwhile, Batchnormolization is applied in each layer and dropout layer between each Dense layer.

The parameter momentum in the BatchNormalization will keep memories of mean and variance of previous batch. Enlarge this parameter will accelerate convergence. Since now we have comparatively deep net so we increase the momentum from 0.75 to 0.85 in the BatchNormalization.