TBMI26 – Computer Assignment Reports
Deep Learning

Deadline – March 14 2021

Author/-s: Mohammed Ali & Simon Alsén

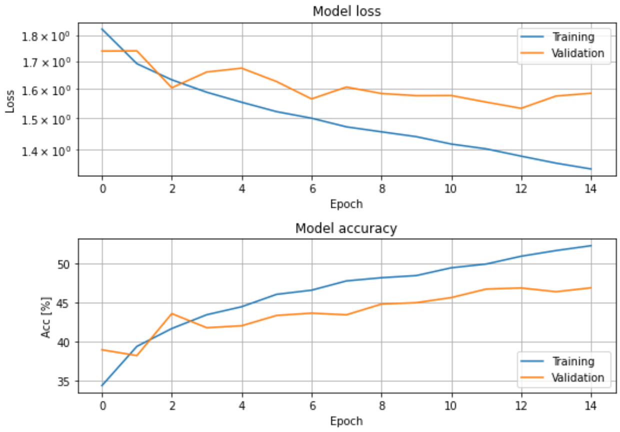
In order to pass the assignment you will need to answer the following questions and upload the document to LISAM. Please upload the document in PDF format. **You will also need to upload the Jupyter notebook as an HTML-file (using the notebook menu: File -> Export Notebook As…)**. We will correct the 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.

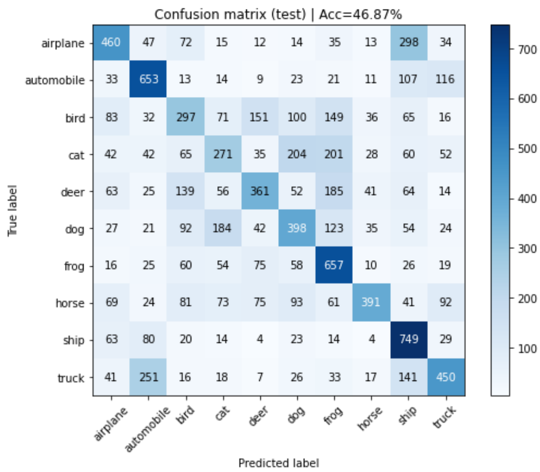
1. **The shape of X_train and X_test has 4 values. What do each of these represent?**
   The first one represents the number of images, the second and third one represents the x and y pixels respectively, and the fourth one represents the RGB channels.

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.**
   Our model uses two layers:

```
Dense(1024, activation='relu')(x)
Dense(10, activation='softmax')(x)
```

The first layer uses the ReLU activation function to avoid the vanishing gradient problem. The layer consists of 1024 neurons, and the number of neurons was chosen through trial and error. The second layer uses a Softmax function which is able to represent probabilities for different classes. Because there are 10 classes, the layer is set to consist of 10 neurons.





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?**

The images in the MNIST dataset are 28x28 pixels, and consists of only two colors (black and white). The images in the CIFAR10 dataset are 32x32 pixels, and uses RGB color.

The variation is also higher in the classes in the CIFAR10 dataset than in the MNIST dataset. The numbers are 2 dimensional, while the objects in the images in the CIFAR10 dataset is 3 dimensional. The backgrounds in the images in the MNIST dataset are all black, while the backgrounds in the images in the CIFAR10 are much more varied.

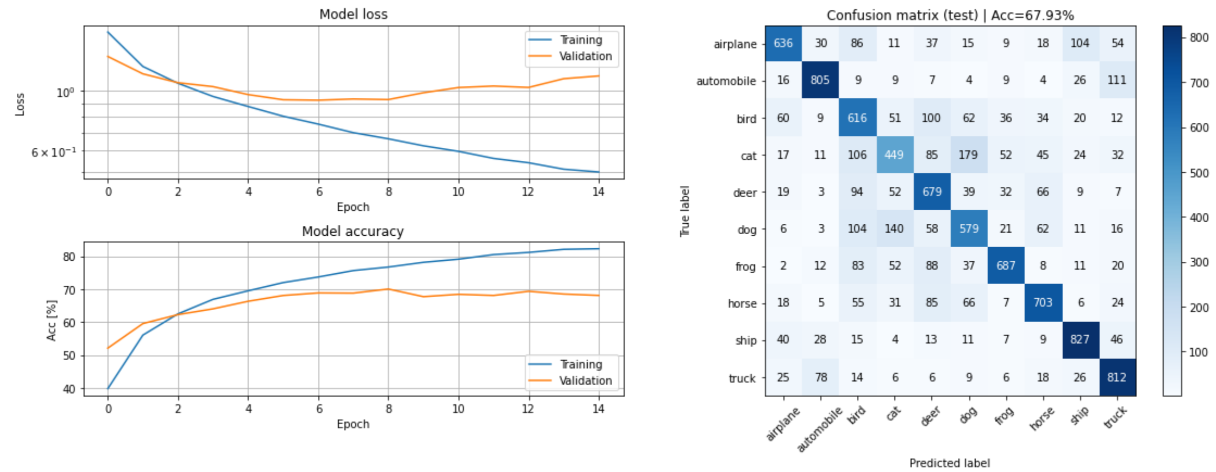
Because of this, the images in the CIFAR10 dataset will be harder to model, and as a result the test accuracy is much lower for the model for the CIFAR10 dataset than for the model for the MNIST dataset.

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

The model consists of:

- A convolutional layer with input shape (32, 32, 3). The kernels (also known as filters) are set to have size 3x3. The layer uses the ReLU activation function.
- A pooling layer which performs down sampling operation by taking the max value over a 2x2 pooling window.
- A convolutional layer with kernel size 3x3, that uses ReLU as the activation function.
- A 2x2 max pooling layer.
- A convolutional layer with kernel size 3x3, that uses ReLU as the activation function.
- A flatten layer that flattens the input.
- A dense layer with 10 neurons that uses the softmax activation function. This layer is used to make the final classification.

By using multiple convolutional layers, the model allows for more non-linearity.



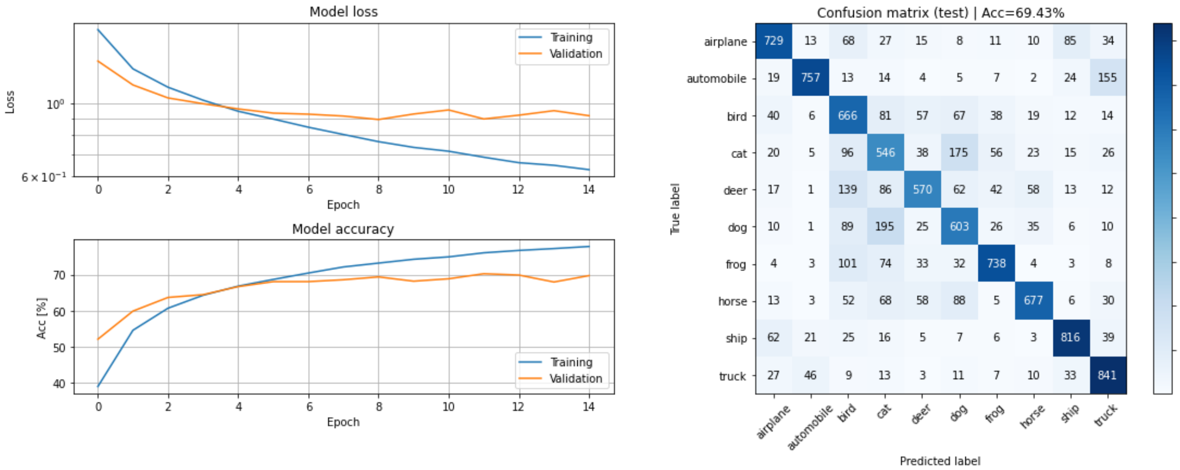
5. **Compare the CNN model with the previous Fully Connected model. You should find that the CNN is much more efficient, i.e. achieves higher accuracy with fewer parameters. Explain in your own words how this is possible.**

The fully connected model has over 3 million parameters, while the CNN model has 66570 parameters. The fully connected model only achieves 46.9% accuracy, while the CNN model achieves 67.9% accuracy.

CNNs uses kernels that are shifted over the images. This allows the CNNs the capability to learn and detect objects in images regardless of the objects positions in the images (i.e., a CNN does not need to learn a feature such as an edge, line or color of an object in all the regions of an image). A fully connected model can learn to recognize objects in images, but they won't be able to recognize the objects if they are moved to another location in the images.

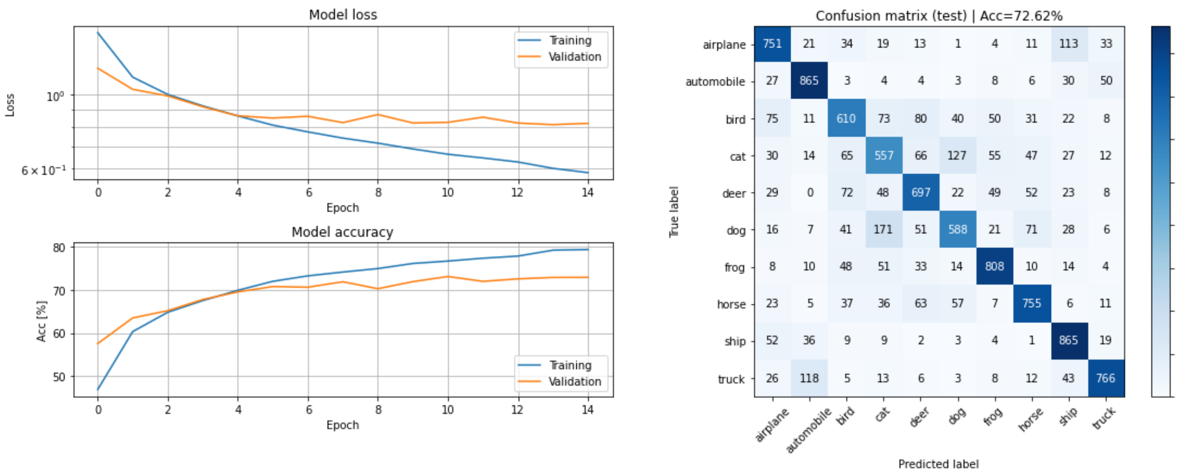
A CNN only needs to learn the coefficients in the kernels, and since the size of the kernels are smaller than the images, a CNN has to learn fewer parameters than a fully connected model. This speeds up convergence, and reduces the risk for overfitting, which in turn will improve the generalization of the models and the test accuracy.

6. **Train the CNN-model with added Dropout layers. Describe your changes and show the evaluation image.**
A dropout layer has been added before the last, dense layer. The dropout layer randomly turns of 20% of the nodes in the previous layer in order to avoid overfitting.



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?
The training accuracy is slightly lower for the model in Q6, but the validation and test accuracy is slightly higher. The dropout layer reduces overfitting, i.e., increasing bias but reducing variance.

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!
   A small minibatch size leads to faster training. However, if the minibatch size is too small, the risk for inaccurate estimates of the statistics (μ and σ^2) increases which might cause unstable results. The noise introduced by a smaller minibatch size can have a regularizing effect on the model, i.e., improving the generalization of the model and thereby also improving the test accuracy.

   A larger minibatch size will instead result in estimates that are closer to the actual mean and variance in the whole population, and subsequently a more accurate estimate of the gradient. However, a large minibatch size will also lead to slower convergence when training the model. A large minibatch size can also cause overfitting, which will have a negative impact on test accuracy.

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.**
    Our model consists of two blocks of two convolutional layers (to allow for non-linearity), followed by a max pooling layer and a dropout layer (the first dropout layer uses a rate of 0.2, and the second dropout layer uses rate 0.5). The convolutional layers use ReLU activation to avoid the vanishing gradient problem. After these two blocks, the input is flattened, and batchnormalization is applied. Finally, two dense layers are used using ReLU and softmax activation respectively.