Andreas Gwyther-Gouriotis 13893627

Example 2 – Neural Networks:

Problem Statement:

A fashion design company wants to analyse what clothes people are wearing this summer. To do this the company has requested a model that can record what types of clothes people are wearing as they walk past the shop. In order to have an automated AI that can analyse what type of clothing someone is wearing a convolutional neural network (CNN) will be developed in order for the AI to learn how to recognise the different types of clothing. The clothing types can be seen below:

The categories are:

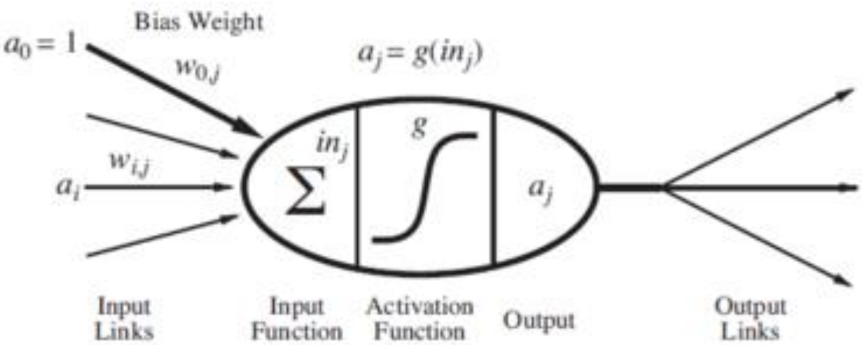
1. T-shirt/top
2. Trouser
3. Pullover
4. Dress
5. Coat
6. Sandal
7. Shirt
8. Sneaker
9. Bag
10. Ankle boot

Dataset Source:
github.com/zalandoresearch/fashion-mnist

The developed AI must be able to accurately identify what type of category of clothing someone is wearing. Once completed the fashion design company can use this information to produce new clothing that aligns with what people are wearing in summer.

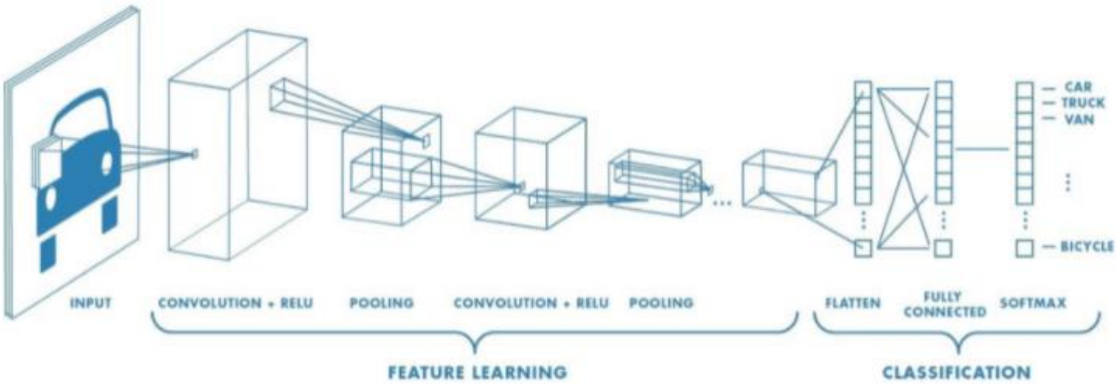
Description of the AI technique:

A neural network is based off the biology of how a human brain works essentially replacing connected neurons in the brain with connected nodes in the neural network. To speak generally, a neural network is a collection of nodes (neurons) connected together. The four components of a neural network include the input links, the input function, the activation function and the output. The following diagram shows these four components:



Lu, H. (2021)

When discussing a CNN, the main components are the input layer, the output layer, and the multiple hidden layers. The hidden layers of a CNN consists of a series of convolutional layers that convolve with a multiplication, which is mathematically a sliding dot product. The typical format for the hidden layers in the neural network begins with a convolutional layer and is then followed by a pooling layer. The first convolutional layer is responsible for capturing the Low-Level features such as edges, colour, gradient orientation, etc while the pooling is used to reduce the spatial size of the convolved feature and extract dominate features that can be used to classify the image. Once the convolution and pooling are completed the CNN goes into its classification phase which flattens the final output as a column vector and uses a series of fully connected layers to classify the images being inputted. The final output being the classification. In order to visualise this technique, the following diagram is provided:

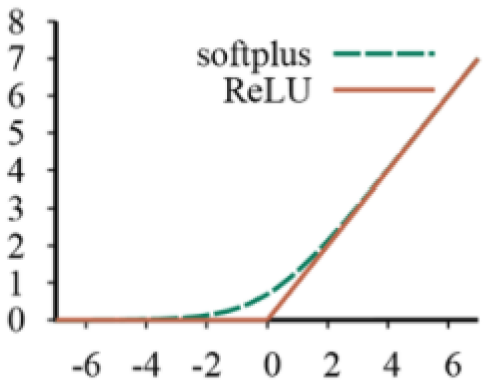


Russell S, Norvig, P. (2020)

In the convolution stage there is an activation function called RELU which stands for unit employing the rectifier. The activation function decides whether a node should be activated or not based on the calculated sum of the weighted inputs from the input function and further adding bias with it. For the RELU activation function the output node will only be activated if the calculated weighted sum is positive, otherwise, it will output zero. This method removes the issue of the vanishing gradient which is a problem in back propagation networks. The mathematical formula for the RELU activation function is as follows:

$$\mathrm{ReLU}(x) = \max(0, x).$$

To visualise this equation graphically the following diagram is provided:



Russell S, Norvig, P. (2020)

In the pooling stage there are two methods that the CNN uses in pooling being max pooling and average pooling. Essentially in the convolutional stage a window will slide across the image selecting the pixels within the window, for example a 3x3 window would select 9 pixels. These pixels then go into the pooling layer where either max or average pooling is used. Max pooling takes the maximum value from the portion of the image covered by the window, while average pooling returns the average of all the values from the portion of the image covered by the window. Max pooling typically works better. Once the process of convolution and pooling is completed the image goes through the classification process as discussed above.

How the AI technique is used to solve the problem:

The CNN was used to solve this problem as images of different types of clothing being t-shirt/top, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags and ankle boots were inputted into the neural network and the CNN had to classify what type of clothing was shown in the image. To set up the CNN the first step was to design the topology of the neural network. In this example two convolutional layers were used which took in one input channel as the images were black and white, used max pooling in the pooling stage, used RELU for the activation function and batch normalisation to optimize the algorithm improve the rate of convergence. For the first convolution layer the input was one as discussed above, the batch normalisation was 16 which results in 16 output channels, the window size is 3x3, the stride is 2 meaning how far the window is moved after each convolution layer and the padding was 1 meaning that 1 layer of pixels with a 0 value is added around the image this prevents the image from shrinking. For the max pooling layer, the window size is set to 2x2 which again focuses down on the pixels sent from the convolutional layer. Once the convolutional layer is finished the 16 nodes are sent to another convolutional layer where the process is repeated with the same variables except for a change in the input which is now 16 channels and the output which will be 32 channels.

After both convolutional layers are completed the fully connected layers must be designed. In this example there are 4 fully connected layers that attempt to flatten the inputted pixels with a column vector of 288. To understand how 288 was chosen, take the final number of output channels which is 32 and times it by the number of values in each channel which in this case will be 9 as the window size is 3x3. In the first layer the pixels will be flattened to 120 with a dropout rate of 0.5 which means half of the nodes in the fully connected layer will be removed. This is done to reduce the test-set error of a network to increase its ability of generalisation, this makes a network harder to fit the training set and will reduce overfitting. The fully connected layer once again uses the RELU activation function, and this process continues until the 288 inputted values are flattened to 20. The final output channels will be 9 due to there being 9 types with each type being assigned a digit.

Once the topology has been designed the CNN is constructed in the code and is ready to be trained. Before the CNN is trained the dataset goes through resizing in this case the image size is 28 by 28, and the data is broken up into batches of 64 to be displayed. Once the data has been loaded the CNN can be trained to preform this the EPOCH and learning rate must be set. The EPOCH is how many times the network goes through the entire data set, in this case it is set to 3 to thoroughly train the data but to prevent overfitting. The learning rate is what updates the weights of the inputted values based on the estimated error each time the model weights are updated. The key is to have a learning rate that isn’t too low so that the CNN improves as the training continues but isn’t too high so that the weights aren’t drastically changed, and the loss function diverges. In this case the learning rate is set to 0.001 which after trial and error produced the best results. After the training is complete the CNN is tested against the test data and an accuracy to the solution is given.

Andreas Gwyther-Gouriotis 13893627

Solution Results:

The final accuracy of the CNN was 0.8945833333333333 or 89%. This is a reasonably high accurate model which once tested against a random image provided the following results:

Actual results: [8 3 5 4 0 5 9 6 7 8 2 0 2 4 3 4 9 4 2 8]

Predicted results: [8 3 5 4 4 5 9 6 7 8 4 0 2 4 3 4 9 4 2 8]

The results show that there was a minimal amount of error with only 2 mistakes being made out of 20 predictions. In order to improve these results more testing could be done on the learning rate to find a rate that provides even better loss reduction. More convolutional layers could be added and an overall better design to the CNN could be made as there are a lot of variables that can be changed such as the dropout rate as well as the window and pooling sizes. Since this problem involved images, the CNN was the appropriate neural network to use as it preforms the best out of all the pure artificial intelligence techniques. Overall, this satisfies the problem statement to high degree and will help the fashion design company analyse what people are wearing this summer.

References:

Lu, H. 2021, 'Module 3 -- AI techniques for ability of learning -- Part II", UTS Online Subject 41040 lecture notes, UTS, Sydney, viewed 24 September 2021, < https://canvas.uts.edu.au/courses/18678/modules>.

Russell, S. J., Norvig, P. (2020). Artificial intelligence: A modern approach, 4th Edition. Pearson.

Zalandoresearch. "Zalandoresearch/Fashion-Mnist: A MNIST-like Fashion Product Database. Benchmark." GitHub, github.com/zalandoresearch/fashion-mnist.