

Part 1 Answers

2a) Train images represents an array of training images, the first value 60000 is the number of images in the array, and the next two values 28x28 represent the size of the image. Training labels just have 60000 elements, all of which are the class of the images. Test images and labels have the same setup as the training arrays, although they only have 10000 images.

3a) The images are fairly pixelated as they are (after looking this up) apparently only 28 x 28 pixels, which isn't a whole lot of data. The images are also all on a gray scale, which makes it harder for me to distinguish some things, especially because white on a clothing item looks just like the background.. Personally, I think I would've misclassified a couple of pictures, such as the bag, one of the tops I would've labeled a dress, and one pair of sandals that looked like a sneaker.

Part 2 Answers

1b) I chose 64, 128, and 256 as my activation units because those are all multiples of 64 which is one stack of items in Minecraft. (Clearly mine wasn't based on any technical reasons!)

Part 3 Answers

1c) The Training and Validation loss graph indicates there may be some overfitting as Validation Loss starts to increase at the end, but training loss is on a constant decrease. This is also somewhat supported by the second graph, as there is a decent gap between Training Accuracy and Validation accuracy, which would also indicate some overfitting.

4a) What stands out is just how much worse/better the model is at classifying different clothing items. For example, my f-score for trousers was 0.98, which is very high, but for shirts it was only 0.70.

5a) My immediate takeaway is that t-shirts/tops and shirts are much lighter than trousers or bags. I think the reason for this is that shirts have a lot more that they look like shape wise, whereas pants are the only clothing item that has the same shape with the two long legs. A shirt could look like a jacket, dress, pullover, or just about any other kind of item you wear on your torso.

5b) Well, the model was a lot better at predicting bags, but that's probably because I just wouldn't have known that was an option. I was right when I said that one of the shirts/dresses I would've mixed up, as shirts were a really low performing category for the model. One thing I was surprised by though is that the model was really accurate for sandals, and I didn't think it would be.

6) I would be pretty satisfied, an accuracy of almost 90%, with many items in the mid-high 90s for f1-score is a pretty solid model. For the next run, I would try changing the activation units on

my hidden layers, and maybe getting rid of or adding one, to try and see if that results in a more accurate model.