Reed McElfresh
IST 718: Big Data Analytics
Lab 9

The code for this paper can be found at: https://github.com/rmcelfresh/IST718Lab9

Introduction

For year the MNIST database was the gold standard for testing computer vision. However, as machine learning has improved, the utility of MNIST has decreased. Many algorithms can achieve 97% accuracy on the database and some have gotten up to 99.7%. In response, Zalando Research has created a next generation MNIST database which has the same structure as MNIST however instead of hand written numbers, the new database includes 70,000 pictures of clothing of ten different classes. This paper will introduce two methods of machine learning that were used to attempt to identify what type of clothing is represented in each picture.

Analysis

The database is split into a training set consisting of 60,000 images and 10,000 images in the test set. In the database there are ten different varieties of clothing represented: T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. The data was accessed from two locations: the Keras data from ?? and the Zalando Research GitHub repository.

A sample of each variety of clothing is show in Image 1. Image 2 shows the differences within the sandal class. Image 3 shows the differences within the sneaker class.

Image 1: Sample images of each type of clothing included in the database.

Image 2: Sample images of sandals

Image 3: Samples of different sneakers

A review of these sample images demonstrates why this database is harder for algorithms to process than the original MNIST database. A computer is going to have a more difficult time to differentiate between a shoe and sandal than it would differentiating among hand drawn numbers. The key to processing will be the relative pixel darkness. Image 4 demonstrates the relative pixel density of a sample of images.

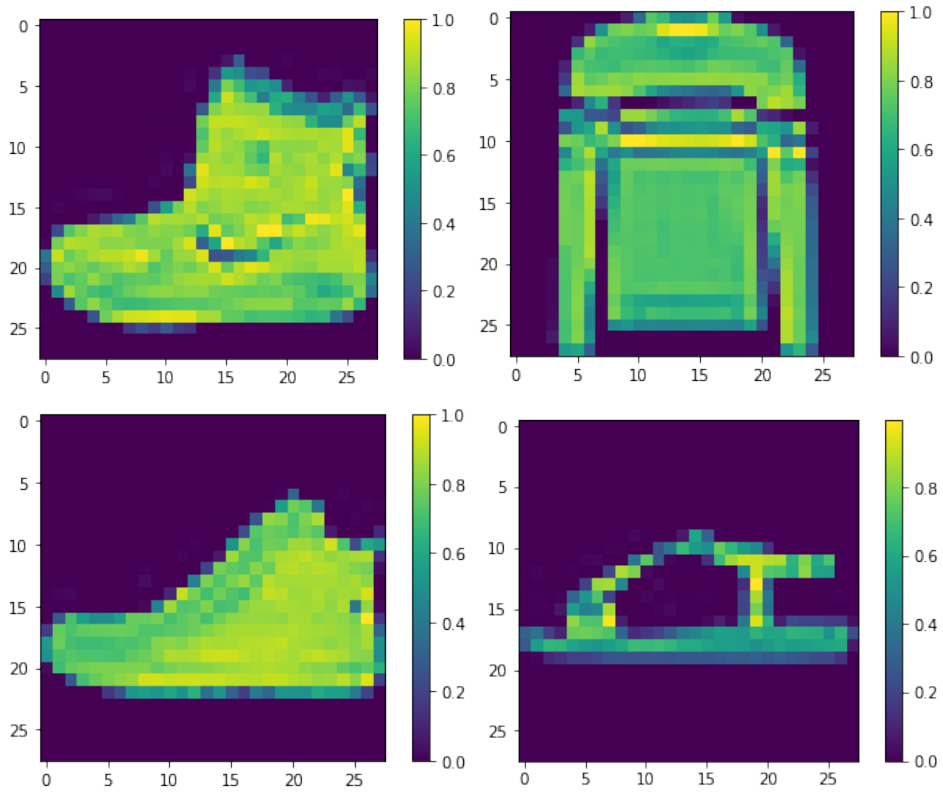


Image 4: Images demonstrating the relative pixel density of four images from the database.

A random forest model and a Keras model were created to categorize the clothing items. The Random Forest model was adopted from https://www.kaggle.com/atorin/mnist-digit-recognition-with-random-forests and the Keras code was adopted from https://www.tensorflow.org/tutorials/keras/basic_classification

Results

Model	Specifications	Time to Run (seconds)	Accuracy on Training	Accuracy on Test
Random Forest	10 Estimators	13.7	100%	85.17%
Random Forest	20 Estimators	28.7	100%	85.25%
Random Forest	50 Estimators	70	100%	87.44%
Random Forest	100 Estimators	139	100%	87.36%
Random Forest	50 Estimators; Criterion = Entropy	88	100%	87.49%
Keras	Activation = Relu & Softmax; Compiler optimizer = adam, loss= sprase_categorica_crossentropy; 10 epochs	4.93	100%	88.7%
Keras	Activation = Relu & Softmax; Compiler optimizer = adam, loss= sprase_categorica_crossentropy; 300 epochs	141	99%	88.25%

All of these models were significantly better predicting on the training data than they were on the test data—all were 99% accurate or better on the training data but less than 90% on the test data. However, this difference was immediately present, for the 10 epoch Keras model, the difference was occurring as seen in Figure 7.

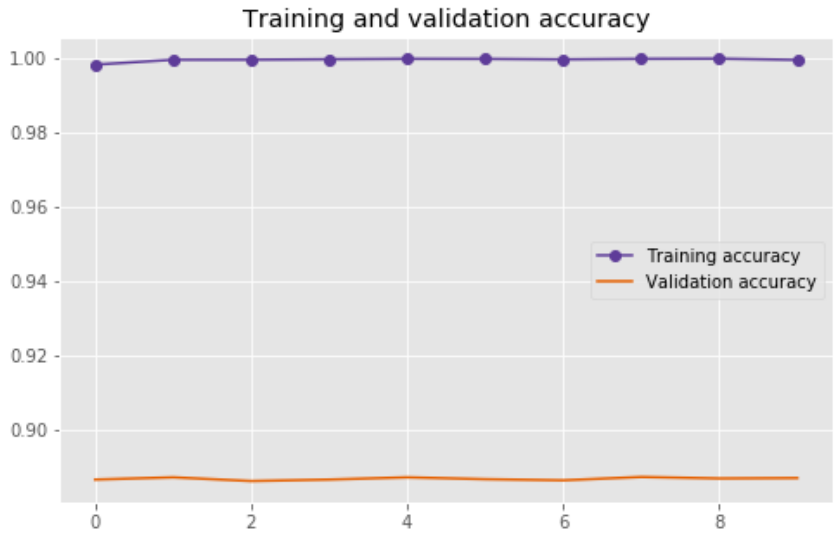


Figure 7: Comparison of testing and training accuracy of 10 Epoch Keras model showing the immediate difference in accuracy.

The Keras models had better accuracy than any permutation of the RandomForest models. The first Keras model was also the fastest to do the predictions, taking less than 5 seconds. The best RandomForest model took 13.7 seconds. Building a true Tensor Flow model without the Keras front end would dramatically improve the processing time.

The confusion matrices for the final RandomForest model and the final Keras model are show as images 5 and 6. Both were fairly accurate identifying trousers accurately, however they both struggled differentiating between the different varieties of shoes, between coats and pull overs, and between t-shirts/tops and shirts.

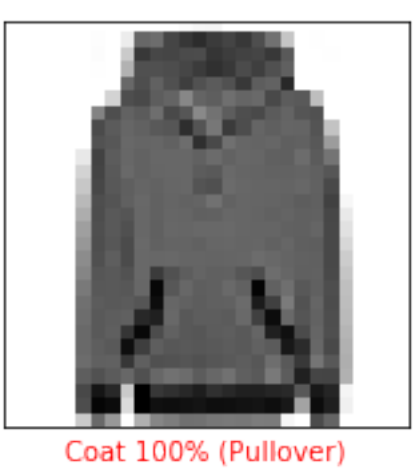
		True Label										
	Row Labels	**Ankle boot**	**Bag**	**Coat**	**Dress**	**Pullover**	**Sandal**	**Shirt**	**Sneaker**	**Trouser**	**T-shirt/top**	**Grand Total**
Predicted Label	Ankle boot	944	1			1	20	1	16			983
	Bag		960	3	3	1		12		2	9	990
	Coat		3	836	39	92		80		5	6	1061
	Dress	2	5	26	876	8	3	27		12	22	981
	Pullover		3	70	15	807		75			17	987
	Sandal	6	3	1		1	935		5		2	953
	Shirt	1	9	63	31	66	1	674			102	947
	Sneaker	46	8				41		979		1	1075
	Trouser		1		11	2		1		975	2	992
	T-shirt/top	1	7	1	25	22		130		6	839	1031
	Grand Total	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**10000**

Image 5: Keras Model Confusion Matrix

		True Label										
		Ankle boot	**Bag**	**Coat**	**Dress**	**Pullover**	**Sandal**	**Shirt**	**Sneaker**	**Trouser**	**T-shirt/top**	**Grand Total**
Predicted Label	Ankle boot	949	1		1		12		32			995
	Bag	2	973	1	1	5	1	17		2	12	1014
	Coat		6	811	28	120		87		5	3	1060
	Dress		3	43	906	10	1	32		21	29	1045
	Pullover		2	97	14	799		126		1	12	1051
	Sandal	8	2				954		12		1	977
	Shirt		8	46	27	53		577		4	84	799
	Sneaker	41	4				32		956			1033
	Trouser		1	1	3			1		965		971
	T-shirt/top			1	20	13		160		2	859	1055
	Grand Total	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**10000**

Image 6: Confusion Matrix for the RandomForest Model.

The Keras model forms its prediction based on the likelihood of it being one type of clothing.Sometimes the final Keras model was 100% confident of its recommendation, even when it was wrong as shown in Image 8 below:



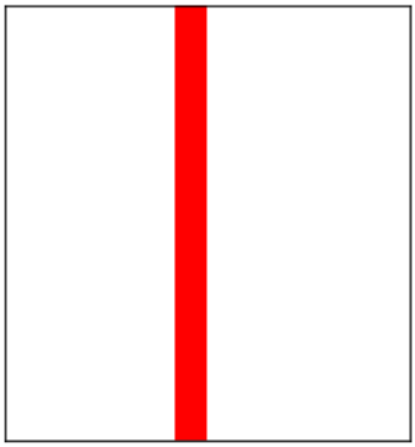


Image 8: Incorrect prediction of a pullover.

Image 9 shows a larger sample of the predictions. Most were strong, only identifying one variety of clothing, however others were more unsure.

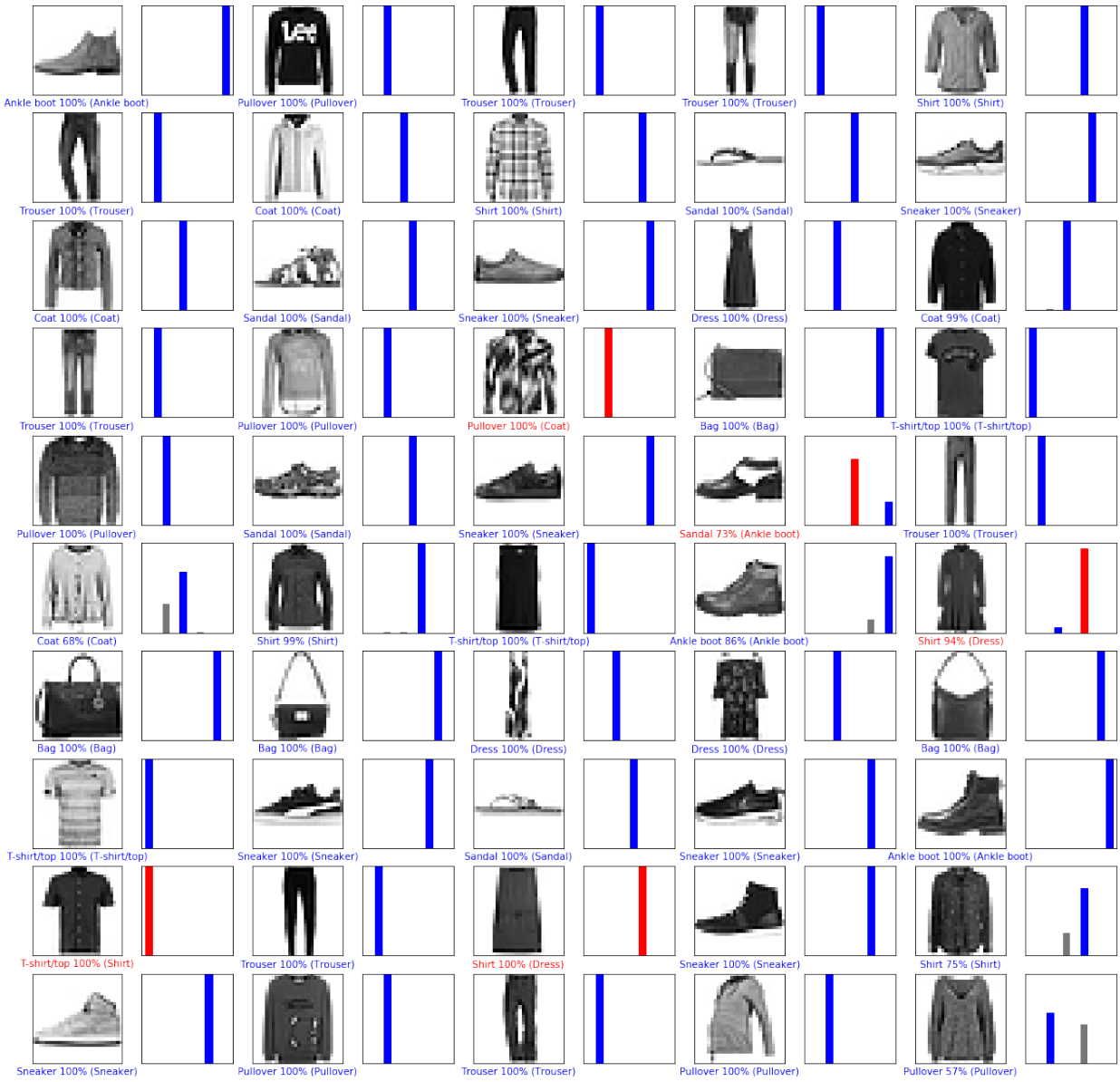


Image 9: Sample of prediction accuracy measures.

Conclusion

The Keras model would be the model that I would recommend to use in a production environment. It is slightly more accurate and depending on the number of epochs used, it can be quicker than the less accurate RandomForest model. Neither model is accurate enough to use consistently, however the first Keras Model could be adopted to scrape a website and determine the number of different varieties of clothing they were using, if we can accept that ~11% of the predictions would be wrong.