Math 10 Final report

Team member: JUNYI LI (11809157), Sean Long (93188739) Alimu Alibotaiken(54084203)

Classification Model	Accuracy
K-Nearest Neighbor	0.90240
Linear Model	0.90640
Support Vector Machine	We do not know
Neural Network	0.098
Kera	0.96

To increase the accuracy of prediction, we try many different algorithms to train the machine and predict by test data. X_train is imported from kmnist-train-imgs.npz and y_train is imported from kmnist-train-labels.npz. After we imported these data, the original shape of X_train is (60000,28,28). Then we use reshape function to change the size to (60000,784).

K-Nearest Neighbor: we directly import KNeighborsClassifier from sklearn.neighbors and set the value of k is 5 to implement the training data. Them we use .fit in KNeighborsClassifier to train the machine by importing X_train and y_train. At the end, the accuracy of K-Nearest Neighbor is 0.90.

Linear Model: To train the machine, we imported LogisticRegression from sklearn.linear_model. Before we trained, we set solver to lbfgs. Due to the size of training data, the accuracy of the alogthrim is not pretty high. Compare with another algorithm is normal.

Others: We tried every algorithm we can think, but we cannot improve the accuracy exceed 0.91. So, we use Kera function to train. Firstly, we transfer y_train to one-hot size so that it is convenience for following calculate. We import Sequential model to construct our neural

network to study mnist. Due to it is a 2*2 matrix, so that we add Convolution2D and MaxPooling2D. And then, we add adam optimizer to optimize the accuracy of the algorithm. The learning rate is 1e^-4. There are too many variables in x, so that we add dropout 0.5 in our neural network. To minimize the loss function, we increase the number of epochs so that this algorithm will run for long time to training the machine. When epochs are 20, the accuracy is 0.95.

Conclusion: In conclusion, it is significant to find the keras and neural network have a good performance in the kmnist study. Compare with other different algorithm, the size of training data they can accept is bigger and they can process efficiency.