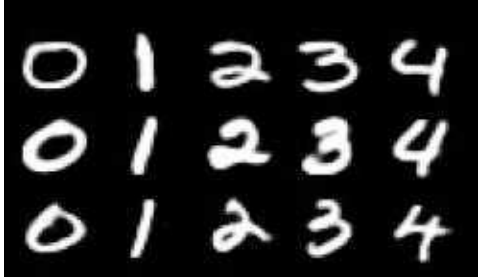


Homework# 6 - Due day 11:59PM, 2nd January, 2018

- Support Vector Machine and Visualization on 2D space
 - ▶ Use SVM models to tackle classification on images of hand-written digits (digit class only ranges from 0 to 4, as figure shown on right).
 - ▶ You are highly suggested to use LIBSVM library, with C++/Matlab/Python supports. There are fruitful official introductions or plenty materials you can find on internet, feel free to use them.
 - ▶ Training data and testing data are both provided:
 - ➔ training:
 - **X_train.csv** is a 5000x784 matrix. Every row corresponds to a 28x28 gray-scale image
 - **T_train.csv** is a 5000x1 matrix, which records the class of the training samples
 - ➔ training:
 - **X_test.csv** is a 2500x784 matrix. Every row corresponds to a 28x28 gray-scale image
 - **T_test.csv** is a 2500x1 matrix, which records the class of the test samples

Homework# 6 - Due day 11:59PM, 2nd January, 2018

- Support Vector Machine and Visualization on 2D space
 - You can choose to use any kernel functions you like (libsvm provides linear, polynomial, and RBF kernels).
 - If you choose to use libsvm, please use C-SVC (you can choose by setting parameters in the function input, C-SVC is soft-margin SVM)

25
bonus

Since there are some parameters you need to tune for, please check how to do grid search for finding parameters of best performing model. For instance, in C-SVC you have a parameter C , and if you use RBF kernel you have another parameter γ , you can search for a set of (C, γ) which gives you best performance in cross-validation. (lots of sources on internet, just google for it)

- Choose one of the models you've trained to do visualization.

25
points

- (first use PCA to project all your data onto 2D space
- use different colors to draw samples with different digit class

25
points

- With all the data samples are shown by “dots”, the “support vectors” should be shown with different symbols, e.g. square, triangle, cross (in libsvm, to get support vectors please refer to [here](#))

50
bonus

- You can also try to plot decision boundary:)

75
points

- submit a report: what you have done (code, testing performance), what you have visualized, what you have learned.