**Friday March 15- with Dr.Diggans (mentor)**

- Last time: train 100 CNNs with 100 subsamples (each of them containing 5000 images). Extract subsample feature spaces using CNN embedding function. Attain singular values and following entropy values.

**Other possible approaches:**

- Dr.Diggans: input limits into the BSIE function other than input singular vectors to attain the entropy value. (for more details read Dr.Diggans's email and comments on BSIE code (the min and max of the singular values of the whole training set)) That is, provide min singular value (can be assumed as 0) and max singular value (of the whole training set) boundary inputs.
- If take a smaller subsample (meaning smaller variance among data), then singular values can be smaller correspondingly.
- Singular values can tell you many things. But if related to PCA, we can take the first few largest singular values (low order approximations of high dimensional space), and the following or the rest tiny singular values are basically just noise. Their directions in the high dimensional spaces don't have much information in them.

- So, it might be interesting to just look at the first k largest singular values in one singular vector after applying SVD to the feature matrix. We can get much larger and diverse entropy values after inputting them through BSIE.

- In this case, we have 64 features in the feature vector space (not the middle diagonal matrix anymore), we can look at the first 5 singular vectors. These vectors have the most variance. For example, the first singular vector that has the highest variance, we use this most important component in our coordinate system (multiply by its corresponding singular values from the middle diagonal matrix) and the second singular vector multiplied by its singular value (second coordinate).

- SVD: besides middle diagonal matrix, also have U and V (feature space and data space). We want to look at things in the feature space to extract vectors from it. It could be the rows of U that are the vectors we are interested in. For example, the

first row of U is the singular vector (or feature vector) that goes along with the first singular value.

- What is the projection of the first singular vector of my sample to the first singular vector of the training set. That is, the projection of one vector onto another vector.

- we can use the projection of my data onto that singular vector to create a smaller coordinate system using dot product.

Next week:
- Change the lower and upper bound of singular values in BSIE function input.
- Try to increase each subsample size from 5000 to 10000 to 20000.
- Expected the test accuracy of models trained with subsamples increases if the entropy values of these subsamples are close to the entropy value of the training set.