# CS480 Assignment 2

**Name:** Gogulashanth Sathiyaseelan, **Student ID:** 20554735

## Exercise 1

Boosted trees with feature extraction (pre-processing) was used to classify the MNIST dataset.

DAISY feature descriptors were first extracted from each individual images. The DAISY detector/descriptor works as follows:

1. Apply Gaussian blur to an image
2. Compute magnitude and direction of gradients for each patch in an image to create an orientation histogram bin
3. Gaussian averaging is used to average the feature descriptor for each location in the image.
4. The feature descriptors are normalized and PCA applied for dimensionality reduction.
5. Quantization is applied to reduce the dynamic range of the feature descriptors.

These DAISY descriptors are used to train a machine learning model. Adaboost is used with random forest as a base classifier. The depth and number of estimators of random forest is limited so that it functions as a 'weak learner' in the Adaboost algorithm. Adaboost algorithm is run with a max number of iterations as 400 (for performance reasons). Then, a set of DAISY descriptors are extracted for the test set and then predicted using the learned model.

## Exercise 2

$\epsilon_{t+1}(h_t) = \frac{1}{2}$. Proof:

$$\epsilon_{t+1}(h_t) = \sum_{i=1}^{n} p_i^{t+1} \cdot |h_t(\mathbf{x}_i) - y_i|$$

Since $(h_t(\mathbf{x}_i), y_i) \in 0, 1$ and $p_i^t = \frac{w_i^t}{\sum_{j=1}^n w_j^t}$,

$$\epsilon_{t+1}(h_t) = \sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^{t+1} = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} w_j^{t+1}}{\sum_{i=1}^n w_i^{t+1}}$$

Noting that,

$$w_i^{t+1} = w_i^t \beta_t^{1-|h_t(\mathbf{x}_i)-y_i|} = \begin{cases} w_i^t & h_t(\mathbf{x}_i) \neq y_i \\ \beta_t w_i^t & h_t(\mathbf{x}_i) = y_i \end{cases} \tag{1}$$

We can write

$$\epsilon_{t+1}(h_t) = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} w_j^t}{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} w_j^t + \beta_t \sum_{k:h_t(\mathbf{x}_k)=y_k} w_k^t} \tag{2}$$

Dividing top and bottom of equation(2) by $\sum_{i=1}^n w_i^t$,

$$\epsilon_{t+1}(h_t) = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} w_j^t / \sum_{i=1}^n w_i^t}{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} w_j^t / \sum_{i=1}^n w_i^t + \beta_t \sum_{k:h_t(\mathbf{x}_k)=y_k} w_k^t / \sum_{i=1}^n w_i^t} = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t}{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t + \beta_t \sum_{k:h_t(\mathbf{x}_k)=y_k} p_k^t} \tag{3}$$

Also,

$$\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$$

Substituting, $\epsilon_t = \epsilon_t(h_t) = \sum_{i=1}^{n} p_i^t \cdot |h_t(\mathbf{x}_i) - y_i|$ and using the fact, $\sum_i^n p_i^t = \sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t + \sum_{k:h_t(\mathbf{x}_k)=y_k} p_k^t = 1$,

$$\beta_t = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t}{1-\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t} = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t}{\sum_{k:h_t(\mathbf{x}_k)=y_k} p_k^t} \tag{4}$$

Substituting equation(4) into equation(3),

$$\epsilon_{t+1}(h_t) = \frac{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t}{\sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t + \sum_{j:h_t(\mathbf{x}_j)\neq y_j} p_j^t} = \frac{1}{2} \tag{5}$$