Standardizing the Data

Support Vector Machines (SVMs) are known to be very sensitive to the initial data formatting. Thus, your first task will be to write a function that standardizes the provided feature vectors. More specifically, as an input, your function will take a feature matrix X, where every row in the matrix depicts a particular data point, and every column denotes a specific feature. Then, your function should return an updated feature matrix as:

$$X_{new} = \frac{X - \mu}{\sigma}$$

where μ denotes the mean across the rows of matrix X, and σ depicts the standard deviation across the rows of matrix X. Note that μ and σ are vectors of the same dimensionality as there are number of features in the dataset (e.g. the number of columns in the matrix X). Such a normalization ensures that the provided features are centered around 0, and that their standard deviation is 1.

Your Function

Save Reset MATLAB Documentation (https://www.mathworks.com/help/)

```
function X_new=StandardizeData(X)
   % standardizes data
   %
   % Input:
   % - X: an n x d dimensional feature matrix where n is the number of observations, and d is the number of fea
   % Output:
   % - X_new: an n x d dimensional normalized feature matrix.

   % get number of rows
   r = size(X,1);

   mu = mean(X,1); %matlab mean(X,0,1);
   mu_mat = repmat(mu,r,1);

   % get standard dev of each feature across rows
   sig = std(X,1); %matlab std(X,0,1);
   sig_mat = repmat(sig,r,1);

   X_new= (X - mu_mat)./sig_mat;

end

```

Code to call your function

Reset

```
load('ImageDataTrain.mat');
Xtrain=StandardizeData(data.trainX);

load('ImageDataTest.mat');
Xtest=StandardizeData(data.testX);
```

▶ Run Function

Previous Assessment: All Tests Passed

Submit

Is Standardization Correct for Training Data?