**IT4060 – Machine Learning**

**Lab 8 – Unsupervised learning**

**Exercise 1: K-means algorithm**

1. Insert code to randomly initialize the clusters in the kMeansInitCentroids.m.

randidx = randperm(size(X, 1));

centroids = X(randidx(1:K), :);

1. Add code to assign the closest centroid to each data point in findClosestCentroids.m.

for i = 1:length(X)

minDist = Inf;

for j = 1:K

dist = norm(X(i,:) - centroids(j,:)) ^ 2;

if dist < minDist

minDist = dist;

idx(i) = j;

end

end

end

1. Add the following code to compute the next set of centroids for the next iteration in computeCentroids.m.

for i = 1:K

count = 0;

for j = 1:m

if idx(j) == i

count = count + 1;

centroids(i,:) = centroids(i,:) + X(j,:);

end

end

centroids(i,:) = centroids(i,:) / count;

end

1. Run the ex7.m file to run the k-means algorithm.

**Exercise 2: Principle Component analysis (PCA)**

1. Insert the following code into projectData.m file.

U\_reduced = U(:, 1:K);

Z = X \* U\_reduced;

1. Insert the following code into the recoverData.m file.

U\_reduced = U(:, 1:K);

X\_rec = Z \* U\_reduced';

1. Insert the following code to pca.m file.

Sigma = (X' \* X) / m;

[U, S, V] = svd(Sigma);

1. Run the ex7\_pca.m file to run the PCA algorithm.

**Submission:**

Upload the code of both exercises as a single zip file to the courseweb link. The file name should be your registration number.

If your submission is greater than 10 MB, you may use dropbox to upload the submission and the link to the shared folder can be mentioned in a text file that should be uploaded to courseweb. The text file name should be the registration no.