

第一部分 K均值

K均值第一步： 将所有样本X按最接近的簇编号分组

```
distance = zeros(K, 1);  
for i = 1:size(X,1),  
    for j = 1:K,  
        distance(j) = sum((X(i,:) - centroids(j,:)).^2);    求X到每一个centroids的距离并存入矩阵distance  
    end  
    idx(i) = find(distance == min(distance));    最小值所在行的行列数即x的最小簇编号  
end
```

K均值第二步： 求每个分组的均值中心，并将中心点移动到新的位置

```
for i = 1 : K,    一共有K个中心点，即数据被分为K组  
    centroids(i,:) = mean(X(find(idx == i), :));    在上一步中所有的X均被编上了号，即idx  
end
```

均值中心的随机初始化：

% Initialize the centroids to be random examples

% Randomly reorder the indices of examples

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

% Take the first K examples as centroids

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

第二部分 PCA

第一步，得出主成分矩阵U

```
sigma = (1/m) .* (X' * X);  
[U, S, V] = svd(sigma);
```

第二步，得出映射后的矩阵Z

```
for i = 1:size(X,1),  
    x = X(i,:);  
    for k = 1:K,  
        Z(i,k) = x' * U(:, k);  
    end
```

end

这里为什么写成 $Z = X * U(:, K)$;不行我也不知道