ECE421 Assignment 1

Hshmat Sahak
1005903710

1) See code attached, either ipynb file or end of this pdf

2) The code below is responsible for importing the UCI ML Breast Cancer dataset:

```
from sklearn.datasets import load_breast_cancer
dataset = load_breast_cancer()
```

The resulting "dataset" variable is a dictionary with keys data, target, feature_names, and more.
Since this is an unsupervised learning task, we only require the input data (input features for all training examples), not any target values. So, the data we need is only dataset.data.
We also need to pass in the number of clusters k. In this problem, k changes from 2 to 7.
So, we call as follows (note: this is pseudocode):

```
data = dataset.data
distortions = []
for k = 2,3,...,7:
    centroids, assignments = kmeans(data, k)
    //calculate distortion
    distortions.append(distortions)
```

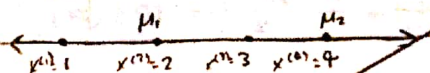
3) See code and figure attached, either ipynb file or end of this pdf

4) I would pick k = 5. From the figure in problem (3), cost decreases as k increases, but seems to hit an inflection point at k = 5. Values of k above 5 minimize cost by relatively little, so they may overfit. Values of k below 5 are associated with high costs, so they will underfit. k = 5 seems right. With 569 training examples, each cluster is expected to have ≈ 110 datapoints on average

Problem 2

IGNORE!

1) A proof that something is "not" true by a counter-example is a valid analytical proof.
Consider $k = 2$ and dataset made of 4 points in $\mathbb{R}$: $x^{(1)} = 1$, $x^{(2)} = 2$, $x^{(3)} = 3$, $x^{(4)} = 4$. Initialize k-means with centroids $\mu_1 = 2$, $\mu_2 = 4$ and assume if $x^{(i)}$ is equally distant to multiple centroids μ_k, the point will be assigned to the centroid whose index is smallest, i.e. k with the smallest value for $k \in \arg\min_k \|x^{(i)} - \mu_k\|^2$.



Running k means with this initialization:

due to assumption stated above

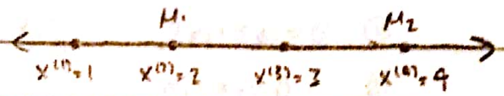
Cluster assignment: $c^1 = 1$, $c^2 = 1$, $c^3 = 1$, $c^4 = 2$

Move centroids: $\mu_1 = \frac{1+2+3}{3} = 2$, $\mu_2 = \frac{4}{1} = 4$

Cluster centroids have not changed, so algorithm terminates.

Problem 2)

1) A proof that something is "not" true by a counter-example is a valid analytical proof.
Consider $k=2$ and dataset made of 4 points in $\mathbb{R}$: $x^{(1)}=1, x^{(2)}=2, x^{(3)}=3, x^{(4)}=4$. Initialize k-means with centroids $\mu_1=2, \mu_2=4$ and assume if $x^{(i)}$ is equally distant to multiple centroids μ_k, the point will be assigned to the centroid whose index is smallest, i.e. k with the smallest value for $k \in \arg\min_k \|x^{(i)}-\mu_k\|^2$.



Running k means with this initialization:

Cluster Assignment: $c^1=1$ since $x^{(1)}=1$ closer to $\mu_1=2$ than $\mu_2=4$

$c^2=1$ since $x^{(2)}=2$ closer to $\mu_1=2$ than $\mu_2=4$

$c^3=1$ since $x^{(3)}=3$ same distance to μ_1 and μ_2, but select μ_1 since index convention

$c^4=2$ since $x^{(4)}=4$ closer to $\mu_2=4$ than $\mu_1=2$

Move centroids: $\mu_1=\frac{1+2+3}{3}=2$, $\mu_2=4/1=4$

Cluster centroids have not changed, so algorithm terminates

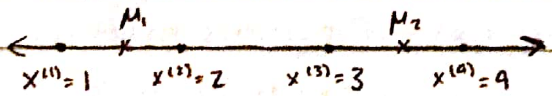
The distortion is: $J=\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}-\mu_{c^{(i)}}\|^2$

For our current cluster centroids: $J_1=\frac{1}{4}\sum\|x^{(i)}-\mu_{c^{(i)}}\|^2=\frac{1}{4}\left((x^{(1)}-\mu_{c^{(1)}})^2+\dots+(x^{(4)}-\mu_{c^{(4)}})^2\right)$

$$=\frac{1}{4}\left((1-2)^2+(2-2)^2+(3-2)^2+(4-4)^2\right)$$

$$=\frac{1}{4}(1+0+1+0)=\frac{1}{2}$$

To show this is not globally optimal solution, we will start with a different initialization and show this converges to a smaller distortion.



Cluster Assignment: $c^{(1)}=1$, $c^{(2)}=1$, $c^{(3)}=2$, $c^{(4)}=2$ (based on minimum distances)

Move centroids: $\mu_1=\frac{1+2}{2}=1.5$ $\quad\mu_2=\frac{3+4}{2}=3.5$

The distortion in this case is:

$$J_2=\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}-\mu_{c^{(i)}}\|^2$$

$$=\frac{1}{4}\left((x^{(1)}-\mu_{c^{(1)}})^2+\dots+(x^{(4)}-\mu_{c^{(4)}})^2\right)$$

$$=\frac{1}{4}\left((\tfrac{1}{2})^2\times 4\right)$$

$$=\frac{1}{4}$$

Since $J_2=\frac{1}{4}<\frac{1}{2}=J_1$, the cluster centroids that we obtain in the first case (after the algorithm ends) have higher distortion than this case, so are not a globally optimal solution.

Problem 3) SVD

1) The rank of A is the number of linearly independent rows of A

Since row 2 cannot be written as a scalar multiple of row 1, the rows of matrix A are linearly independent.

Alternatively, we can show the rows are linearly independent by showing:

$$\alpha_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = 0 \quad \text{implies} \quad \alpha_1 = \alpha_2 = 0.$$

Do do this, we have:

$$\alpha_1 + 2\alpha_2 = 0 \quad ①$$

$$2\alpha_1 + 3\alpha_2 = 0 \quad ②$$

$$\alpha_1 + \alpha_2 = 0 \quad ③$$

① - ③: $\boxed{\alpha_2 = 0}$

② ③: $\alpha_1 + 2\alpha_2 = 0 \Rightarrow \alpha_1 + 2(0) = 0 \Rightarrow \boxed{\alpha_1 = 0}$

So, the rows are independent. 2 linearly independent rows means $\boxed{\text{rank}(A) = 2}$

2) $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 1 & 1 \end{bmatrix}$

$$A^T A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 8 & 3 \\ 8 & 13 & 5 \\ 3 & 5 & 2 \end{bmatrix}$$

Eigenvalues of $A^T A$:

Set: $\det(A^T A - \lambda I) = 0$

$$\det\left(\begin{bmatrix} 5-\lambda & 8 & 3 \\ 8 & 13-\lambda & 5 \\ 3 & 5 & 2-\lambda \end{bmatrix} \right) = 0$$

$$(5-\lambda)\det\left(\begin{bmatrix} 13-\lambda & 5 \\ 5 & 2-\lambda \end{bmatrix}\right) - 8\det\left(\begin{bmatrix} 8 & 5 \\ 3 & 2-\lambda \end{bmatrix}\right) + 3\det\left(\begin{bmatrix} 8 & 13-\lambda \\ 3 & 5 \end{bmatrix}\right) = 0$$

$$(5-\lambda)[(13-\lambda)(2-\lambda) - 25] - 8(8(2-\lambda) - 15) + 3(40 - 3(13-\lambda)) = 0$$

$$(5-\lambda)(26 - 15\lambda + \lambda^2 - 25) - 8(16 - 8\lambda - 15) + 3(40 - 39 + 3\lambda) = 0$$

$$(5-\lambda)(\lambda^2 - 15\lambda + 1) - 8(1 - 8\lambda) + 3(1 + 3\lambda) = 0$$

$$5\lambda^2 - 75\lambda + 5 - \lambda^3 + 15\lambda^2 - \lambda - 8 + 64\lambda + 3 + 9\lambda = 0$$

$$-\lambda^3 + 20\lambda^2 - 3\lambda = 0$$

$$\lambda^3 - 20\lambda^2 + 3\lambda = 0$$

$$\lambda(\lambda^2 - 20\lambda + 3) = 0$$

$$\lambda = 0 \quad \text{or} \quad \lambda = \frac{20 \pm \sqrt{20^2 - 4(1 \times 3)}}{2} = \frac{20 \pm \sqrt{400-12}}{2} = 10 \pm \sqrt{97}$$

Singular values are square root of nonzero eigenvalues:

$$\boxed{\begin{aligned} \sigma_1 &= \sqrt{10 + \sqrt{97}} \\ \sigma_2 &= \sqrt{10 - \sqrt{97}} \end{aligned}}$$

3) $A = U\Sigma V^T$

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^2 V^T$$

We have σ_1 and σ_2. So, we can write $\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{10+\sqrt{97}} & 0 & 0 \\ 0 & \sqrt{10-\sqrt{97}} & 0 \end{bmatrix}$

The columns of V are eigenvectors of A^TA

Let's first find eigenvectors associated with eigenvalues found in (2):

$\lambda_1 = 10+\sqrt{97}$

$$A - \lambda_1 I = \begin{bmatrix} 5-\lambda & 8 & 3 \\ 8 & 13-\lambda & 5 \\ 3 & 5 & 2-\lambda \end{bmatrix} = \begin{bmatrix} -5-\sqrt{97} & 8 & 3 \\ 8 & 3-\sqrt{97} & 5 \\ 3 & 5 & -8-\sqrt{97} \end{bmatrix}$$

Compute row-reduced form:

Divide row 1 by $-\sqrt{97}-5$: $R_1 = \frac{R_1}{-\sqrt{97}-5}$

$$\begin{bmatrix} 1 & \frac{5-\sqrt{97}}{9} & \frac{5-\sqrt{97}}{24} \\ 8 & 3-\sqrt{97} & 5 \\ 3 & 5 & -\sqrt{97}-8 \end{bmatrix}$$

Subtract row 1 multiplied by 8 from row 2: $R_2 = R_2 - 8R_1$

$$\begin{bmatrix} 1 & \frac{5-\sqrt{97}}{9} & \frac{5-\sqrt{97}}{24} \\ 0 & \frac{-\sqrt{97}-13}{9} & \frac{\sqrt{97}+10}{3} \\ 0 & \frac{\sqrt{97}+10}{3} & \frac{-7\sqrt{97}-69}{8} \end{bmatrix}$$

Multiply row 2 by $-\frac{9}{\sqrt{97}+13}$: $R_2 = -\frac{9}{\sqrt{97}+13}R_2$

$$\begin{bmatrix} 1 & \frac{5-\sqrt{97}}{9} & \frac{5-\sqrt{97}}{24} \\ 0 & 1 & \frac{-\sqrt{97}-11}{8} \\ 0 & \frac{\sqrt{97}+10}{3} & \frac{-7\sqrt{97}-60}{8} \end{bmatrix}$$

Add row 2 multiplied by $\frac{-5+\sqrt{97}}{9}$ to row 1: $R_1 = R_1 + \frac{-5+\sqrt{97}}{9}R_2$

$$\begin{bmatrix} 1 & 0 & \frac{-3-\sqrt{97}}{8} \\ 0 & 1 & \frac{-\sqrt{97}-11}{8} \\ 0 & \frac{\sqrt{97}+10}{3} & \frac{-7\sqrt{97}-64}{8} \end{bmatrix}$$

Subtract row 2 multiplied by $\frac{\sqrt{97}+10}{3}$ from row 3: $R_3 = R_3 - \frac{\sqrt{97}+10}{3}R_2$

$$\begin{bmatrix} 1 & 0 & \frac{-3-\sqrt{97}}{8} \\ 0 & 1 & \frac{-\sqrt{97}-11}{8} \\ 0 & 0 & 0 \end{bmatrix}$$

Solving:

$$\begin{bmatrix} 1 & 0 & \frac{-3-\sqrt{97}}{8} \\ 0 & 1 & \frac{-\sqrt{97}-11}{8} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$z = t$

$y = \frac{11+\sqrt{97}}{8}t$

$x = \frac{3+\sqrt{97}}{8}t$

So, $v_1 = \begin{bmatrix} 3+\sqrt{97} \\ 11+\sqrt{97} \\ 8 \end{bmatrix}$ is an eigenvector

$\lambda_2 = 10 - \sqrt{47}$

$$A - \lambda_2 I = \begin{bmatrix} 5-\lambda & 8 & 3 \\ 8 & 13-\lambda & 5 \\ 3 & 5 & 2-\lambda \end{bmatrix} = \begin{bmatrix} -5+\sqrt{47} & 8 & 3 \\ 8 & 3+\sqrt{47} & 5 \\ 3 & 5 & -8+\sqrt{47} \end{bmatrix}$$

Compute row-reduced form:

Divide row 1 by $-5+\sqrt{47}$: $R_1 = \frac{R_1}{-5+\sqrt{47}}$

$$\begin{bmatrix} 1 & \frac{5+\sqrt{47}}{9} & \frac{5+\sqrt{47}}{24} \\ 8 & 3+\sqrt{47} & 5 \\ 3 & 5 & -8+\sqrt{47} \end{bmatrix}$$

Subtract row 1 multiplied by 8 from row 2 : $R_2 = R_2 - 8R_1$

$$\begin{bmatrix} 1 & \frac{5+\sqrt{47}}{9} & \frac{5+\sqrt{47}}{24} \\ 0 & \frac{-13-\sqrt{47}}{9} & \frac{10-\sqrt{47}}{3} \\ 3 & 5 & -8+\sqrt{47} \end{bmatrix}$$

Subtract row 1 multiplied by 3 from row 3: $R_3 = R_3 - 3R_1$

$$\begin{bmatrix} 1 & \frac{5+\sqrt{47}}{9} & \frac{5+\sqrt{47}}{24} \\ 0 & \frac{-13+\sqrt{47}}{9} & \frac{10-\sqrt{47}}{3} \\ 0 & \frac{10-\sqrt{47}}{3} & \frac{-69+7\sqrt{47}}{8} \end{bmatrix}$$

Multiply row 2 by $\frac{9}{-13+\sqrt{47}}$: $R_2 = \frac{9}{-13+\sqrt{47}} R_2$

$$\begin{bmatrix} 1 & \frac{5+\sqrt{47}}{9} & \frac{5+\sqrt{47}}{24} \\ 0 & 1 & \frac{-11+\sqrt{47}}{8} \\ 0 & \frac{10-\sqrt{47}}{3} & \frac{-69+7\sqrt{47}}{8} \end{bmatrix}$$

Subtract row 2 multiplied by $\frac{5+\sqrt{47}}{9}$ from row 1 : $R_1 = R_1 - \frac{5+\sqrt{47}}{9} R_2$

$$\begin{bmatrix} 1 & 0 & \frac{-3+\sqrt{47}}{8} \\ 0 & 1 & \frac{-11+\sqrt{47}}{8} \\ 0 & \frac{10-\sqrt{47}}{3} & \frac{-69+7\sqrt{47}}{8} \end{bmatrix}$$

Add row 2 multiplied by $\frac{-10+\sqrt{47}}{3}$ to row 3: $R_3 = R_3 + \frac{-10+\sqrt{47}}{3} R_2$

$$\begin{bmatrix} 1 & 0 & \frac{-3+\sqrt{47}}{8} \\ 0 & 1 & \frac{-11+\sqrt{47}}{8} \\ 0 & 0 & 0 \end{bmatrix}$$

We solve:

$$\begin{bmatrix} 1 & 0 & \frac{-3+\sqrt{47}}{8} \\ 0 & 1 & \frac{-11+\sqrt{47}}{8} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$z = t$

$y = \frac{11-\sqrt{47}}{8}$

$x = \frac{3-\sqrt{47}}{8}$

$$v_2 = \begin{bmatrix} 3-\sqrt{47} \\ 11-\sqrt{47} \\ 8 \end{bmatrix}$$ is an eigenvector

$\lambda_3 = 0$:

$$\begin{bmatrix} 5 & 8 & 3 \\ 8 & 13 & 5 \\ 3 & 5 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$x \begin{bmatrix} 5 \\ 8 \\ 3 \end{bmatrix} + y \begin{bmatrix} 8 \\ 13 \\ 5 \end{bmatrix} + z \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By observation, $x = 1, y = -1, z = 1$ works

$v_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ is an eigenvector.

Now, unnormalized V is

$$V^{\text{not normalized}} = [v_1 \; v_2 \; v_3]$$

$$V^{\text{not normalized}} = \begin{bmatrix} 3+\sqrt{97} & 3-\sqrt{97} & 1 \\ 11+\sqrt{97} & 11-\sqrt{97} & -1 \\ 8 & 8 & 1 \end{bmatrix}$$

$$\Rightarrow \boxed{V = \begin{bmatrix} \frac{3+\sqrt{97}}{2\sqrt{97+7\sqrt{97}}} & \frac{3-\sqrt{97}}{2\sqrt{97-7\sqrt{97}}} & \frac{\sqrt{3}}{3} \\ \frac{11+\sqrt{97}}{2\sqrt{97+7\sqrt{97}}} & \frac{11-\sqrt{97}}{2\sqrt{97-7\sqrt{97}}} & -\frac{\sqrt{3}}{3} \\ \frac{4}{\sqrt{97+7\sqrt{97}}} & \frac{4}{\sqrt{97-7\sqrt{97}}} & \frac{\sqrt{3}}{3} \end{bmatrix}}$$

$AV = U\Sigma$

Since Σ is of form $\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$, $u_i = \frac{1}{\sigma_i} A v_i$

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{10+\sqrt{97}}} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} \frac{3+\sqrt{97}}{2\sqrt{97+7\sqrt{97}}} \\ \frac{11+\sqrt{97}}{2\sqrt{97+7\sqrt{97}}} \\ \frac{4}{\sqrt{97+7\sqrt{97}}} \end{bmatrix} = \begin{bmatrix} \frac{3(\sqrt{97}+11)}{2\sqrt{167\sqrt{97}+1694}} \\ \frac{47+5\sqrt{97}}{2\sqrt{167\sqrt{97}+1694}} \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{10-\sqrt{97}}} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} \frac{3-\sqrt{97}}{2\sqrt{97-7\sqrt{97}}} \\ \frac{11-\sqrt{97}}{2\sqrt{97-7\sqrt{97}}} \\ \frac{4}{\sqrt{97-7\sqrt{97}}} \end{bmatrix} = \begin{bmatrix} \frac{33-3\sqrt{97}}{2\sqrt{1694-167\sqrt{97}}} \\ \frac{47-5\sqrt{97}}{2\sqrt{1649-167\sqrt{97}}} \end{bmatrix}$$

Therefore, $\boxed{U = \begin{bmatrix} \frac{3(\sqrt{97}+11)}{2\sqrt{167\sqrt{97}+1649}} & \frac{33-3\sqrt{97}}{2\sqrt{1694-167\sqrt{97}}} \\ \frac{47+5\sqrt{97}}{2\sqrt{167\sqrt{97}+1649}} & \frac{47-5\sqrt{97}}{2\sqrt{1649-167\sqrt{97}}} \end{bmatrix}}$

We have now found both U and V, so we are done.

1) Implement kmeans function- this performs one trial of the k-means algorithm, starting with a random initialization of k data points

In [3]:

```
import numpy as np
def kmeans(data, k):
    num_examples, num_features = data.shape
    old_centroids = np.zeros((k, num_features))
    random_indices = np.random.choice(num_examples, size=k, replace=False)
    cluster_centroids = data[random_indices, :]
    cluster_assignments = [0]*num_examples

    repeat = True
    while repeat:
        # Cluster Assignment Step
        clusteridx_to_examples = dict()
        for i in range(num_examples):
            # Find cluster point closest to data point
            mindist, minidx = np.linalg.norm(data[i]-cluster_centroids[0]), 0
            for j in range(1, k):
                if np.linalg.norm(data[i]-cluster_centroids[j]) < mindist:
                    mindist, minidx = np.linalg.norm(data[i]-cluster_centroids[j]), j

            # Store data point in index_to_examples dictionary
            cluster_assignments[i] = minidx
            if minidx in clusteridx_to_examples:
                clusteridx_to_examples[minidx].append(i)
            else:
                clusteridx_to_examples[minidx] = [i]

        # Move centroids step
        for clusteridx, examples_in_cluster in clusteridx_to_examples.items():
            cluster_centroids[clusteridx] = np.mean(np.array([data[i] for i in examples_in_cluster]), axis=0)

        # Repeat iff cluster centroids have changed
        comparison = old_centroids == cluster_centroids
        repeat = not comparison.all()
        old_centroids = np.copy(cluster_centroids)

    return cluster_centroids, cluster_assignments
```

2) Run kmeans for breast cancer data for 2 <= k <= 8

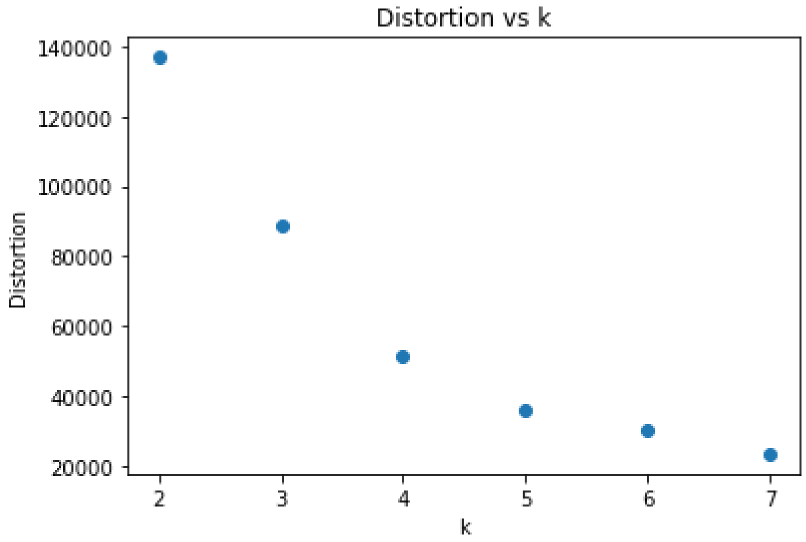
In [6]:
```
from sklearn.datasets import load_breast_cancer
dataset = load_breast_cancer()
data = dataset.data
num_examples = data.shape[0]
distortions = []
for k in range(2,8):
    cluster_centroids, cluster_assignments = kmeans(data, k)
    distortion = 0
    for i in range(num_examples):
        distortion += np.linalg.norm(data[i] - cluster_centroids[cluster_assignments[i]])**2
    distortions.append(distortion/num_examples)
distortions
```

Out[6]:
```
[136982.6008405956,
 88783.42629047789,
 51364.748070263326,
 36331.63689693994,
 30099.055519437814,
 23468.086724960584]
```

3) Plot distortion vs k

In [10]:
```
import matplotlib.pyplot as plt
plt.scatter(list(range(2,8)), distortions)
plt.title("Distortion vs k")
plt.xlabel("k")
plt.ylabel("Distortion")
```

Out[10]:
```
Text(0, 0.5, 'Distortion')
```


Distortion vs k
Distortion
140000
120000
100000
80000
60000
40000
20000
2
3
4
5
6
7
k