2.a.

We are using the square of euclidean distance here. It will give the same result as euclidean distance.

Running for $\mu_1 = [20, 30]$ and

$\mu_2 = [20, -10]$.

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 1000$ and to $\mu_2 = 200$.

$$z_1 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_1 = 1000$ and to $\mu_2 = 200$.

$$z_2 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 200$ and to $\mu_2 = 1000$.

$$z_3 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 100$ and to $\mu_2 = 900$

$$z_4 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_4) - \mu_k\|^2 = 1$$

$$\mu_1 = \text{average}(\phi(x_3), \phi(x_4)) = (15, 20)$$

$$\mu_2 = \text{average}(\phi(x_1), \phi(x_2)) = (20, 0)$$

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 425$ and to $\mu_2 = 100$

$$z_1 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_2 = 100$ and to $\mu_1 = 625$

$$z_2 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 25$ and to $\mu_2 = 500$

$$z_3 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 25$ and to $\mu_2 = 400$

$$z_4 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_4) - \mu_k\|^2 = 1$$

Therefore we have convergence.

$\phi(x_1), \phi(x_2) \in$ cluster 2. $\phi(x_3), \phi(x_4) \in$ cluster 1.

Running for $\mu_1 = [0, 10]$ and $\mu_2 = [30, 20]$

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 200$ and to $\mu_2 = 800$

$$z_1 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_1 = 1000$ and to $\mu_2 = 400$

$$z_2 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 200$ and to $\mu_2 = 400$

$$Z_3 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 500$ and to $\mu_2 = 100$

$$Z_4 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_4) - \mu_k\|^2 = 2$$

$\mu_1 = \text{average}(\phi(x_1), \phi(x_3)) = (10, 10)$
$\mu_2 = \text{average}(\phi(x_2), \phi(x_4)) = (25, 10)$

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 100$ and to $\mu_2 = 325$

$$Z_1 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_1 = 500$ and to $\mu_2 = 125$

$$Z_2 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 100$ and to $\mu_2 = 325$

$$z_3 \longleftarrow \underset{k=1,\dots,K}{\arg\min} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 200$ and to $\mu_2 = 125$

$$z_4 \longleftarrow \underset{k=1,\dots K}{\arg\min} \|\phi(x_4) - \mu_k\|^2 = 2$$

So we have convergence

$\phi(x_1), \phi(x_3) \in$ cluster 1.

$\phi(x_2), \phi(x_4) \in$ cluster 2.

1.e.

Run your linear predictor with feature extractor extractCharacterFeatures. Experiment with different values of n to see which one produces the smallest test error. You should observe that this error is nearly as small as that produced by word features. Why is this the case?

We get the lowest error when n = 5. It is similar to the error we can get with whole words as features. The reason of this similarity is in the average word length in English language, which is equal to 4.7 characters.

A review in which n-grams probably outperform word features is

"One of the best movie with the greatest videography". The word like videography is rare and may have low weight in the word features, but 5 grams like video would be common.

2.a.

We are using the square of euclidean distance here. It will give the same result as euclidean distance.

Running for $\mu_1 = [20, 30]$ and
$\mu_2 = [20, -10]$.

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 1000$ and to $\mu_2 = 200$.

$$z_1 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_1 = 1000$ and to $\mu_2 = 200$.

$$z_2 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 200$ and to $\mu_2 = 1000$.

$$z_3 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 100$ and to $\mu_2 = 900$

$$z_4 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_4) - \mu_k\|^2 = 1$$

$$\mu_1 = \text{average}(\phi(x_3), \phi(x_4)) = (15, 20)$$

$$\mu_2 = \text{average}(\phi(x_1), \phi(x_2)) = (20, 0)$$

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 425$ and to $\mu_2 = 100$

$$z_1 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_2 = 100$ and to $\mu_1 = 625$

$$z_2 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 25$ and to $\mu_2 = 500$

$$z_3 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 25$ and to $\mu_2 = 400$

$$z_4 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_4) - \mu_k\|^2 = 1$$

Therefore we have convergence.

$\phi(x_1), \phi(x_2) \in$ cluster 2. $\phi(x_3), \phi(x_4) \in$
$\in$ cluster 1.

Running for $\mu_1 = [0, 10]$ and $\mu_2 = [30, 20]$

Euclidean distance squance for $\phi(x_1)$ to $\mu_1 = 200$ and to $\mu_2 = 800$

$$z_1 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_1 = 1000$ and to $\mu_2 = 400$

$$z_2 \leftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 200$ and to $\mu_2 = 400$

$$z_3 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 500$ and to $\mu_2 = 100$

$$z_4 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_4) - \mu_k\|^2 = 2$$

$$\mu_1 = \text{average}(\phi(x_1), \phi(x_3)) = (10, 10)$$
$$\mu_2 = \text{average}(\phi(x_2), \phi(x_4)) = (25, 10)$$

Euclidean distance squared for $\phi(x_1)$ to $\mu_1 = 100$ and to $\mu_2 = 325$

$$z_1 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_1) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_2)$ to $\mu_1 = 500$ and to $\mu_2 = 125$

$$z_2 \longleftarrow \arg\min_{k=1,\dots,K} \|\phi(x_2) - \mu_k\|^2 = 2$$

Euclidean distance squared for $\phi(x_3)$ to $\mu_1 = 100$ and to $\mu_2 = 325$

$$z_3 \leftarrow \underset{k=1,\dots,K}{\arg\min} \|\phi(x_3) - \mu_k\|^2 = 1$$

Euclidean distance squared for $\phi(x_4)$ to $\mu_1 = 200$ and to $\mu_2 = 125$

$$z_4 \leftarrow \underset{k=1,\dots K}{\arg\min} \|\phi(x_4) - \mu_k\|^2 = 2$$

So we have convergence

$\phi(x_1), \phi(x_3) \in$ cluster 1.

$\phi(x_2), \phi(x_4) \in$ cluster 2.

1.e.

Run your linear predictor with feature extractor extractCharacterFeatures. Experiment with different values of n to see which one produces the smallest test error. You should observe that this error is nearly as small as that produced by word features. Why is this the case?

We get the lowest error when n = 5. It is similar to the error we can get with whole words as features. The reason of this similarity is in the average word length in English language, which is equal to 4.7 characters.

A review in which n-grams probably outperform word features is „One of the best movie with the greatest videography. The word like videography is rare and may have low weight in the word features, but 5 grams like video would be common.