Name:

Giovanni Zavalza

Netid:

gzava3

CS 441 - HW1: Instance-based Methods

Complete the sections below. You do not need to fill out the checklist.

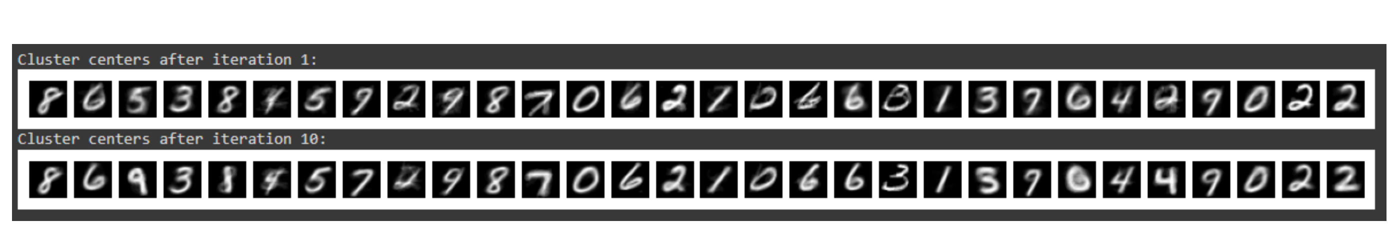
Total Points Available	**[ ] / 145**
1. Retrieval, K-means, 1-NN on MNIST	
a. Retrieval	[ ] / 5
b. K-means	[ ] / 15
c. 1-NN	[ ] / 10
2. Make it fast	
a. K-means plot	[ ] / 15
b. 1-NN error plots	[ ] / 8
c. 1-NN time plots	[ ] / 7
d. Most confused label	[ ] / 5
3. Temperature Regression	
a. RMSE Tables	[ ] / 20
4. Conceptual questions	[ ] / 15
5. Stretch Goals	
a. Evaluate effect of K for MNIST	[ ] / 15
b. Evaluate effect of K for Temp Reg.	[ ] / 15
c. Compare Kmeans more iterations vs. restarts	[ ] / 15

1. Retrieval, K-means, 1-NN on MNIST

a. What index is returned for x_test[1]?

31117

b. Paste the display of clusters after the 1st and 10th iteration for K=30.



```
Cluster centers after iteration 1:
```

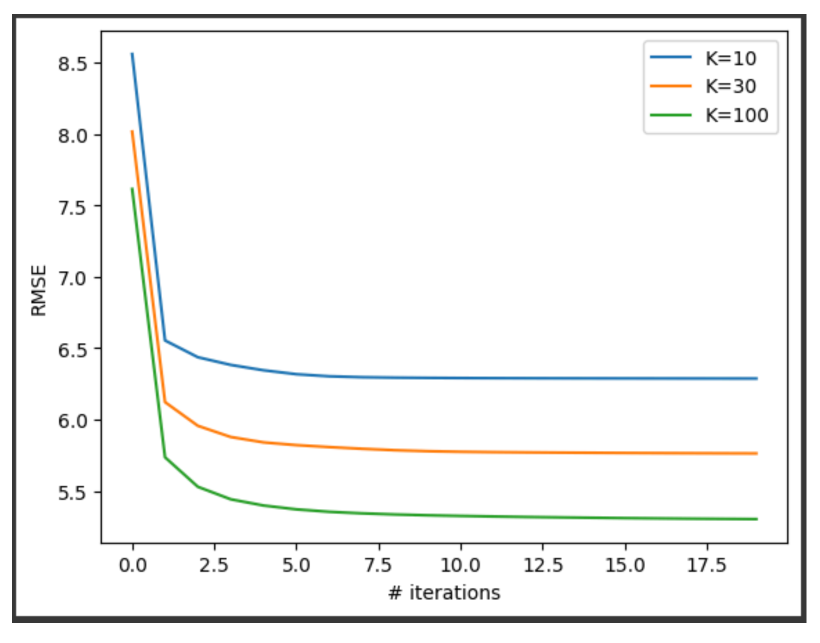
```
Cluster centers after iteration 10:
```

c. Error rate for first 100 test samples, using first 10,000 training samples (x.x%)

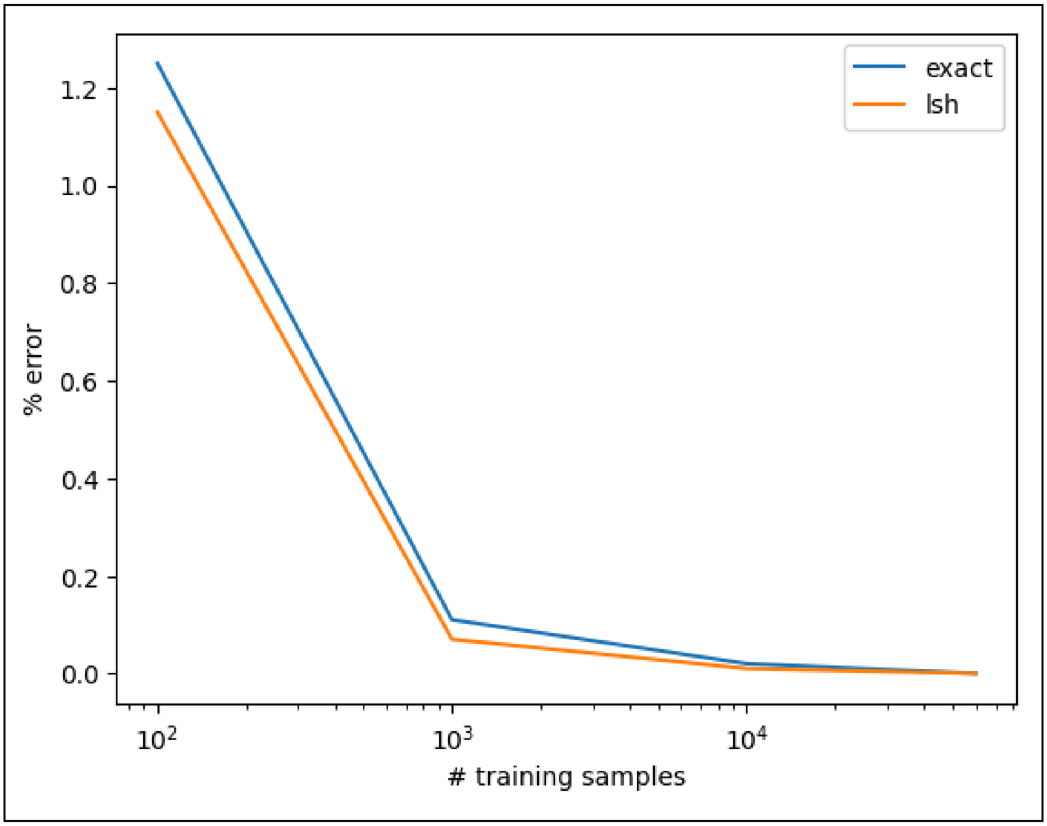
7.0%

2. Make it fast

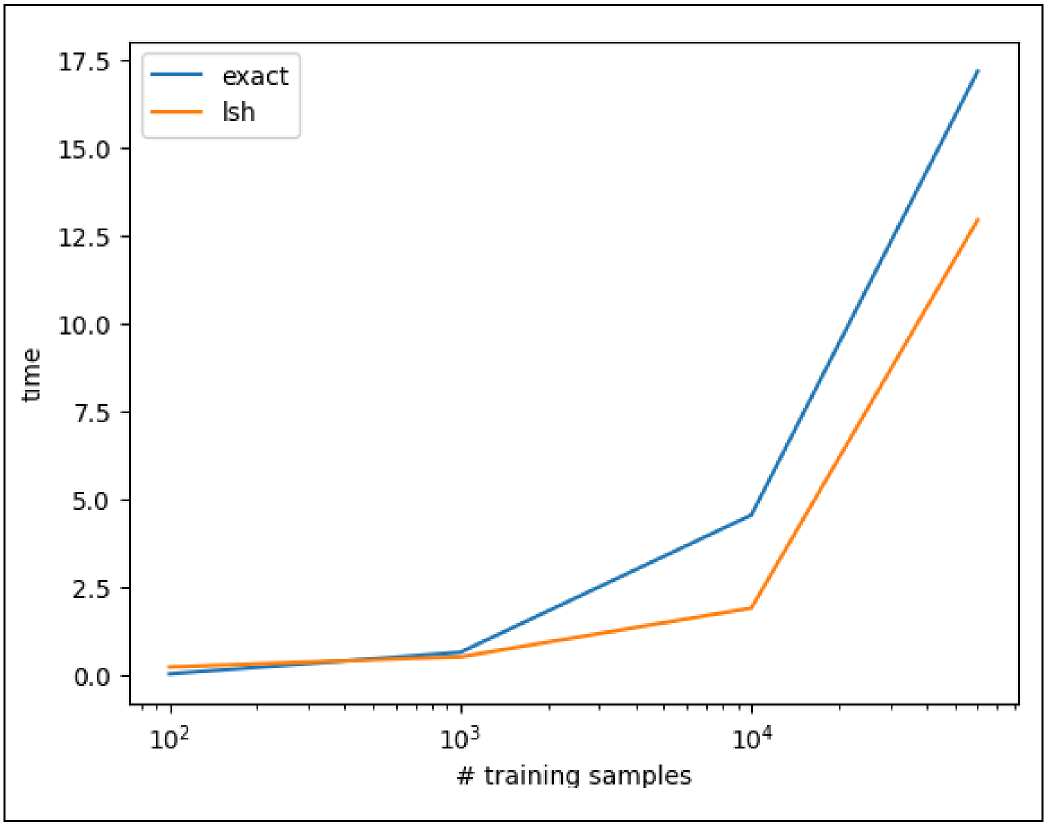
a. KMeans plot of RMSE vs iterations for K=10, 30, 100



b. Nearest neighbor error vs training size plot



c. Nearest neighbor time vs training size plot



d. What label is most commonly confused with ‘3’?

5

3. Temperature Regression

a. Table of RMSE for KNN with K=5 (x.xx)

	KNN (K=5)
Original Features	3.19
Normalized Features	2.9

4. Test your understanding

Fill in the letter corresponding to the answer. If you’re not sure, you can sometimes run small experiments to check.

1. Is K-means guaranteed to decrease RMSE between each sample and its nearest cluster center in each iteration until convergence?

a. No
b. Yes

b

Mathematically proven that it always approaches a local minimum of error

2. If you increase K, is K-means expected or guaranteed to achieve lower RMSE?
    a. Guaranteed
    b. Expected but not guaranteed
    c. Not expected

b

Expected, but is there are 3 clusters of data, you want K = 3 and not K = 4 which risks 2 Ks “sharing” a cluster, increasing error chances depending on centroid initialization

3. In K-NN regression, for training labels y, what is the lowest target value that can possibly be predicted for any query?
    a. Min(y)
    b. Mean(y)
    c. Can’t be determined

a

Target value always calculated from K training labels, each training label can be at min(y), making target value capable of reaching min(y) despite taking mean of K labels

4. Would you expect the “training error” for 1-NN to be higher or lower than 3-NN for classification? Training error is the error if you test on the training data.
    a. Lower
    b. Higher
    c. It’s problem-dependent

a

It overfits to training data the lower the K

5. Would you expect the test error for 1-NN to be higher or lower than for 3-NN for regression?
    a. Lower
    b. Higher
    c. It’s problem-dependent

b

The higher the K, the more generalize it becomes, making it more reliable on passing a test

5. Stretch Goals (optional)

a. Select best K parameter for K-NN MNIST classification in K=1, 3, 5, 11, 25. (x.xx)

Validation Set Performance	K=1	K=3	K=5	K=11	K=25
% error	3.04	2.85	3.02	3.50	4.30

Best K:

3

Test % error (x.xx)

2.83

b. Select best K parameter for K-NN temperature regression in K=1, 3, 5, 11, 25. (x.xx)

Validation Set RMSE	K=1	K=3	K=5	K=11	K=25
Original Features	6.23	5.07	4.86	4.62	4.47
Normalized Features	3.94	3.26	3.08	2.92	2.92

Best Setting (K, feature type):

25, Normalized

Test RMSE (x.xx)

2.77

c. Kmeans, MNIST: compare average and standard deviation RMSE based on number of iterations and number of restarts

(4 digit precision)

K=30	RMSE avg	RMSE std
20 iterations, 1 restart	5.7862	0.0107

4 iterations, 5 restarts	5.8261	0.0065
50 iterations, 1 restart	5.7800	0.0082
10 iterations, 5 restarts	5.7842	0.0086

Acknowledgments / Attribution

List any outside sources for code or ideas or “None”.

ChatGPT for debugging and optimizations.
I mention where I used ChatGPT in code comments

https://www.w3schools.com/python/python_ml_confusion_matrix.asp