Lecture 11 (Nov 23, 2020)

Assignment 11 (due Nov 30)

**Quiz #5 (Nov 27-29)**

DO PART I ONLY

k-means

1. do problems described in pdf

PART II (optional no grade)

1. take the best k\* from part 1 and implement your own kmeans with the following three metrics:
2. Euclidean
3. Manhattan (Minkowski p=1)
4. Minkowski (p=1.5)

What distance mtric gives you the most “pure” clusters?

Project Presentations: Dec 7 (Monday)

Final exam: Dec 12-13 (Sat, Sun)

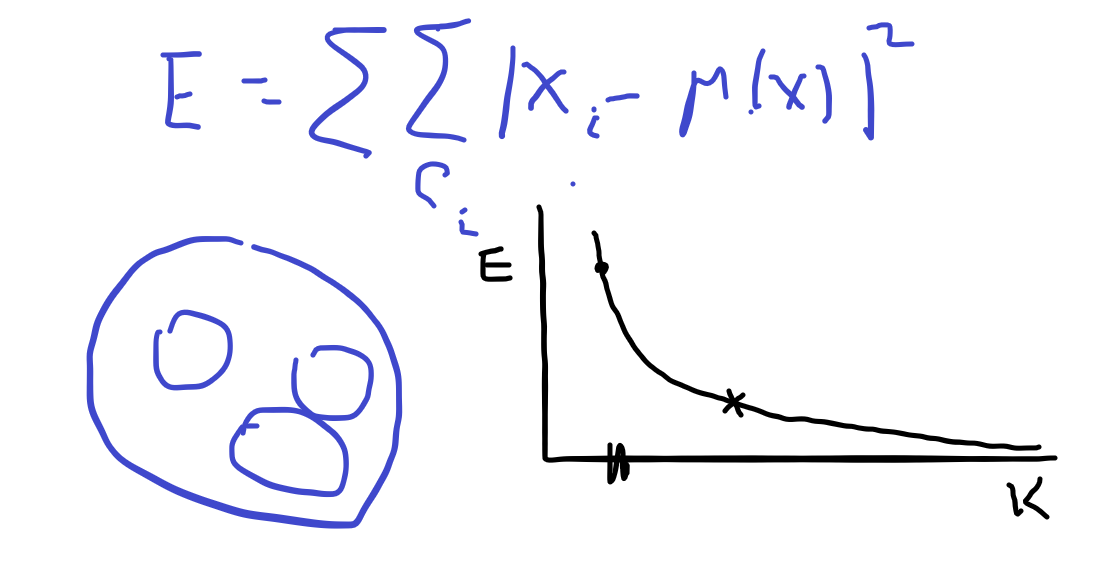
Code due by Dec 10, 2020

powerpoint presentation on Dec 7

* 1-2 page description how to run
* code
* - video recording about your project

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | appl |  | netflix |  | TSLA |  |
| Classifier | Acc | val | Acc | val | Acc | val |
| Buy and Hold |  | $191 |  | 128 |  | 2,400 |
| kNN with best k | 88 (k=3) | 264 | 62 (k=11) | 200 | 99 | 264 |
| logistic | 82 | 282 | 83 | 391 |  |  |
| Naïve bayesian | 81 | 275 | 89 | 332 |  |  |
| Linear models (best degree) | 58 (d=2) | **197** | 60 (d=1) | 131 |  |  |
| Decision tree | 75 | 275 | 100 | 357 |  |  |
| Random Forest | **58** | 275 | 100 | 357 |  |  |
| Linear discriminant | 85 | 278 | 74 | 313 |  |  |
| Quadratic discriminant | 87 | 251 | 91 | 341 |  |  |
| SVM (linear) | 85 | 248 | 83 | 386 |  |  |
| Ada Boost | **89** | **276** | 96 | 348 |  |  |
|  |  |  |  |  |  |  |

CLUSTERING:



We compute (“decide on” ) k by visualization

* in k-means clusters will have different number if elements

Suppose you want clusters to about the same size. How would you do this?

n\_1 + N\_2 = 10

|  |  |  |
| --- | --- | --- |
| N\_1 | N\_2 | N\_1\*\*2 + N\_2\*\*2 |
| 0 | 10 | 100 |
| 1 | 9 | 82 |
| 2 | 8 | 68 |
| 3 | 7 | 58 |
| 4 | 6 | 52 |
| 5 | 5 | **50** |
| 6 | 4 | 52 |
| 7 | 3 | 58 |
| 8 | 2 | 68 |
| 9 | 1 | 82 |
| 10 | 0 | 100 |

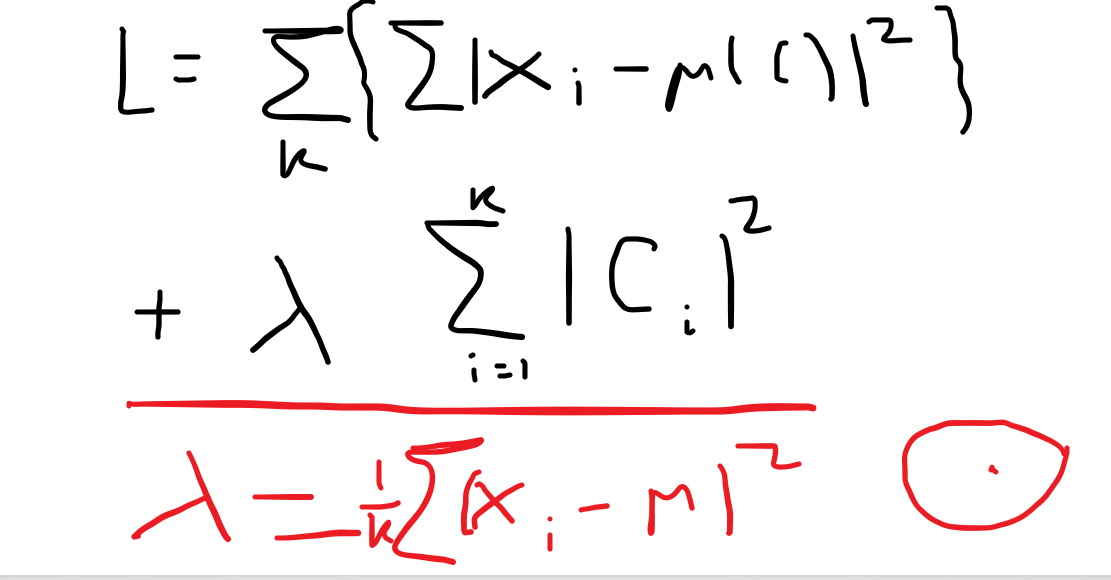
General result:

if N=n1+n2+…+ nk then

n1\*\*2 + n2\*\*2 + … + n2\*\*2 is minimized

when n1=n2=…=nk = N/k

How do we use this to help k-means find clusters of about the same size?



WISH ALL OF YOU A HAPPY THANKSGIVING!

Stay healthy and well

no quiz 6 in CS-677 !!!!