Analysis of Drought in the U.S. Between 2010-2017

Xingyao Chen & Cassidy Le

May 6th, 2019

MATH189, Spring 2019

Outline

- Dataset overview
- Dataset visualization
- County drought classification
- Identification of Dissimilar Drought Counties
- Discussion

Dataset

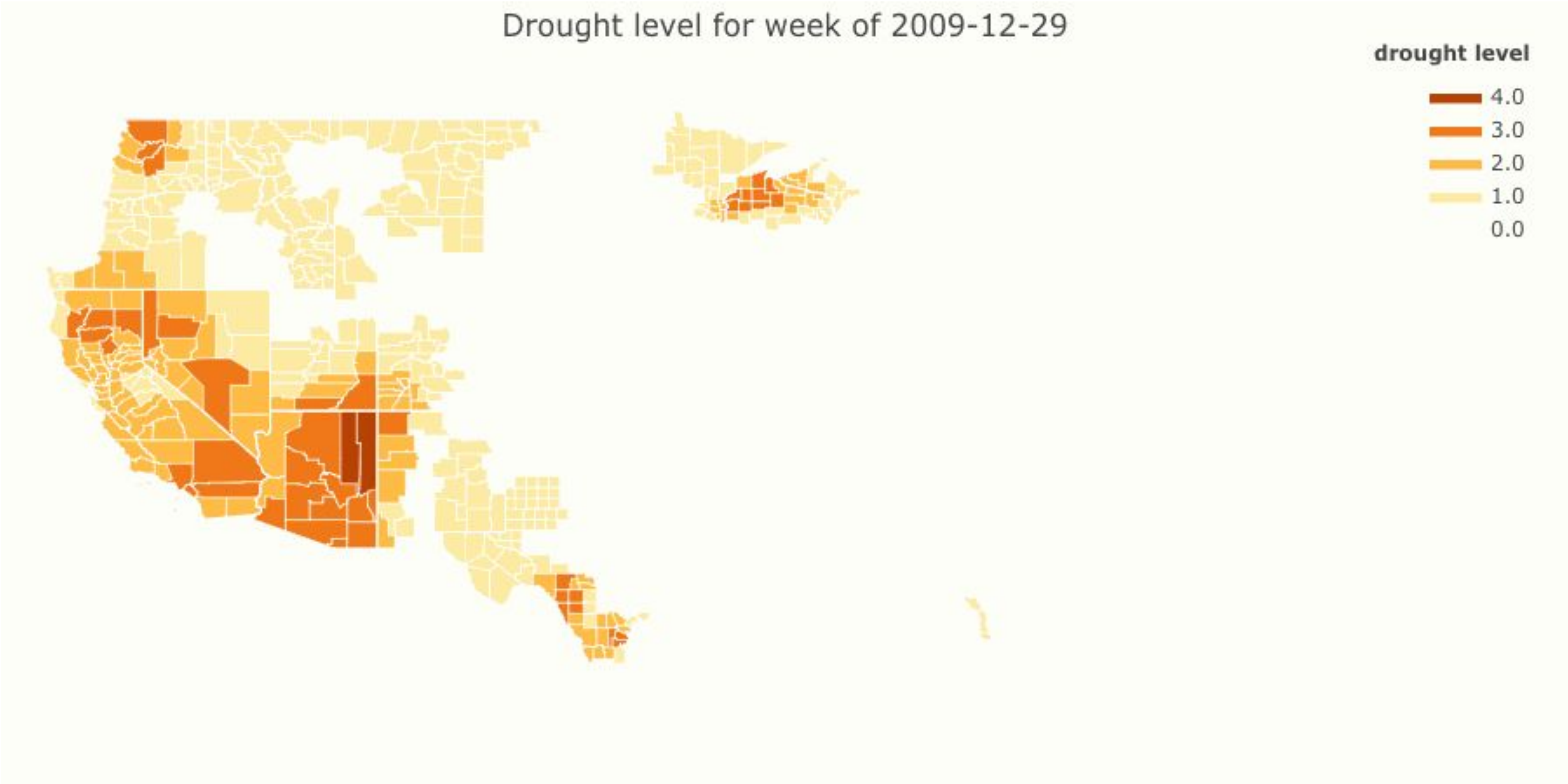
droughts

Data containing the particular percentage of various range of drought severities, indexed by counties for particular start-end periods throughout the United States. *~1.35 million rows & 11 columns.* Size: ~100MB. Source: U.S. Drought Monitor.

Field	Type	Description
fips	INTEGER	FIPS code for the particular county
county	STRING	County name
state	STRING	State code (2-letters)
Classification of Droughts (6 total)	FLOAT	Percentage of population affected by severity of droughts by: no drought (none), D0 = abnormally dry, D1 = moderate, D2 = severe, D3 = extreme, D4 = exceptional Additional Information
valid_start	STRING	Start date of event in form of YYYY-MM-DD
valid_end	STRING	End date of event in form of YYYY-MM-DD

```
"Autauga County, AL"    01001   "Autauga County, AL"    01001
                        "Chilton County, AL"    01021
                        "Dallas County, AL"     01047
                        "Elmore County, AL"     01051
                        "Lowndes County, AL"    01085
                        "Montgomery County, AL" 01101
"Baldwin County, AL"    01003   "Baldwin County, AL"    01003
                        "Clarke County, AL"     01025
                        "Escambia County, AL"   01053
                        "Mobile County, AL"     01097
                        "Monroe County, AL"     01099
                        "Washington County, AL" 01129
                        "Escambia County, FL"   12033
"Barbour County, AL"    01005   "Barbour County, AL"    01005
                        "Bullock County, AL"    01011
                        "Dale County, AL"       01045
                        "Henry County, AL"      01067
                        "Pike County, AL"       01109
                        "Russell County, AL"    01113
                        "Clay County, GA"       13061
                        "Quitman County, GA"    13239
                        "Stewart County, GA"    13259
```

Data Visualization



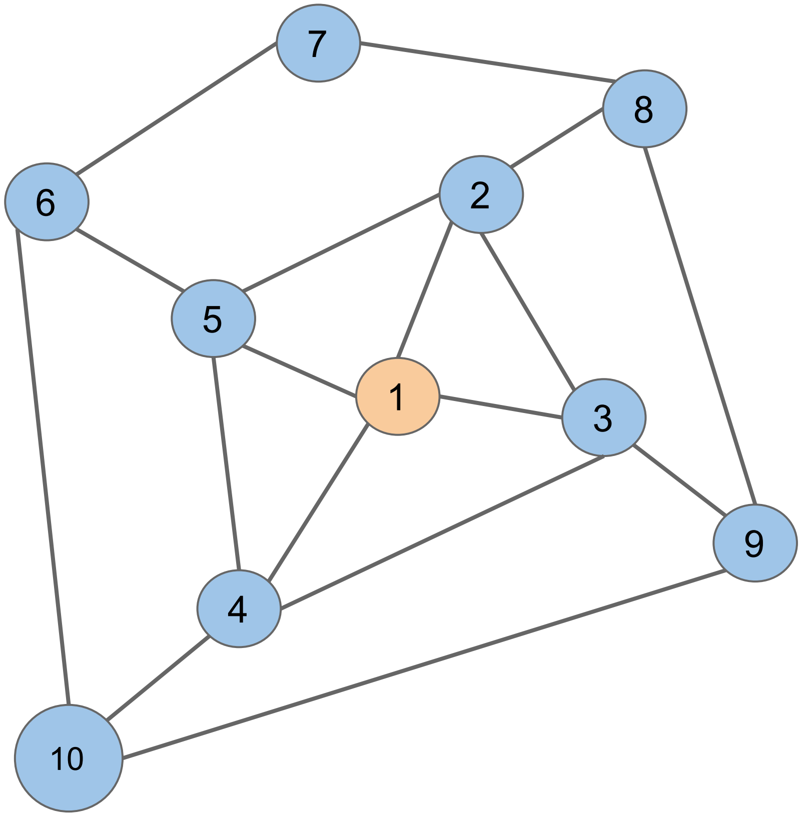
Can we predict a county's drought level?

Implementation

- “K Adjacent Neighbor" classifier (KAN)

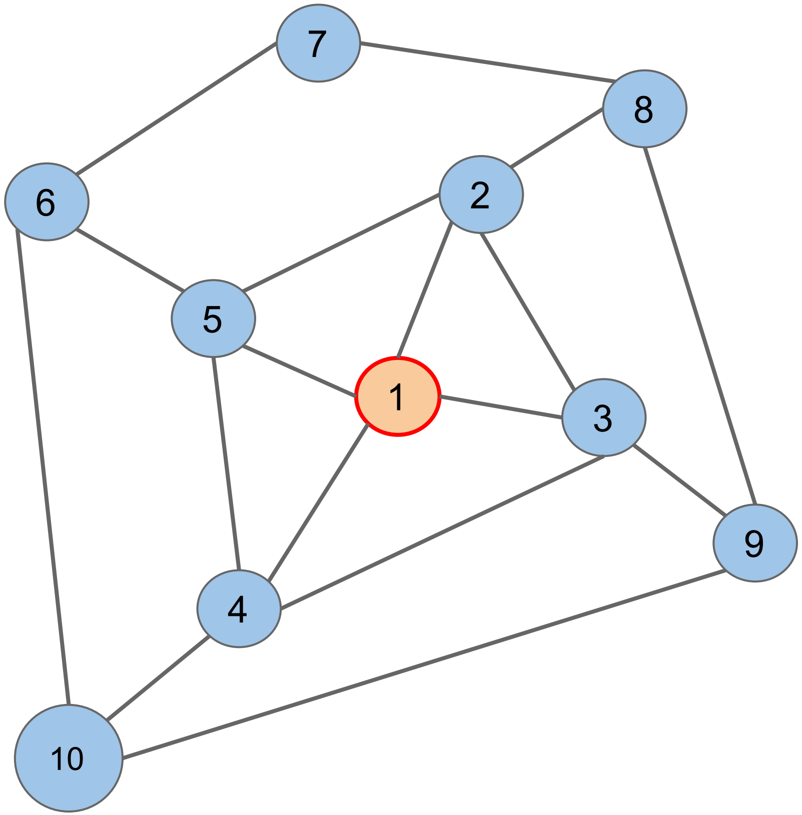
Implementation

- “K Adjacent Neighbor” classifier (KAN)



Implementation

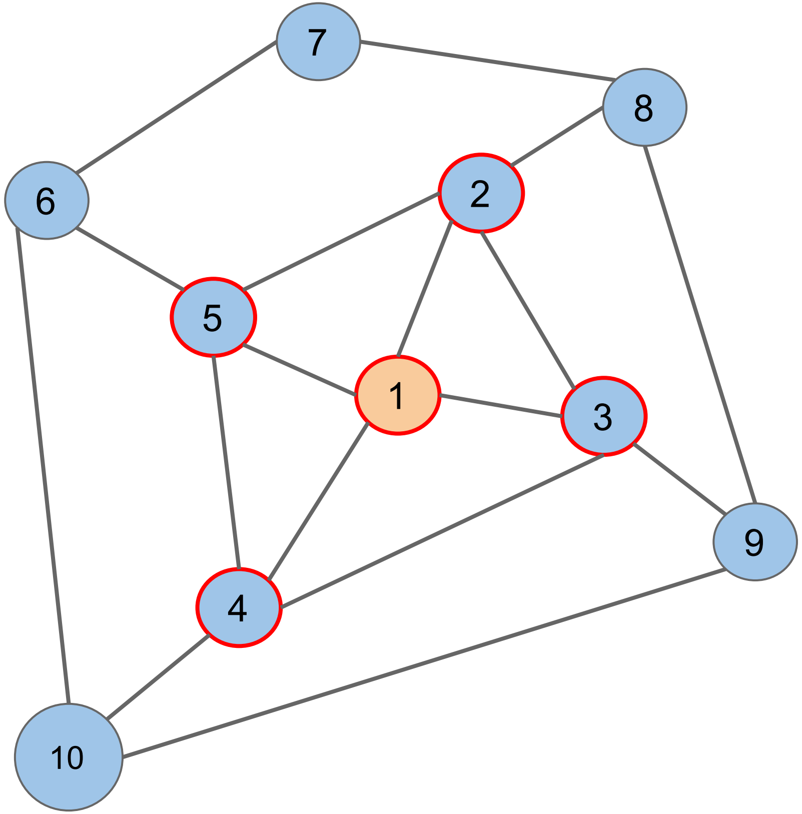
- "K Adjacent Neighbor" classifier (KAN)
- To classify the drought level of a county *c*, KAN takes the majority vote of its neighbors that are *n* degrees of separation from *c*



Degree of separation = 0

Implementation

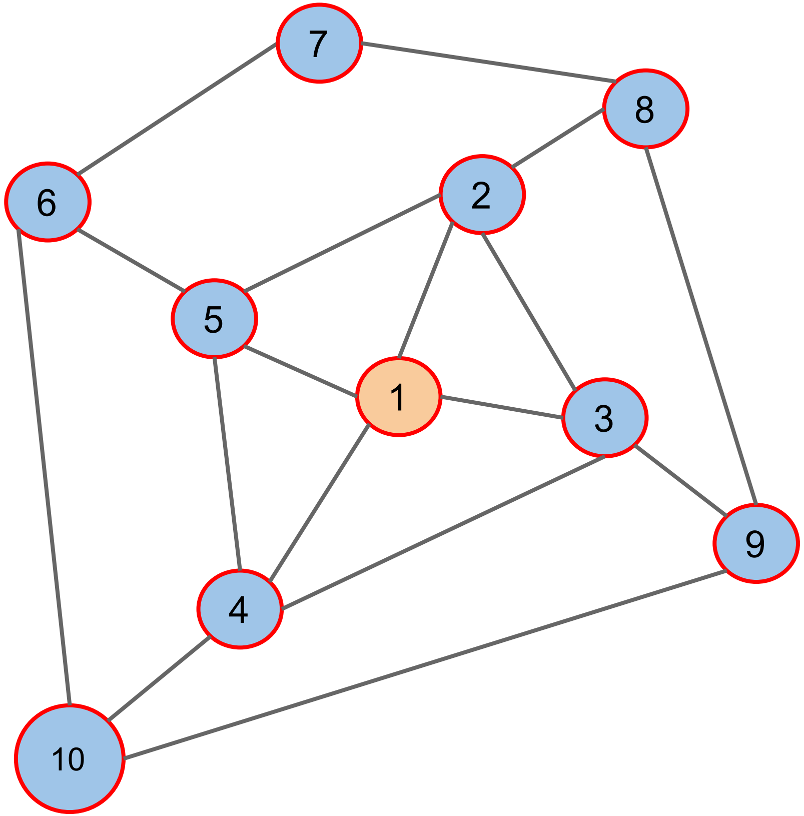
- “K Adjacent Neighbor” classifier (KAN)
- To classify the drought level of a county *c*, KAN takes the majority vote of its neighbors that are *n* degrees of separation from *c*



Degree of separation = 1

Implementation

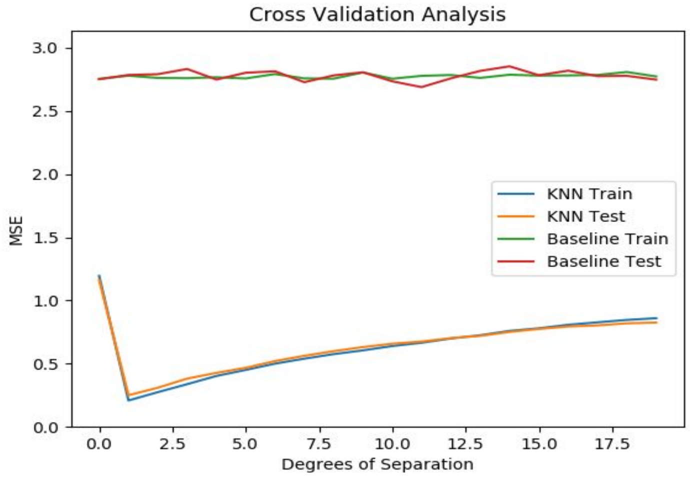
- “K Adjacent Neighbor” classifier (KAN)
- To classify the drought level of a county *c*, KAN takes the majority vote of its neighbors that are *n* degrees of separation from *c*

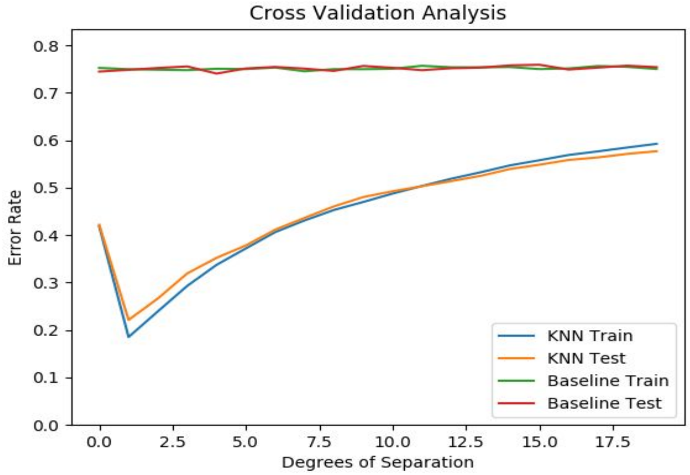


Degree of separation = 2

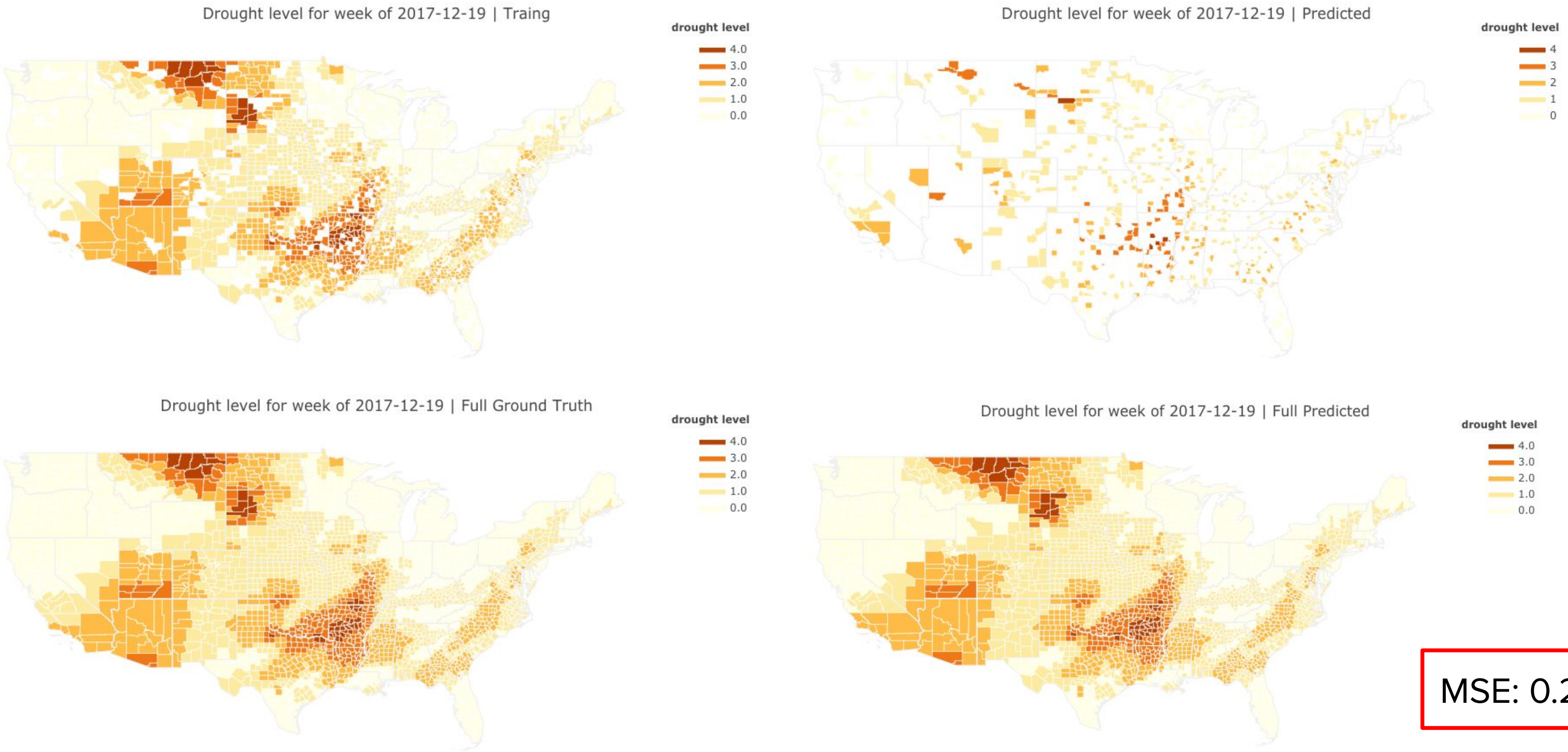
Cross Validation

- 8-fold Cross Validation
  - Data for each fold is drought data of all U.S. counties from a randomly sampled week
  - Training set: randomly sampled 80% of all U.S. counties
  - Validation set: randomly sampled 20% of all U.S. counties
- Optimal Degree of Separation: 1
- Minimum MSE: 0.25





Testing Set Results



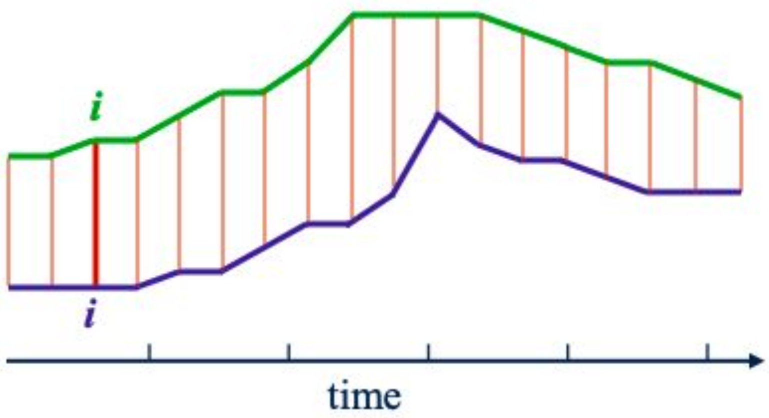
MSE: 0.20

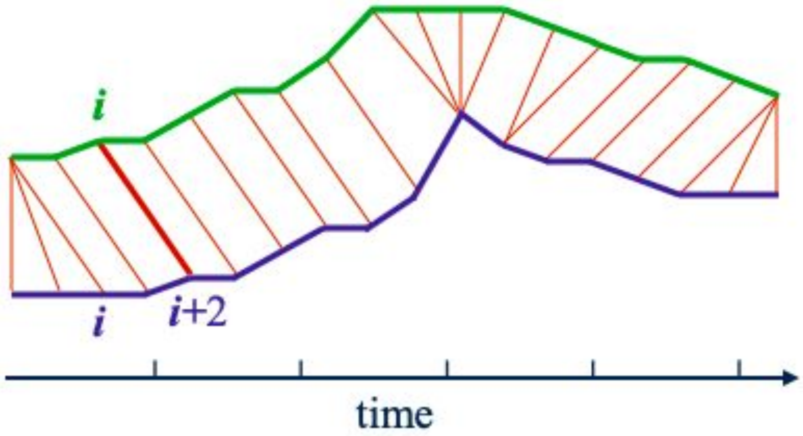
How To Find Dissimilar Drought Counties?

Method

1. For each group of adjacent counties, we measure pairwise distances across all counties using drought data across all seven years (2010-2017)
    - Distance metric used: Dynamic Time Warping
2. Extract groups have at least two pairwise distances that is above some threshold *t*
3. We label these groups of adjacent counties Dissimilar Drought Counties (DDC), which are adjacent counties that have the largest disparities in drought levels across a significant portion of time

Digression: Dynamic Time Warping (DTW)





Linear Alignment Distance Metric vs. Elastic Alignment Distance Metric

http://www.mathcs.emory.edu/~lxiong/cs730_s13/share/slides/searching_sigkdd2012_DTW.pdf

Digression: Dynamic Time Warping (DTW)

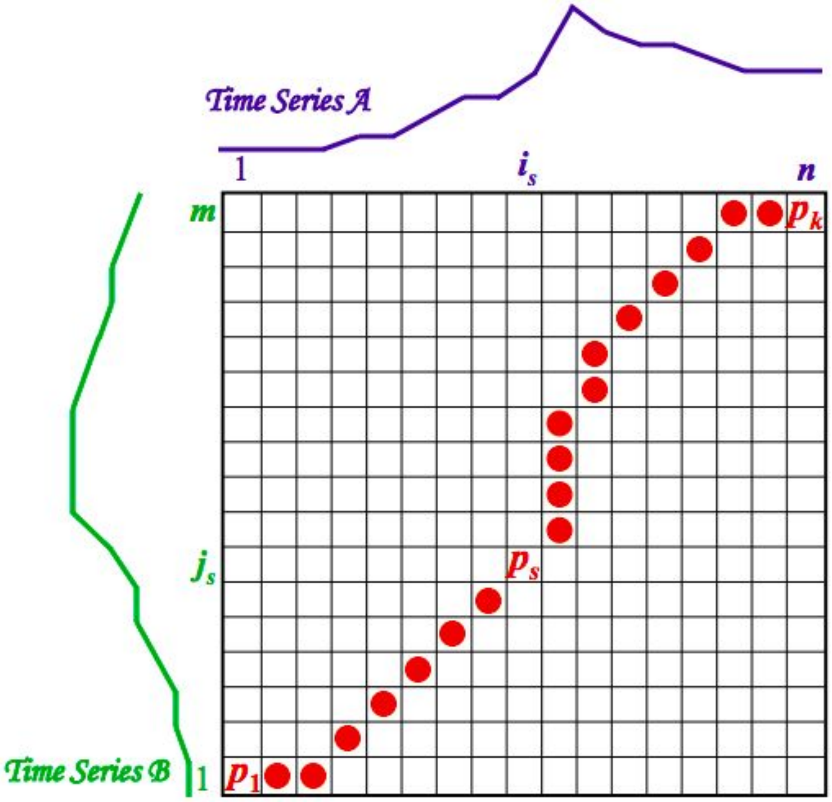
$$DTW(S,T) = min_w\left(\sum_{k=1}^{p}\delta(w_k)\right),$$

where S and T are sequences and W is the warping path defined as $W = w_1, w_2, ..., w_k$. The distance function $\delta(w_k)$ is defined as

$$\delta(S,T) = \left[\frac{\sum_{x=1}^{k} d(p_x)\cdot w_x}{\sum_{x=1}^{k} w_x}\right],$$

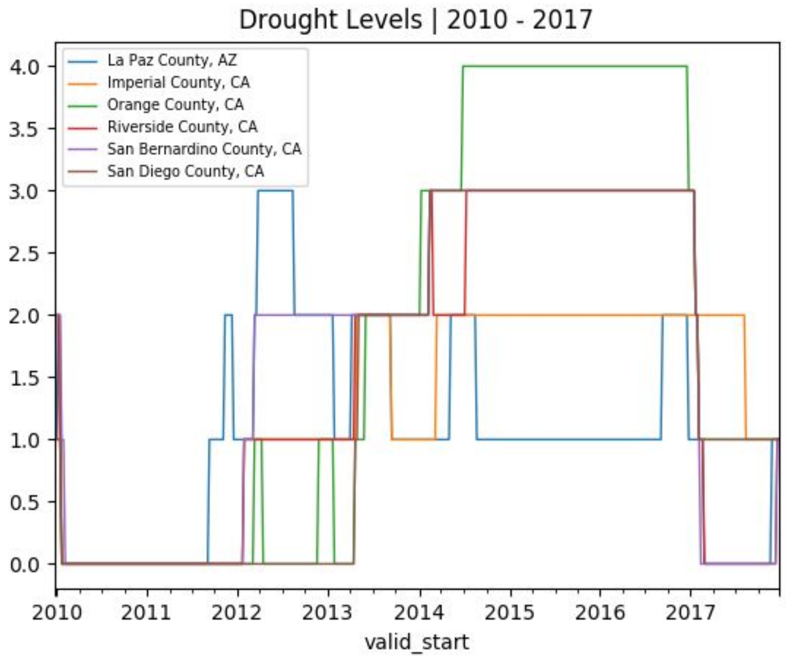
where $d(p_x)$ is the distance between two distinct points and $w_x > 0$ is a weighting coefficient.

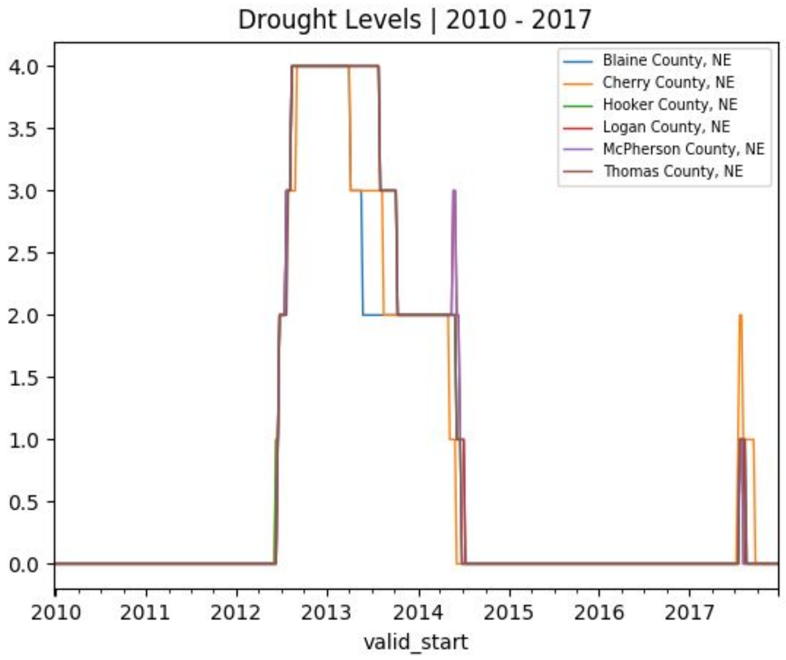
Finding the DTW Path



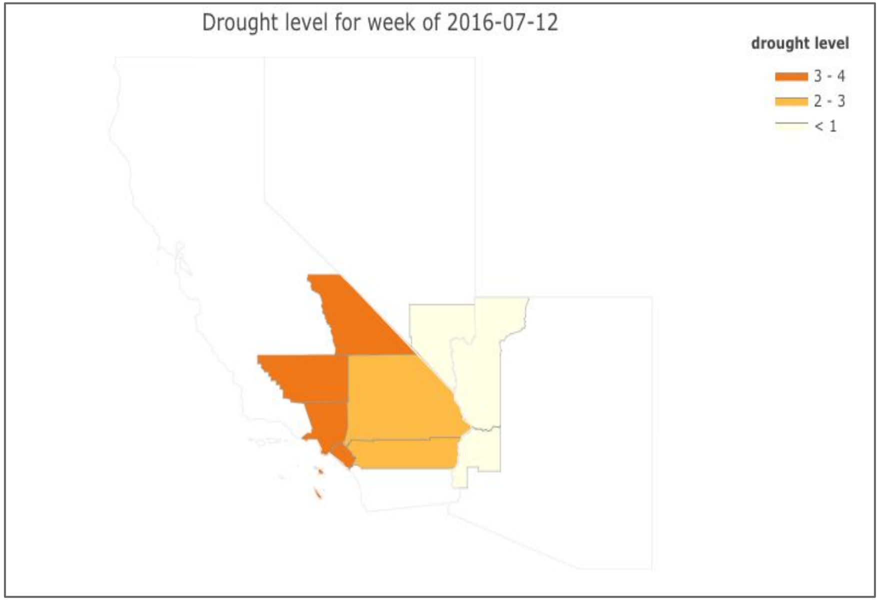
http://www.mathcs.emory.edu/~lxiong/cs730_s13/share/slides/searching_sigkdd2012_DTW.pdf

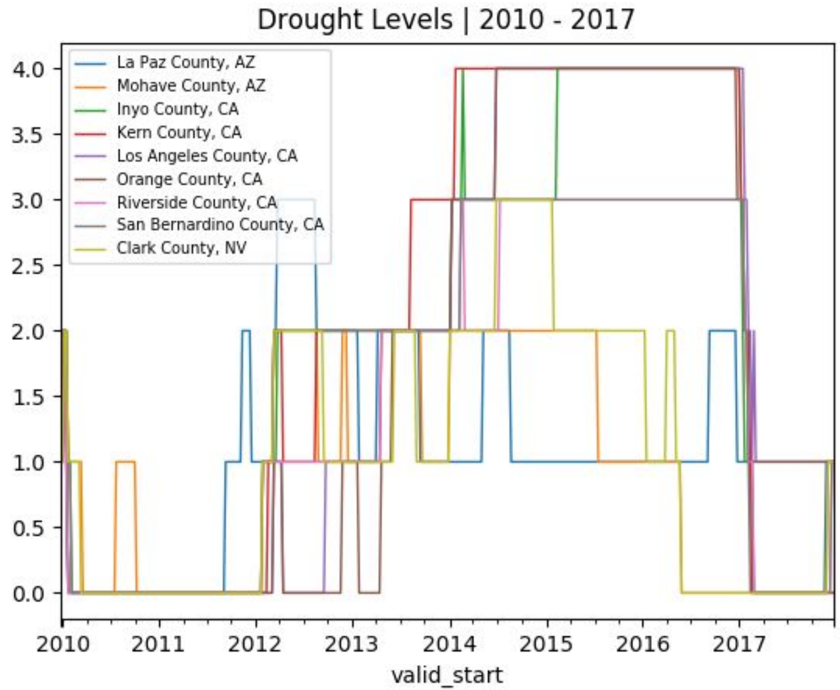
Assessment of Distance Metric



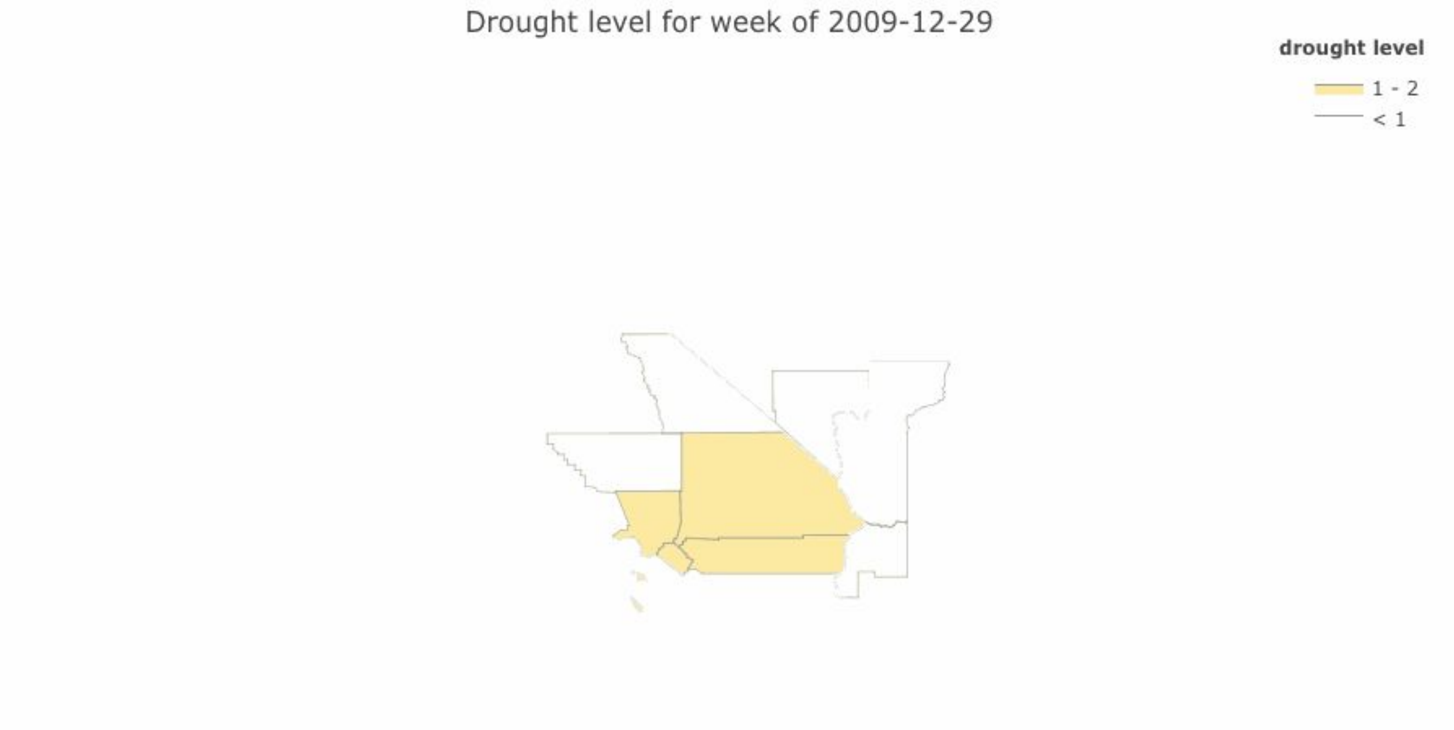


Results

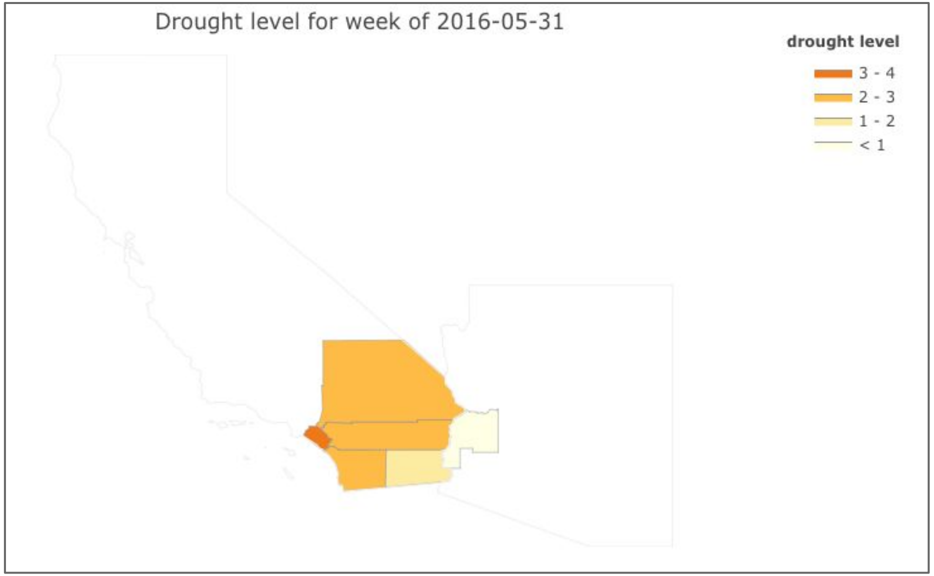

Drought level for week of 2016-07-12
drought level
3 - 4
2 - 3
< 1


Drought Levels | 2010 - 2017
La Paz County, AZ
Mohave County, AZ
Inyo County, CA
Kern County, CA
Los Angeles County, CA
Orange County, CA
Riverside County, CA
San Bernardino County, CA
Clark County, NV
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2010
2011
2012
2013
2014
2015
2016
2017
valid_start

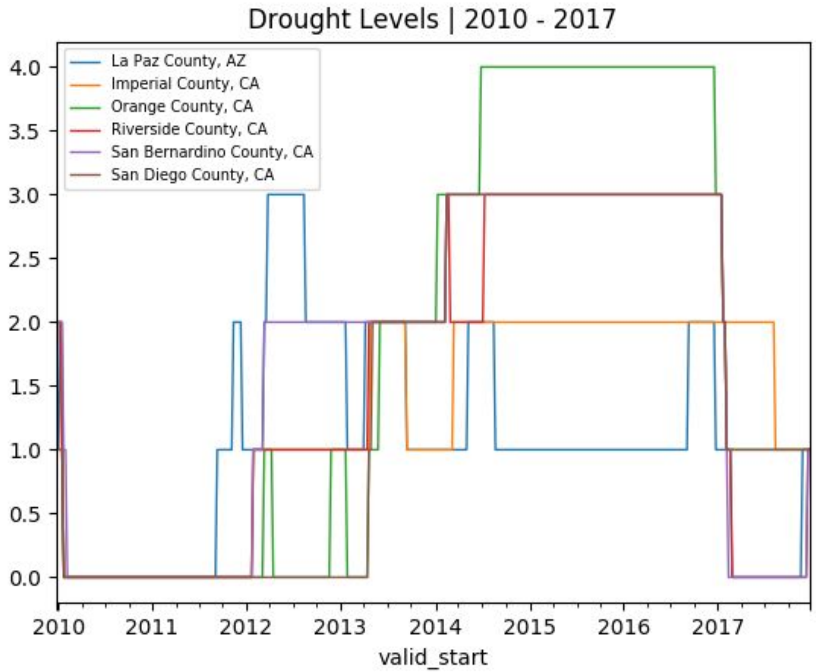
Results



Results

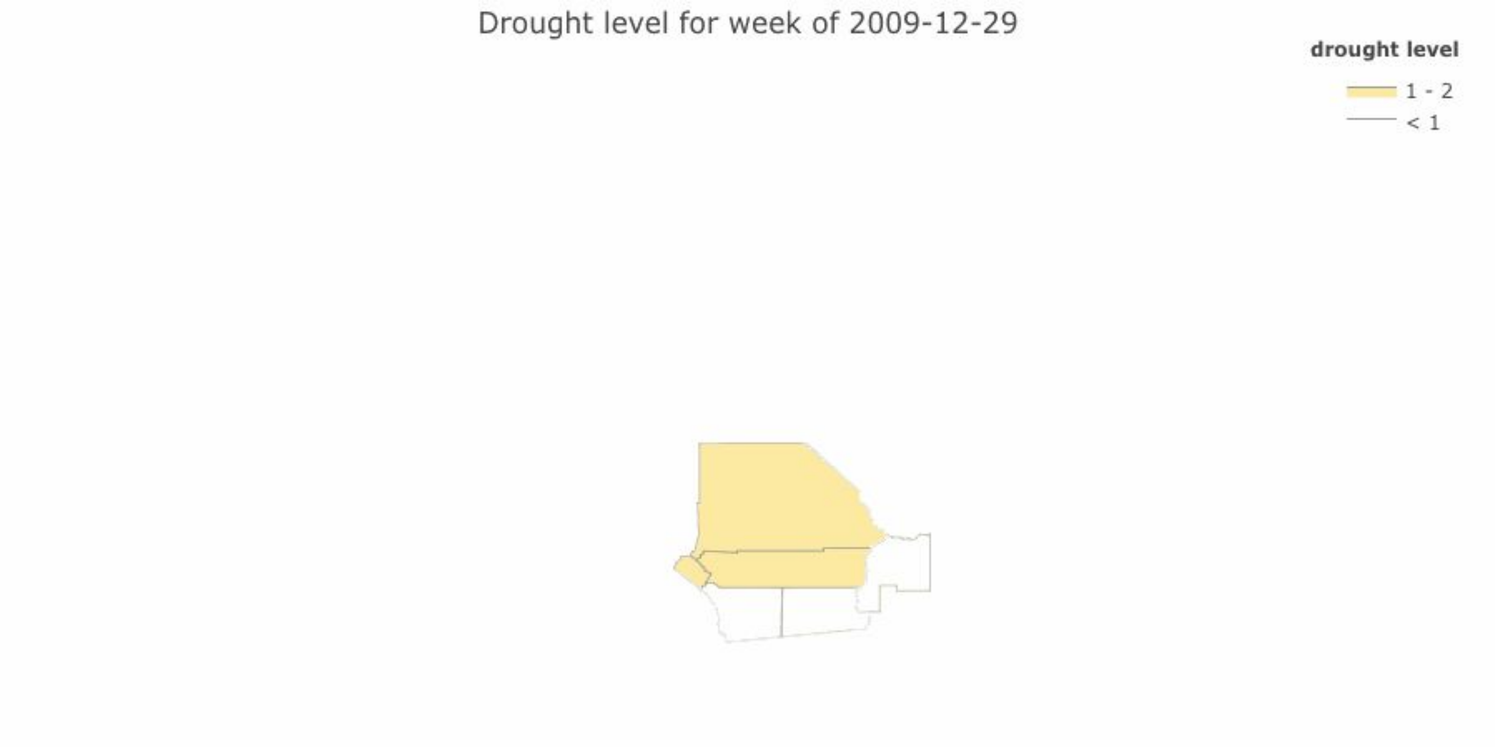


Drought level for week of 2016-05-31
drought level
3 - 4
2 - 3
1 - 2
< 1



Drought Levels | 2010 - 2017
La Paz County, AZ
Imperial County, CA
Orange County, CA
Riverside County, CA
San Bernardino County, CA
San Diego County, CA
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2010
2011
2012
2013
2014
2015
2016
2017
valid_start

Results



Discussion

- For any given county, the counties that are directly adjacent to it offer the best drought prediction
- Visually, our algorithm was effective at finding Dissimilar Drought Counties
- A total of 45 groups of Dissimilar Drought Counties were identified
- Our algorithm may be able to be help local water companies make informed and drought-conscious decisions for allocating water to various counties within its domain

References

Berndt, Donald J, and James Clifford. "Using Dynamic Time Warping to Find Patterns in Time Series." *AAAI Technical Report*, 26 Apr. 1994, pp. 359–370.

Tsiporkova, Elena. "Dynamic Time Warping Aglorithm." *Emory University Mathematics & Computer Science*, www.mathcs.emory.edu/~lxiong/cs730_s13/share/slides/searching_sigkdd2012_DTW.pdf.

KAN Learning Curve