Computer Science 4013/5013: Artificial Intelligence
Homework 4
Due March 23

Because this assignment is a bit deeper than most homework assignments, you are allowed to pair up and turn in a single assignment. You are not required to do this but it is fine if you want to do so. Groups may be no more than two students. Be sure to put the name of both students on your document that you turn in.

This assignment uses real-world weather data obtained from the Oklahoma Mesonet http://www.mesonet.org. This is a unique weather observation platform for the state of Oklahoma that monitors the weather in each county every 5 minutes. For this homework, I have provided one year of data for the Norman station. The description of the variables can be found at http://www.mesonet.org/index.php/site/about/daily_summaries. I have provided two files on D2L, one for January and February 2015 and one for January and February 2016.

1. **(60 points, all students)** Write a program that uses linear regression to predict the daily maximum temperature for January and February 2016 given all of the information provided for January and February 2015. You should treat each column as a separate f_i in your feature vector. Do not worry about making fancy features that use prior time steps. The goal of this exercise is to learn how regression works. You may write this in any language that you choose but you must provide your code and you must **not** use existing linear regression libraries (there are a lot of them but this is not a lot of code in any language so write it yourself and learn more about regression!). **Make sure you use a small alpha to keep this from diverging! My code worked well for any value of alpha $\leq$ 0.00001. It worked for higher values of alpha but the learning was less smooth.**

   - **(30 points)** for correct code that implements linear regression
   - **(15 points)** for a learning curve that shows the summed squared error versus the iterations. You should run for at least 10 iterations.
   - **(15 points)** for a graph that shows your predictions against the actual observed high temperatures for each day

2. **(40 points, 5013 students only)** Write a program that clusters the 2015 data using k-means clustering. Use the 2015 clusters to predict the daily high-temperature value for each day in 2016 and provide a graph showing your predictions and the observed

values. Again, you may implement this in any language you choose, provided that you write your own code and do not use existing libraries. Turn in your code along with your graphics. Explain why you chose the value of k you chose.

- **(20 points)** for correct code that implements k-means clustering
- **(10 points)** for a learning curve that shows the summed squared error versus k. You should test k for values ranging from 2 to 7 (inclusive)
- **(10 points)** for a graph that shows your predictions against the actual observed high temperatures for each day