UC Berkeley
Python class
AY250

Homework 5

Weather Prediction

due Monday Mar 12, 2018 @ 2pm

In this assignment we will create a database to analyze historical weather data and discovery the relationships between major cities.

1) Find a list of the 50 most travelled airports in the US and make a table containing relevant information, just as name, nearest city, latitude, longitude

2) Build another table that will hold historical weather information, such as min/max temperature, humidity, and precipitation

3) Build a web crawler that will pull historical data from weather underground from 2008 until now and populate your tables accordingly

4) For each pair of cities/airports determine how the daily change of temperature high and precipitation from one city predicts the daily change of the other city 1,3, & 7 days in advance

5) Plot the correlation strengths for the 10 top pairs for all three dates, for temperature and precipitation (separately) as a function of distance. Also make a plot as a function of longitude different. What trends do you see?

Notes:

- You can scrape the weather underground webpages (rather than through the non-free API) like (SFO, 1 year): `https://www.wunderground.com/history/airport/KSFO/2017/3/5/CustomHistory.html?dayend=5&monthend=3&yearend=2018`
- the CSV files can be messy, be careful with what comment character your CSV reader assumes/is set to.
- some data will be missing for some fields. You can use interpolation to fill in that data (eg. `DataFrame.fillna`)
- for correlations between pairs, try `np.corrcoef`