Suggested Weather Data Investigations

Pattern Detection & Correlations

1. Apply a FFT to the hourly temperature time series for one city and plot it as frequency vs frequency amplitude. Discuss whether the frequency content is expected, i.e. does it show annual, season, and daily. (Should we have it more open ended? Like ”Discuss the frequency content you found, what do you think they represent?”)

2. Compare the temperature time series for Charlotte with the other weather information (humidity, wind, etc) in Charlotte. What metrics have the highest correlation with temperature?

3. Compare the temperature time series for Charlotte with the time series for the other cities. What cities have the highest correlation with the Charlotte temperature data?

Model Development & Selection

4. Data reduction using all the metrics for Charlotte and two nearby cities:

- Standardize the data and perform Principal Component Analysis (PCA).
- Plot the variance explained by each principal component and determine how many components are needed to capture most of the variability in the data.
- Examine and interpret what physical or meteorological processes the leading principal components might represent.

5. Using all the data for Charlotte, perform mutual information analysis:

- Compute the mutual information between hourly temperature in Charlotte and each of the other weather variables (humidity, wind speed, pressure, etc.).
- Rank the variables according to their mutual information with temperature.
- Compare the mutual information results with the correlation results from Question 3.
- Are the same variables identified as most strongly related to temperature?
- Are there variables with low correlation but relatively high mutual information?
- Discuss what mutual information reveals about the relationships between weather variables that correlation alone may miss.

6. Linear Regression Using the Full Weather Dataset using all available weather data

- Optional (recommended): Apply the data reduction approach from Problem 4, or another suitable data reduction technique, to retain the most relevant variables.
- Implement two of the following methods: Ordinary Least Squares (OLS), LASSO, Ridge, or Elastic Net. Obtain the fitted models.
- Use your models to predict the weather for the next 2 months and upload the resulting predictions as a CSV file to the leaderboard.
- Evaluate model performance. Examine the learned coefficients and try to interpret them.

7. SINDy Analysis Using the Full Weather Dataset

- Optional (recommended): Apply the data reduction approach from Problem 4, or another suitable data reduction technique, to retain the most relevant variables.
- Apply the SINDy discovery framework to discover the governing equations and obtain the fitted model.
- Use your model to predict the weather for the next 2 months and upload the resulting predictions as a CSV file to the leaderboard.
- Evaluate model performance. Examine the discovered coefficients and try to interpret them.

8. Weather Prediction Using a Neural Network

- Optional: Apply the data reduction approach from Problem 4, or another suitable data reduction technique, to retain the most relevant variables.
- Train a Recurrent Neural Network (RNN) or another machine learning–based regression model of your choice using the dataset.
- Use your model to predict the weather for the next 2 months and upload the resulting predictions as a CSV file to the leaderboard.
- Evaluate model performance. Discuss whether the absence of explicit coefficients affects interpretability compared to previous methods.

9. Be Creative!

- You may use any method of your choice, including combinations of approaches, advanced preprocessing techniques, or methods not covered in this course.
- Clearly describe your methodology and evaluate how well it performs.

Whichever team is ranked first on the leaderboard at the submission deadline will receive 30% extra credit on the course project.