ST440/540 – Exam 2 - Due Monday, April 15

THIS IS AN EXAM - DO NOT DISCUSS THE PROBLEM WITH ANYONE (INCLUDING OTHER STUDENTS OR THE TA)! If you have questions, please email me.

The data you will analyze are from the paper A Feed Forward Neural Network Based on Model Output Statistics for Short-Term Hurricane Intensity Prediction. The data and variable descriptions are on the course website. This paper uses deep learning to improve 24-hour ahead forecasts of hurricane intensity (maximum wind velocity, VMAX). The main prediction model is HWRF, which is a mathematical model based on differential equations. In addition to the forecast, HWRF has many other state variables such as sea surface temperature, longitude, time of year, etc, that are usually discarded. In this analysis, you will determine if including these state variables can improve the HWRF prediction. In the analysis, the most important variables are the observations, `VMAX`, and the HWRF predicitons, `HWRF`. All variables other than `VMAX`, including `HWRF`, can be used as predictors. Your goal is the impute the missing values of `VMAX` using a Bayesian analysis.

1. **Model description:** Describe your model, including prior, and argue this is a reasonable approach for this task.
2. **Model comparisons:** Compare at least two models and select a final model for prediction.
3. **Goodness of fit**: Verify that your choosen model fits the data well. If it does not fit well, select a more appropriate model and repeat steps 2 and 3.
4. **Variable importance**: Determine which predictors are the most important and summarize their effect.
5. **Prediction**: For each observation with a missing `VMAX`, compute a posterior predictive mean and 95% credible interval. Using cross-validation, give an estimate of the mean absolute error (i.e., the average of $|VMAX_i - \widehat{VMAX}_i|$ for observations i) and coverage of 95% intervals.
6. Submit a csv file with four columns:
   (1) `HWRF`: This is exactly column 22 of the original dataset
   (2) `VMAX`: This is column 23 of the original dataset except with posterior mean predictions replacing missing values
   (3) `L`: This is the lower endpoint of the prediction interval (leave empty for non-missing observations)
   (4) `U`: This is the upper endpoint of the prediction interval (leave empty for non-missing observations)

   Save the file as `LastnameFirstname.csv` and do not include any other rows or columns. Submit the file following the instructions on the moodle page.

Your paper should be written as a professional document with full sentences, clearly labeled figures and tables and few spelling/grammar errors. Organize your report with subsections corresponding to the questions above. Summarize your analysis in a PDF document that is **no more than two pages long** (12 font, single spaced, one-inch margins). Append your code to the end of this document and submit a single PDF. **In-class students should turn in the exam in class on Monday, April 15. Online students should submit the exam on moodle.**

HAVE FUN!