University of Colorado
Department of Civil, Environmental and Architectural Engineering
Advanced Data Analysis Techniques
CVEN 6833
Homework 1
Due: 10/09/2018

Topics: Surface Fitting – *Linear, GLM, Local Polynomials, Linear and Local Polynomial GLMs, GAM, Spatial Models, Hierarchical and Bayesian methods*

Please present your work neatly. Organization of R-commands, functions will fetch 15% of points.

1. Derive the link function, Fischer score and Information matrix for Poisson distribution.
As a bonus do the same for Gamma distribution.

2. Mean January and July precipitation at 491 locations in Colorado based on data for the period 1980-2002 can be obtained from http://cires.colorado.edu/~aslater/CVEN_6833/colo_pcp_monthly.html
The first column in the file colo_precip.dat is the latitude, the second the longitude, the third the elevation (m) and the last column is the value of the average precipitation (mm).
The location of stations, topography map etc. can be found at http://cires1.colorado.edu/~aslater/CVEN_6833/colo_pcp.html

i. Display the average precipitation for January and July along with the topography as spatial maps.

The first objective is to obtain a predictive model for burned area for each fire as a function of co-variates. For January and July separately:

ii. Fit a 'best' linear regression model (use one of the objective functions - GCV, AIC, SBC or PRESS; you can also try a couple of them to see any differences). This entails fitting the model with all possible combinations of covariates (Latitude, Longitude and Elevation) and selecting the model with the minimum objective function.
Show the scatterplot of observed and modeled precipitation along with the 1:1 line.
iii. Perform ANOVA (i.e. model significance) and model diagnostics (i.e., check the assumptions of the residuals – Normality, independence, homoskedasticity).
iiv. Compute drop-one cross-validated estimates from the best model and scatterplot them against the observed values with the 1:1 line. This and the scatterplot in (ii) above, is to visually see how the model performs in a fitting and cross-validated mode.
v. Drop 10% of observations, fit the model (i.e., the 'best' model from i. above) to the rest of the data and predict the dropped points. Compute RMSE and correlation and show them as boxplots.
vi. Spatially map the model estimates and the standard error.
vii. Briefly discuss what you find [bullet points are fine]

3. Repeat 2. by fitting a ***GLM*** with appropriate link function.

4. Repeat 2. with ***Local polynomial method***. For step vi.
For bonus points, write a code that estimates the value of the function using the L vector and compare with the estimates from the 'predict' command.

5. Repeat 4 with Local Polynomial method but using the appropriate link function (i.e., '***Local GLM***').
[For the Local Polynomial approach the 'best model' involves fitting the best subset of predictors and the smoothing parameter, alpha. You can also compare the GCV from these four different methods.]
Briefly discuss the results from the local polynomial approach and compare them to linear regression.

6. Repeat 4. by fitting a ***Generalized Additive Model*** (GAM) and compare with the GAM fitted in a local polynomial framework.

7. Daily Precipitation at the 491 locations in Colorado is available for three days January 11, 12 and 13, 1997 from http://cires.colorado.edu/~aslater/CVEN_6833/colo_pcp_daily.html .
Select one of the days (try Jan 11 or 12), convert the daily rainfall into at each station to a binary variable of no precipitation (0) and a nonzero precipitation amount (1).
i. Fit a 'best' GLM with the appropriate link function using one of the objective functions. Test the model goodness using ANOVA
ii. Estimate the function on the DEM grade and plot the surface. Also plot the standard error.
Compare them with the surface plot of the elevation and also with the results from Slater and Clark (2006) Figure 4. They use a quasi-local GLM approach but I would imagine the results should be similar.

8. Repeat 7. with a ***Local GLM Logistic Regression***

9. Estimate the spatial surface of January precipitation using ***Kriging***
i. Fit a variogram
ii. Repeat iii – vi of problem 1.

10. Repeat 9. with a ***Hierarchical Spatial Model. [Fit a best linear model as in problem 1 and perform Kriging on the residuals]***

11. Repeat 9. with a ***Bayesian Hierarchical Spatial Model*** (see Verdin et al. 2015, for tips)
[In the Bayesian models, plot the posterior historgram/PDF of the parameters; spatial maps of posterior mean and standard error]

12. Develop a ***Bayesian Hierarchical Spatial Model*** for problem 7

13. I want you to reflect on the suite of analyses performed above on the spatial precipitation data – in particular, the relative performance of the methods, their advantages and disadvantages and potential application of these methods on a problem/data set of your interest. Keep this short and crisp (bullet points are fine).