ST433/533 Applied Spatial Statistics

Lab activity for 9/9/2020

A. CLARIFICATION QUESTIONS

(1) In the air pollution example, why do we get big prediction standard errors when there are limited observations?

The information used to make spatial prediction is the response of the nearby observations, and if there are no nearby observations there is no information, i.e., large uncertainty.

(2) Is the equation we continue to fit, with the lon, lat, lon*lat, lon^2, lat^2, an example of an actual exponential model, or just an example of a model with real covariates substituted in

This is an actual model. But if you had other covariates like elevation or distance to a highway you could include these too.

(3) Can you explain what the proc.time() function is doing in the Kriging predictions code? I'm confused by the "tock-tick" code since they are the same and by subtracting one from the other would equal zero.

The function proc.time gives the current time, so these lines measure how long it take to execute the code in between the “tick” and “tock” lines.

(4) What is the difference between Kriging and using weighted least squares?

Weighted least squares is usually used to estimate a parameter, whereas Kriging is used for prediction.

(5) In lecture, it was said that kriging doesn't assume gaussian data but can be derived from the properties of the multivariate normal distribution. If it's derived from the MVN distribution, why doesn't it assume that the data are normally distributed (gaussian)? I was just confused if kriging could be derived from other distributions or not.

The original Kriging derivation only assumes that you know the mean and covariance functions and it does not assume normality. It turns out that if you assume normality you get the same prediction formula.

(6) Will there be office hour during the exam week?

Yes.

(7) Why is Universal Kriging the most common/popular for spatial predictions than Local Kriging?

Local Kriging is an approximation to universal Kriging. It is faster, but slightly less accurate.

(8) How to decide which type of Kriging should I apply.

See student discussion question (4)

(9) We know kriging is the best unbiased linear predictor (BLUP). Does that mean we cannot (or should not) use lat^2 and long^2 as observation values?

I assume that by ``observation values'' you mean covariates. Kriging is optimal given you have the correct mean and covariance models. It may be that you need to include lat^2 in the mean to have (or at least a reasonable approximation to) the correct model.

(10) On slide 8, there is an equation for prediction. On slide 9, there is also an equation for prediction. Are these two equations equivalent to each other? To me, the equation on slide 8 uses mean value of y0, but equation on slide 9 doesn't use it.

These are similar expressions, but Slide 9 ignores beta-hat (i.e., pretends there are no covariates) to show the relationship between the kriging weights and the covariance function in a simple case.

B. STUDENT DISCUSSION QUESTIONS

(1) If we do not know Y_0, how can I know cov(Y_0,Y_i)

We need to make assumptions! Then we can estimate the covariance.

(2) At what point is there too little local data too perform Kriging prediction at a specific point?

More data the better…Kriging is extrapolation/interpolation so w/o data it will be a large gradient. It also depends on the correlation parameters (large range/small nugget need less data).

(3) In the Kriging predictions map of the ozone (ppm), would the predictions appear less smooth if the point density of the prediction sites (s0) was lower? Or is the smoothness determined by the model fit of the data regardless of the number of prediction sites?

Smoothness is largely determined by the estimated parameters, but in places with no data the prediction surface will always be smooth.

(4) What are real world examples for when it would be best to use i) Simple Kriging, ii) Ordinary Kriging, iii) Universal Kriging, and iv) Local Kriging. Why is simple Kriging necessary if the mean is already known?

i) Air pollution after removing the mean

ii) Case with no known covariate…say the satellite data

iii) The is the more general example with some covariates, say Zillow data

iv) Large datasets, say the satellite data

(5) What's the best method to go about adjusting how high to make the nugget or sill?

Variogram or MLE

(6) Would it be possible to use global sensitivity analyses to find which covariates are contributing most to the uncertainty in kriging weights? What would an analysis like this look like?

Yes, the weights are a function of covariates and parameter estimates and this could tell us how sensitive the Kriging variance is to these values.

(7) Like any other spatial models, extreme values (too large or too small readings) can impact the prediction in kriging. Are there any ways to address extreme values while performing kriging analysis?

Remove outliers for model fitting, but keep them for prediction. Or use local Kriging.

(8) How to balance between mse, coverage, etc. when selecting the models?

It depends on your priorities.

(9) Should covariate terms be dropped if the p-values show they are insignificant?

Yes, dropping it reduces the risk of overfitting. (forward/backward selection.)

(10) What would be a valid way to compare different spatial prediction methods (ie. polynomial/basis functions of s in covariate matrix vs. random forests with s as covariate vs. Kriging)?

Cross validation, and plots of fitted values.

(11) What are the options for spatial prediction if the covariance is unknown.

Can't use Kriging, prediction requires some assumptions and if spatial correlation isn't the best assumption to leverage you will have to find something else.

C. BRIAN'S DISCUSSION QUESTIONS

(1) Can you apply Kriging with

(a) an anisotropic covariance? BR: Yes, but you have to estimate the anisotropy parameters (medium hard)

(b) a nonstationary covariance? BR: Yes, but you have to estimate the covariance function (very hard)

(c) non-Gaussian data? BR: Technically yes, as long as you have the covariance function. You might be better off if you first transform the data to be approximately Gaussian. Kriging is not the best thing to do for say binary data or count data. We'll cover better methods in the generalized linear model section.

(2) What are the relative advantages of cross-validation versus goodness-of-fit criteria like AIC/BIC?

BR: AIC/BIC are nice because they only require fitting each model once, but they require nested models. CV is slow, but more flexible and I think more to the point.

(3) Say you fit a Kriging model and perform cross validation, but the empirical coverage of the 95% prediction intervals is only 80%. What are some things you might try next?

Check the covariance by variogram, check for outliers, check for stationarity, look at covariance in subregions, Bayes (to account for uncertainty in standard error).

(4) What are the main assumptions behind the Kriging algorithm and how would you examine each assumption using data?

We have the right spatial covariance function and mean function. Check mean and covariance as in (3).

(5) **Informative sampling** refers to selecting the sample location $s_1,\ldots,s_n$ based on prior knowledge of the response. For example, we record the traffic density at n intersections in the city, but we only sample intersections that have a recent report of high congestion. We then use these data to build a spatial prediction model to be applied to all intersections in the city.

(a) What problems would this cause for Kriging? Bias! (would be true for any model)

(b) How might informative sampling manifest in a cross-validation analysis? No! Both the training and testing set would have this bias.

(c) How might you adjust for informative sampling? Add a covariate that explains this bias, or maybe collect additional data.