Applying for the data science position

Reza

April 07, 2016

Dear

I am writing to express my interest in working with Zapier as data scientist. I am doing the last steps of doctorate in Mathematics in FSU Jena Germany. I have plenty of experience in research and teaching. Alongside pure math, I cooperated with experts in other fields including economists and urban planners.

To name some, I developed an algorithm for assessing accessibility in cities using big graphs and programmed it in MatLab. I also am familiar with, by nature statistical, techniques of remote sensing, and help urban planners running them through R programming. Together with economists, we studied financial derivatives with aim of understanding their nature and their relationship with gambling.

Now after several years of mainly doing academic jobs, I would like to enter to a more practical yet active sector and found the data analysis really interesting and perfectly fitted to my theoretical background. In the university and other training, I did several courses in Statistics, Statistical inference, operational research, philosophy of science, machine learning, programming and data structures as well as data banks and SQL. I am already advanced in R programming and familiar with python with interest in learning it them more deeply. Some works of mine are in my github page.

In terms of transferable skills, I learned how to translate real world problems into mathematical statements and vice versa. I am good in analyzing and problem solving as key skills in mathematics. I did cooperated with teams of different scientific and social backgrounds. Also for several years I thought mathematics to students of various scientific and social backgrounds through which you have to be enormously patient, friendly and understanding as well as being able to express complicated problems in understandable manners. Last but not least, I am a good independent fast leaner and great in finding the way the jobs are done simply using online resources.

In R I love the "dplyr"" and "data.table" packages for their fast and efficient capability of cleaning and summarizing the data. "raster" is a very nice package for working with grid ?data?. For plotting one should not underestimate the base plotting system of R but whenever it is too much coding needed using it, I prefer to use "ggplot2". Rmarkdown in combination with LaTex is a great tool for making reports. In machine learning techniques, neural network is really good however sometimes I found it is too slow. Random forest worked quite good.

Because of my thesis work, I can not yet work full time. Plus working with the interesting, creative and creditable Zapier is a huge oportunity to earn invaluable experiemce and credit for my career prospect. So I would like to propose a two-month unpaid remote trainee program. I would do the jobs you expect from a data analyst. I am sure I can rapidly learn whatever needed for the job and I haven't learned yet.

Enclosed please find my resume and a sample report of a work of mine. I am looking forward to hearing from you and ready to start from April.

With highest regards

Reza Hosseini

Explanation of a work

MODIS is a satellite which provides mainly two kind of products. First, land surface temperature (LST) and second Normalized Difference Vegetation Index (NDVI), an index between -1 and 1 of which higher values indicate more dense vegetation. This data are provided in weekly intervals. The problem is that it is really coerce, with resolution of 960m. On the other hand we had an image of the SPOT satellite of 1.5m resolution

which unfortunately does not provide LST information but fortunately provides enough data for computing NDVI. On the other hand, it is known that NDVI and LST have an inverse relationship. We use this fact to compute LST in the resolution of SPOT.

Briefly, I computed two linear models. One to convert NDVI of SPOT to NDVI of MODIS and another to convert NDVI to LST. The first one is quite straightforward. But the second model had to be based on not all pixels but just the so called hot edge pixels. i.e. pixels which are among the hottest 5 percent when NDVI is binned in 0.3 unit intervals. Scatter plot below shows NDVI vs LST. Red dots are hot edges and green squares indicate the fitted linear model.
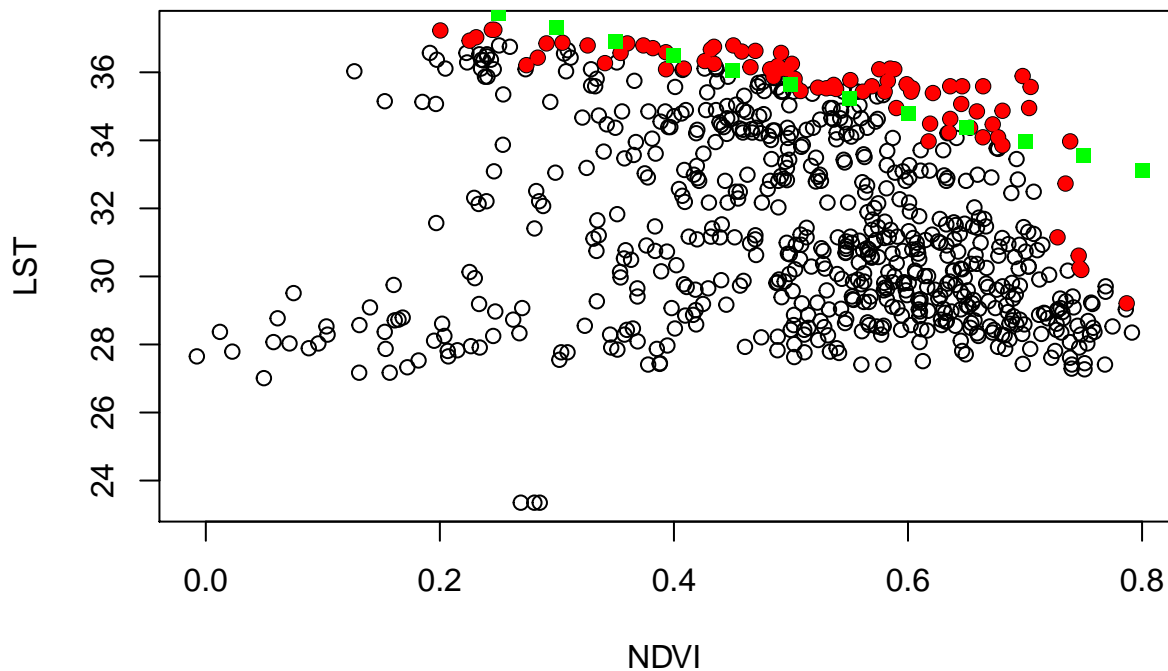
To refine the results, I used the fact that the temperature of a point is dependent to its surrounding points. To take into account this fact, I computed the residuals of the model above with respect to all pixels. Then trained a neural network with NDVI values of each pixel and its eight surrounding pixels as input and the pixels residual as output.

The