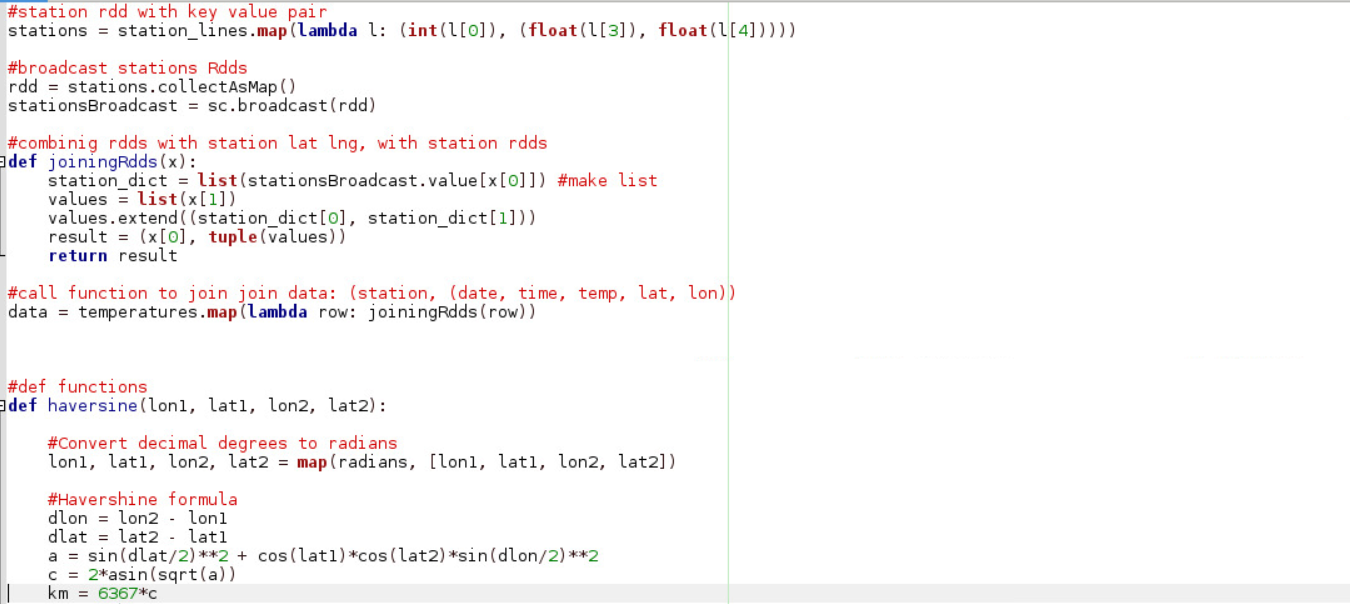
BDA3

Rabnawaz jansher & Saman Zahid

5/26/2018

# BDA3 - Machine Learning with Spark - Exercises

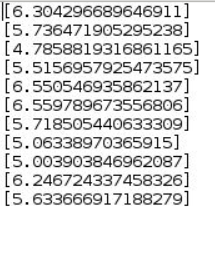












***Q.1 - Show that your choice for the kernelsí width is sensible, i.e. it gives more weight to closer points. Discuss why your definition of closeness is reasonable.***

By running the code on different values of h, we have observed the fact that higher the values of h, lesser will be the effect on kernel. It is because of the fact that when we take higher values of h, we get the wider kernel which means that more points are being considered for calculation. Thus, we have chosen a reasonably smaller value of h so that more weight is assigned to the close points which will be more effective in calculation thus giving better predictions.

***Q.2 - It is quite likely that the predicted temperatures do not differ much from one another. Do you think that the reason may be that the three Gaussian kernels are independent one of another? If so, propose an improved kernel, e.g. propose an alternative way of combining the three Gaussian kernels described above.***

The predicted temperatures do not differ much from one another, it could be because the we are summing over the three independent kernels, which might gives result on the basis of all measures that is if distance is small but time difference is large, then time alone will also effect the prediction (because of kernel independency).

The alternate way could be by combining kernels by taking product of three kernels instead of summing them up and then averaging them. By combining the kernels via multiplication, proportions of the three different kernels is preserved.