EP2300: Management of Networks and Networked Systems
Estimating Conformance to Service Level Agreements (SLAs) using Machine Learning
Task I - Introduction to statistical computation using R

Surya Seetharaman

September 10, 2016

1. Compute the following statistics for each component of X and Y: mean, maximum, minimum, 25th percentile, 90th percentile, standard deviation, and variance.

    Ans : These values were obtained by using the built-in functions in R and the values obtained are as specified in the Table 1.

Table 1: Statistics of the components in X and Y datasets

Title	mean	MIN	MAX	25th PCTL	90th PCTL	SD	VAR
all_..idle	9.064981	0	69.54	0	38.621	16.122822	259.945398
X..memused	89.137517	73.03	97.84	82.965	96.77	8.183662	66.972324
proc.s	7.683303	0	48	0	20	8.532606	72.805357
cswch.s	54045.87	11398	83880	31302	72135.1	19497.81	380164654.86
file.nr	2656.3333	2304	2976	2496	2880	196.1107	38459.4254
sum_intr.s	19978.041	10393	35536	16678	28228.4	4797.271	23013812.168
ldavg.1	75.87577	11.13	147.47	28.2	127.993	43.86244	1923.91404
tcpsck	48.99750	21	87	34	71	15.87116	251.89358
pgfree.s	72872.15	15928	145874	61601.75	97532.5	19504.32	380418544.51
DispFrames	18.818394	0	30.39	13.39	24.61	5.219756	27.245855

2. Compute the following quantities:
    a) the number of observations with memory usage larger than 80% : **2875**
    b) the average number of used TCP sockets for observations with more than 18000 interrupts/sec : **46.34731**
    c) the minimum memory utilization for observations with CPU idle time lower than 20% : **73.03**

3. Produce the following plots:

a) Time series of percentage of idle CPU and of used memory (both in a single plot);
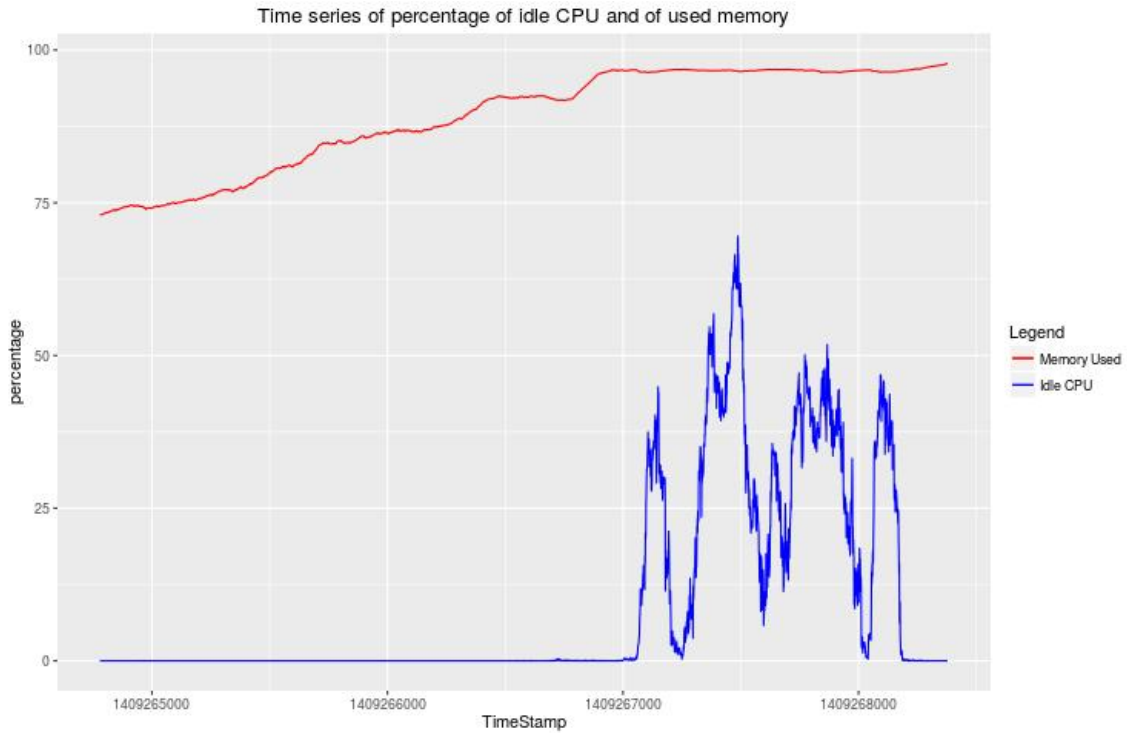

Figure 1: Time series of percentage of idle CPU and of used memory

When plotting the TimeStamp versus the idle CPU time and used memory in R, the graph as seen in figure 1 was obtained. I used the library ggplot() to plot this graph. As we can see, the memory used is high almost above 75% always, and keeps increasing with time (with little irregularities) while the idle CPU plot can be seen to be stagnant at zero for the initial timeperiod and then shows an irregular behaviour.

b) Density plots of idle CPU and of used memory.

The density plots for idle CPU and Used memory are shown in Figures 2 and 3. The plot for idle CPU is having its peak around zero while that for the used memory is having medium and high peaks throughout with the highest being at the end.

Note : The program can be run using the command **R --vanilla < task1.R --args X.csv Y.csv** , task1.R being the name of the source code file.
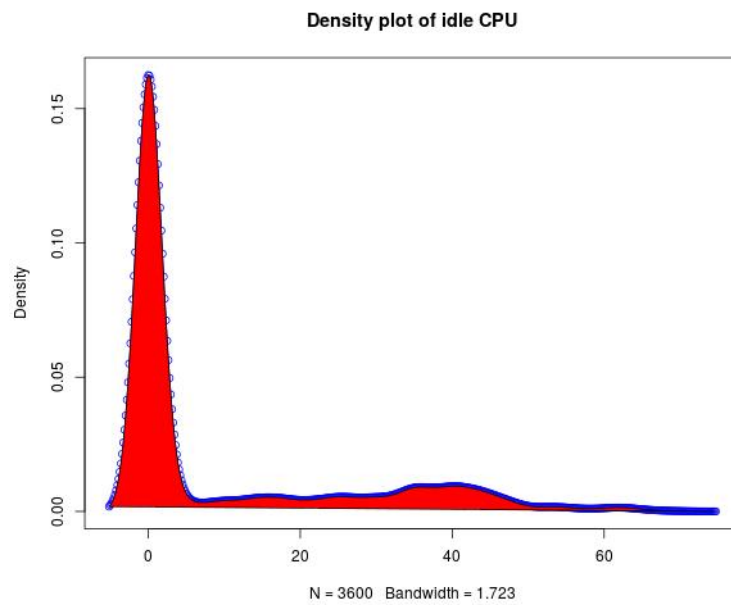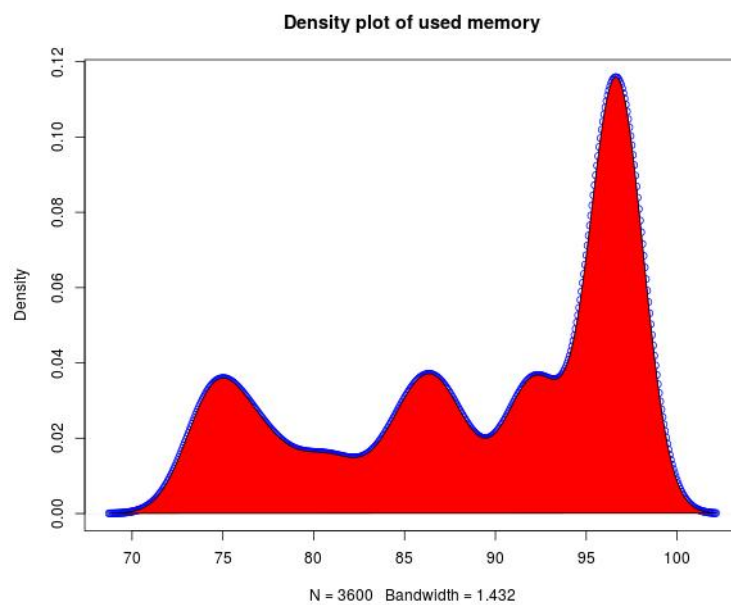
Figure 2: Density plot of idle CPU

Figure 3: Density plot of used memory