OBAFEMI AWOLOWO UNIVERSITY	EEE 526 Machine Learning and Intelligent Control	Dr. Tokunbo Ogunfunmi
Mid Term Exam (Take-Home Exam, Open-Book)		

Name: KOMOLAFE, ELISHA AYOBAMI

I will not give or get unpermitted aid in this exam. and I will report any honor code violations observed by me.

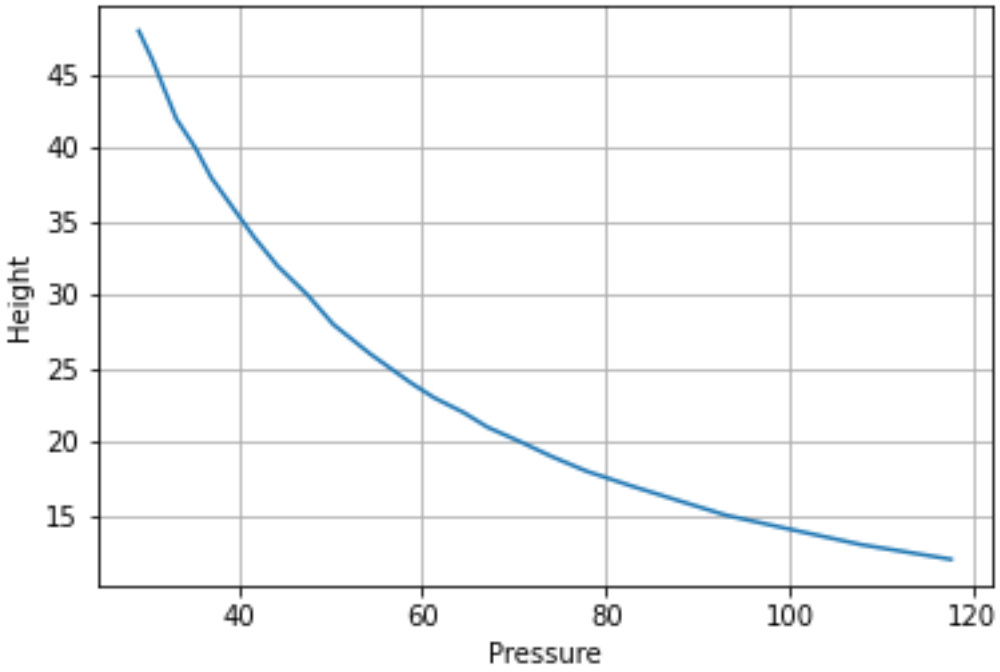
Signature: EAKomolafe ← Your signature means you really

Grades:

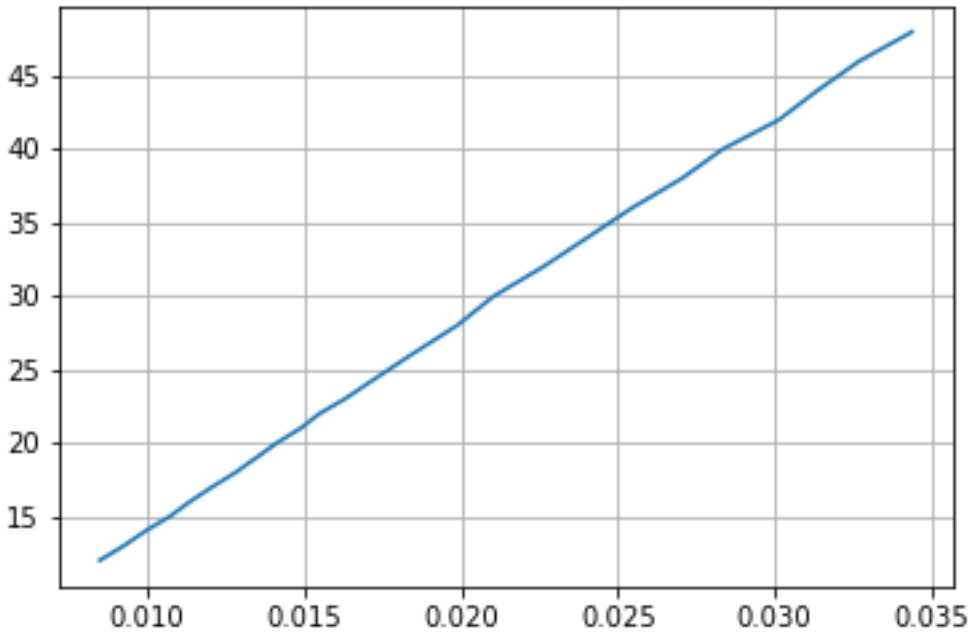
	Possible Score	Your Score
Problem 1	100	
Problem 2	200	
TOTAL	**300**	

Problem 1:

a. Plot of Pressure against Height:



b. Plot of the reciprocal of Pressure against Height:



c. The plots show a clear relationship among the 2 variables,
in the first plot from a. it can been seen that the as the Pressure of the gas increases the

Height decreases showing the relationship is in the inverse, but not linear
While the second plot shows the relationship between the height and inverse of the pressure. In this case as the reciprocal of the pressure increases, the height of the gas also increases showing a linear relationship that is positive as well. Proving that Boyle's hypothesis is correct.

d. Intercept value=-6.476×10^{-05}, p-value = 0.278.
R^2 Value: 1

e. The hypothesis can be tested by using the t test method, available in the summary of the model made from the data of the Boyle experiment. The function t_test can also be used with any hypothesis of choice, the hypothesis chosen for this test is assuming the intercept is equal to 0, the answer is -1.111

f. Boyle's hypothesis is that the Pressure and Volume of a gas are inversely proportional to each other. From the results of the simple linear regression model it is seen that the model has a inverse proportional relationship among the pressure and the Height and this supports Boyle's Hypothesis.

Problem 2:

a. Patterns in the Data:
from the scatter matrix of all the quantitative data we can see that the GRE, CGPA and TOEFL scores have correlative relationship with all other predictors, but not uniform.
The output has a correlation also with most of the predictors except research which is a qualitative value but normalized.
It can also be noticed that the qualitative normalized values have no correction with the rest of the predictors.

b. In the Jupiter notebook, a multiple linear regression model with all the predictors in it

c. To test for the statistical significance of the predictors, the T and F tests can be used, but the ANOVA can show all the results together, and the code for this is also shown in the Jupiter notebook.

d. We can use the t_test to determine if a predictor is statistically significant to the model, if the value is higher the parameter is significant to the models and vice versa. The hypothesis for the t_test for the code assumes all the values are 0.
the most significant predictors are from the highest magnitude: CGPA, Letter of Recommendation, research, GRE Score, TOEFL Score and University rating. The statement of purpose had a negative effect on the model.

e. A predictor is significant when the t value is higher than the others and has a positive value, and with the F-test it shows whether the group of variables are significant together.

f. Determine if multicollinearity will become a problem for this problem:
when none of the t-values are not significant and the F-test is, it can indicate collinearity. Since most of the predictors are significant and the F-test also as well there will not be a problem of collinearity.

g. An suitable hypothesis is that all the significant parameters equals to 0, i.e. gre=0,toefl=0,ur=0,LOR=0,CGPA=0,Research=0 and the alternate hypothesis being the parameters less than 0 being there are not significant.

h. From the linear regression results, I noticed that the statement of purpose parameter does not contribute to the model and when removed the model is the same and the hypothesis is supported since the other parameters are significant making the hypothesis accepted.

i. Comparing forecasts using a model using all the significant variable and a model that has the most significant variables. From the first 10 values of the result compared to the models, between the models the result is very similar apart from some few difference due to significant factors not being used.
and comparing both to the real data both models have different predicted values to the real one and there are some values that have a large difference especially at 10, 8 and 7.