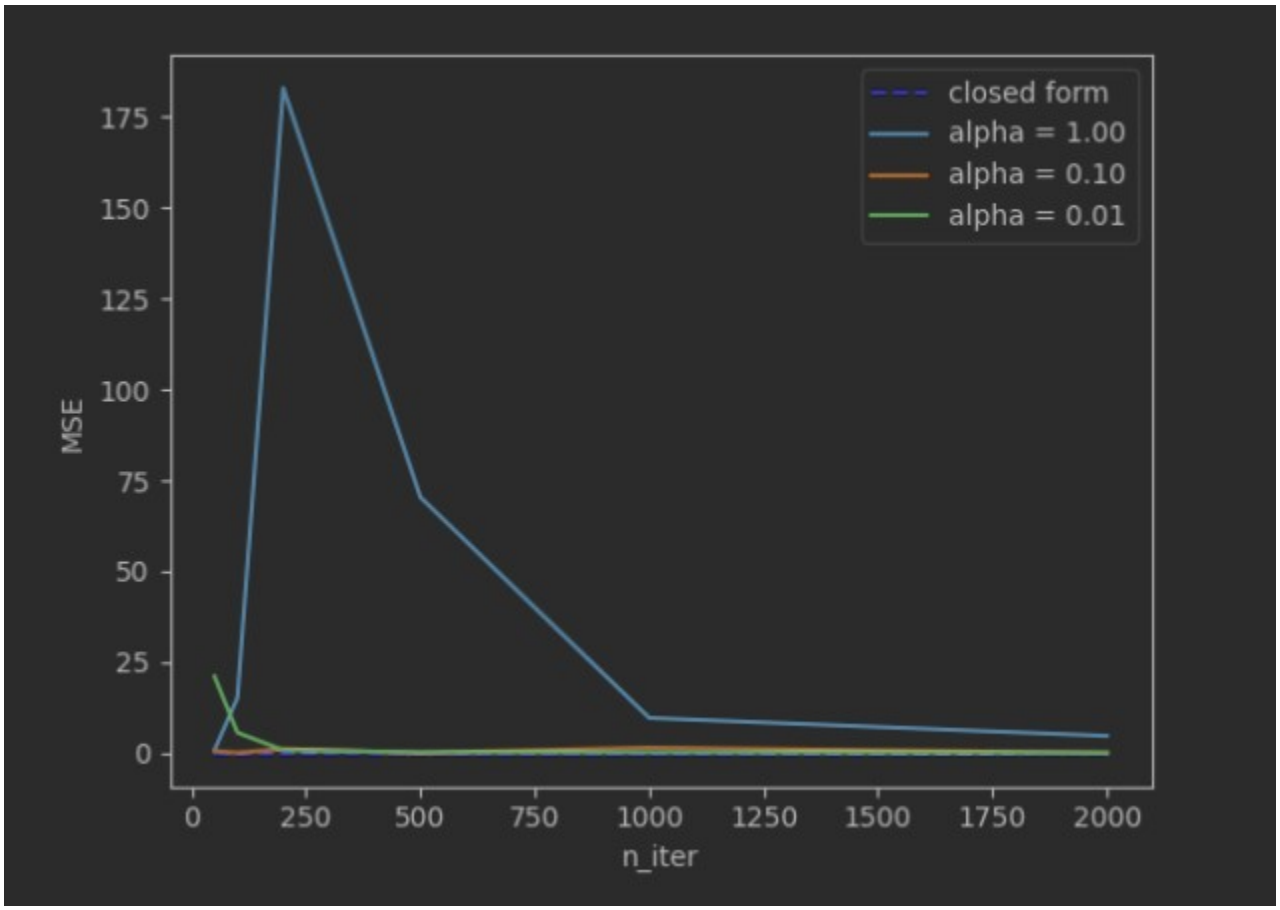
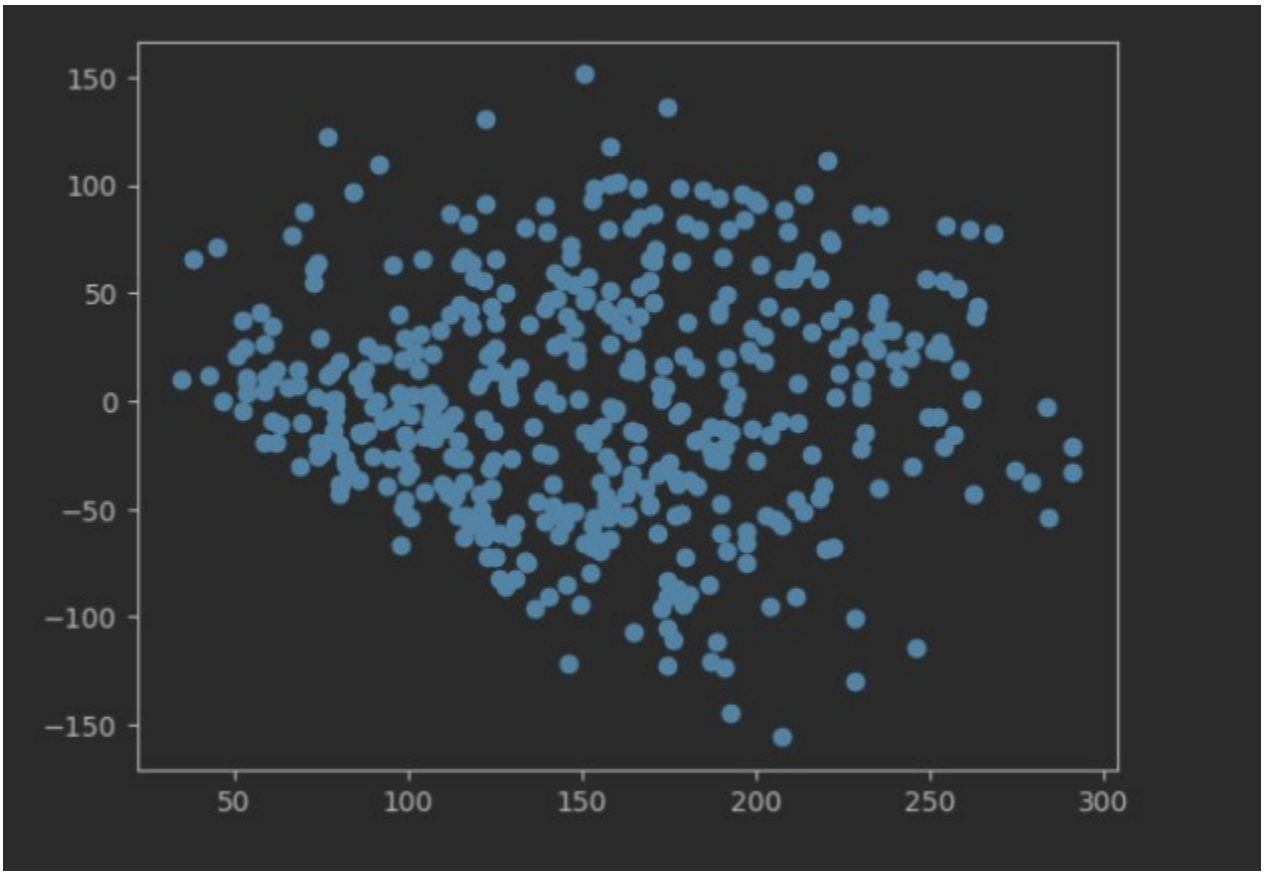
Task 3



I assume that the MSE shoots up that way because if you use an alpha that's too large it might overshoot in one of the iterations and deviates way too far. In turn this means with an alpha that's too large you need a lot more iterations to bring down the MSE to an acceptable level than with a smaller alpha. An alpha that's too small also takes a while to bring down the MSE as you can see by looking at the green curve.

Task 4



Interpretation of residuals

Linearity: is not fulfilled, there is no linear relationship

Independence: They're independent. You cannot predict if the next sample is higher or lower

Normality: is not fulfilled, the deviations do not follow a normal distribution

Equality of variance: is not fullfilled, the distribution is larger the higher the x value

Answers to questions Task 4b

The further the coefficient from 0, the greater the linear relationship. This lets me assume that these three features are the most influental: s1, s5, bmi

The sign means either a negative linear relationship or a positive linear relationship

I would exclude the feature with the least linear relationship which is age