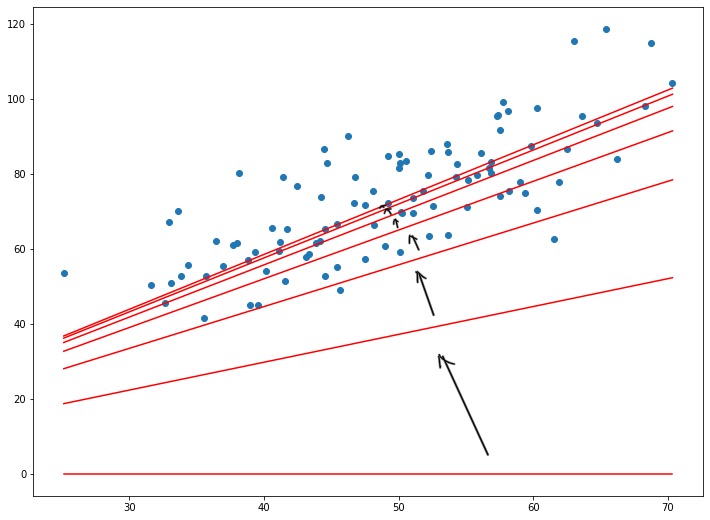
**Discussion Questions**

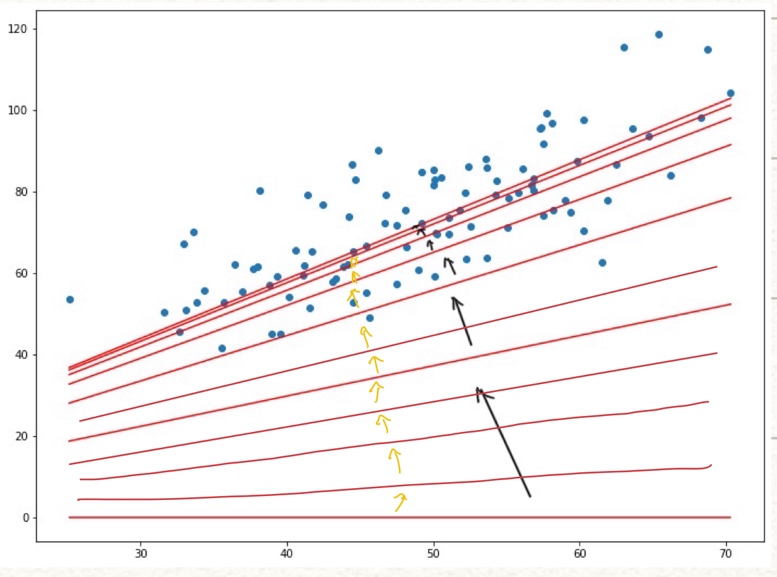
1.What validation have you provided to show that your gradient descent implementation works? Discuss the plots you are submitting.



From the plot we can see that the regression line started at a location really far from the points of data set. Then after each iteration, it got closer and closer to the center of these points. By visualizing, we can stated that the program worked since the regression line was converging to be the best-fit line.

2.Discuss the difference in performance between the fixed step-size *α* and the analytic *α* in terms of the numbers of iterations to converge to the solution. How did you choose the fixed step-size?

The iterations when we use the analytic alpha is more efficient (i.e its looped in a smaller number of time). Because when we use the fixed step-size the line only can move in a small step of the d direction instead of go all the way along that direction.



We can see the concept of fixed steps demonstrated in yellow arrow compared to black arrow.

I started at alpha = 0.1 and tried different steps size until it get to a nice plot. I also print out a0 and a1 along side so I know I should increase or decrease the step size for next try.

3.Did your algorithm converge for every choice of fixed step-size? If not, what causes the algorithm to diverge?

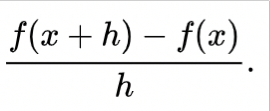
My algorithm did not converge for every choice of fixed step-size. The algorithm can be diverge when the step size is either too small or too big. The reason is that, if the steps size get too big then it could just pass through the right region without noticing and cannot go back. On another hand, if the step size is too small, If step size is too small it doesn’t diverge, but it can not converge because it hits max iterations before reaching the min. Assuming you can derive the analytic step size it’s always usable (even if you’re right and it’s difficult to derive), the issue comes up when it’s not differentiable at every point (discontinuous)

4.Can we always use an analytic value of *α* when minimizing a general function *f* (*x*)? Why or why not?

In some complicated situation, we cannot always use an analytic value of *α* when minimizing a general function *f* (*x*). If analytic value of alpha is provided then we can use it to minimize the general function f(x). However, there will be some situation when the function is too complicated too find an analytic value of alpha.

5.Suppose you didn’t have a closed-form expression for the objective function, that is you only have a black box that you can pass inputs into and get function evaluations out of. Can you still apply Gradient Descent? If so, how? (Hint: can you approximate derivatives numerically? If so how?).

If we only have a black box, we will still be able to apply the Gradient Descent. We can approximated derivatives using the definition of derivative method where we have the derivative can be calculated as1



**Appendix:**

* I denoted my *x* as *x\_* , *b* as *b\_* and *AT* as *At* and *rT* as *Rt*
* I have 2 version of code since one of them is not work properly so I try different way.

**Reference:**

1. https://en.wikipedia.org/wiki/Numerical\_differentiation#/media/File:Derivative.svg

2. https://towardsdatascience.com/linear-regression-using-gradient-descent-97a6c8700931