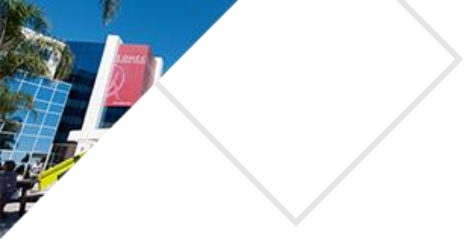
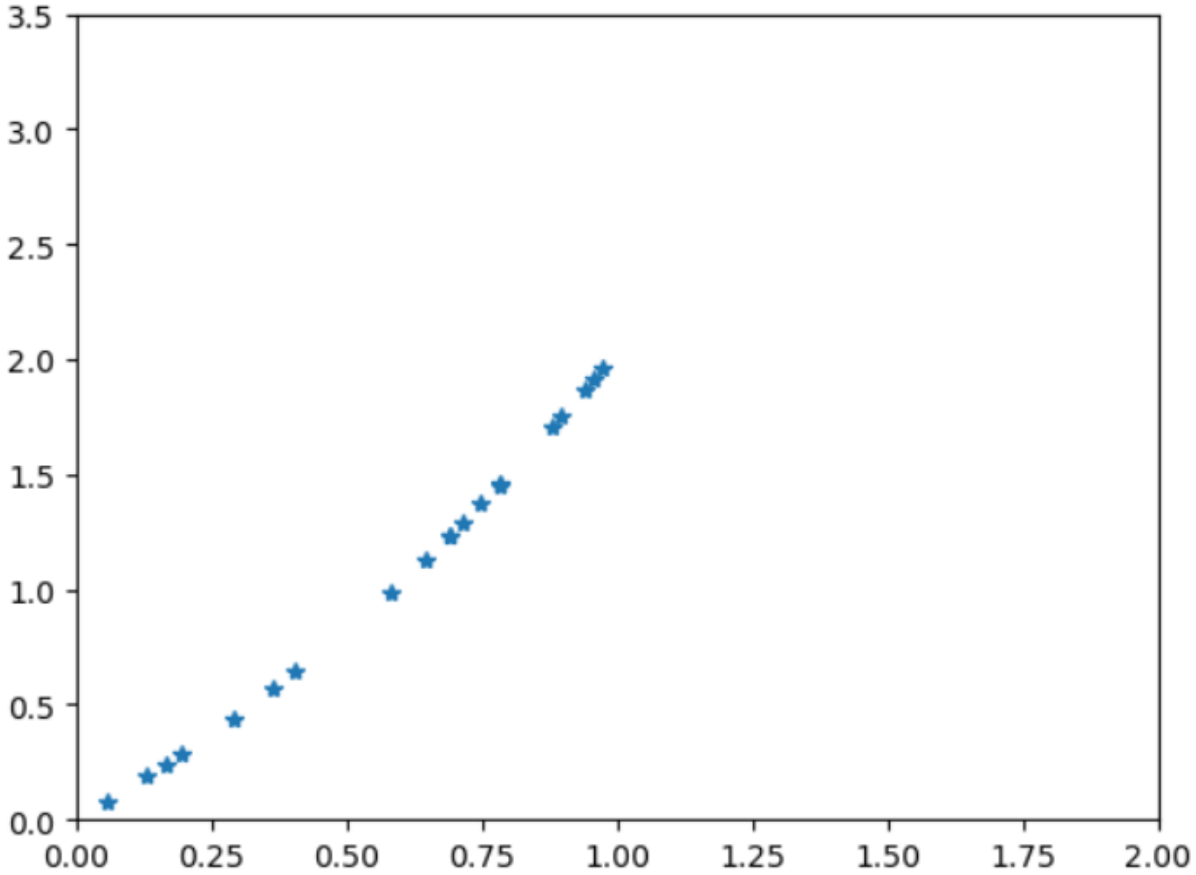


INTERPOLATION VS EXTRAPOLATION

Interpolation

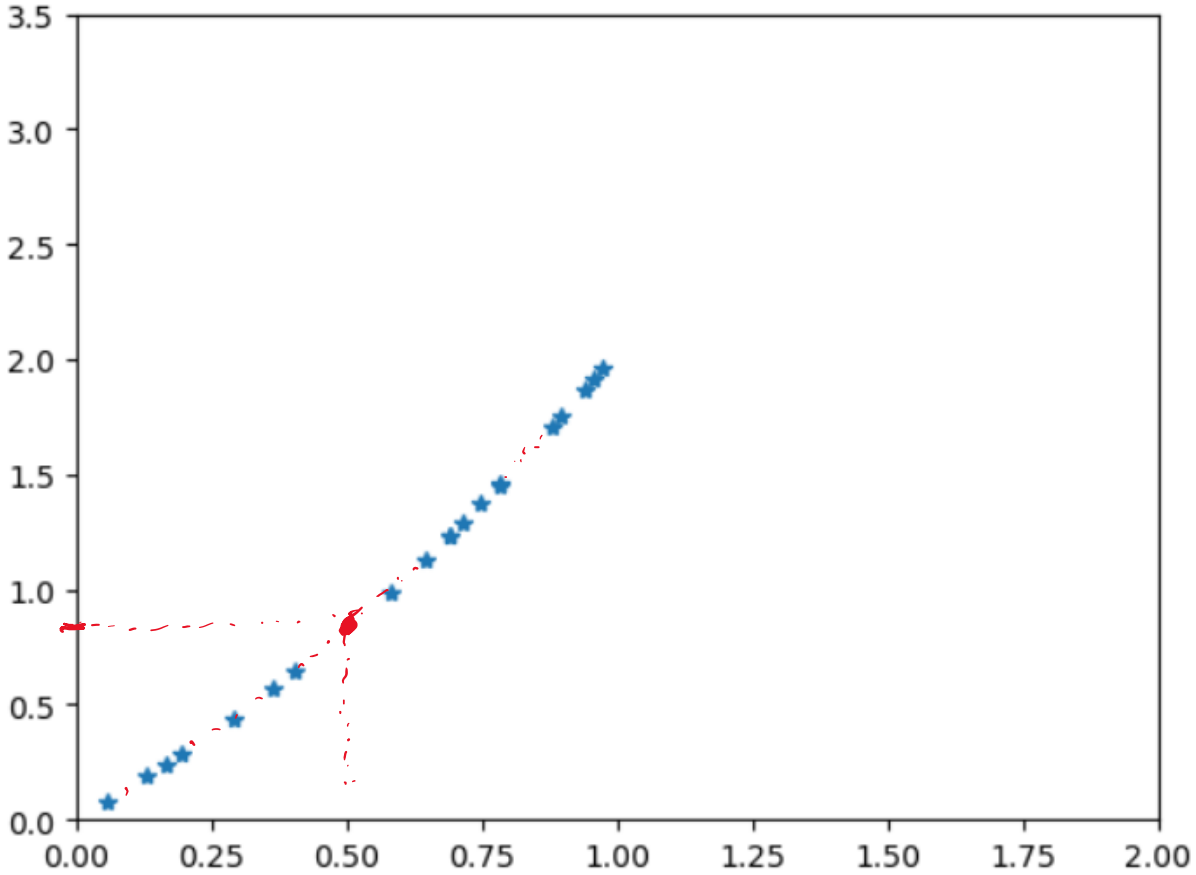


Suppose that we have some data points, but with gaps between the points.

Interpolation consist in estimating the values of the missing points, probably using a reasonable model.

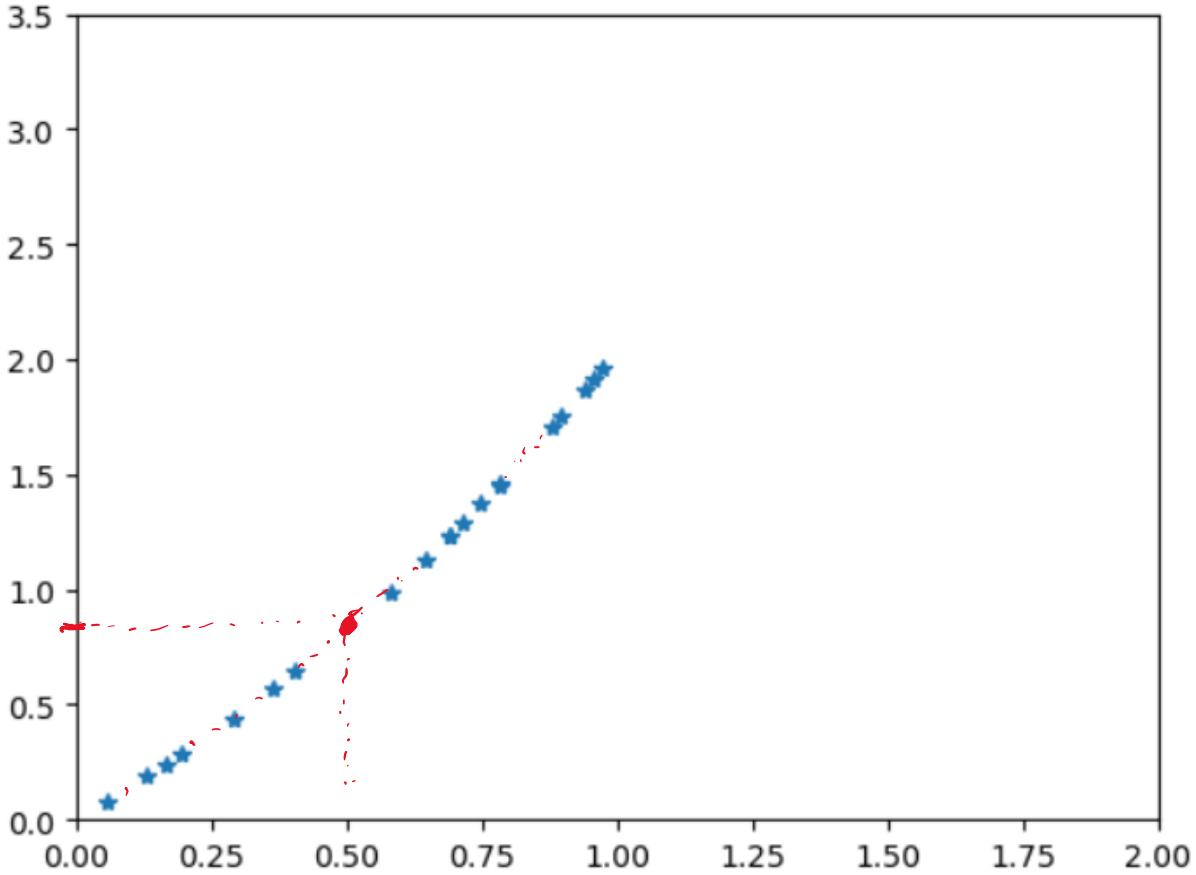
For instance, we could try to make a nonlinear regression to these data points using a quadratic or a cubic polynomial, to find out the plausible value at x=0,5

Interpolation



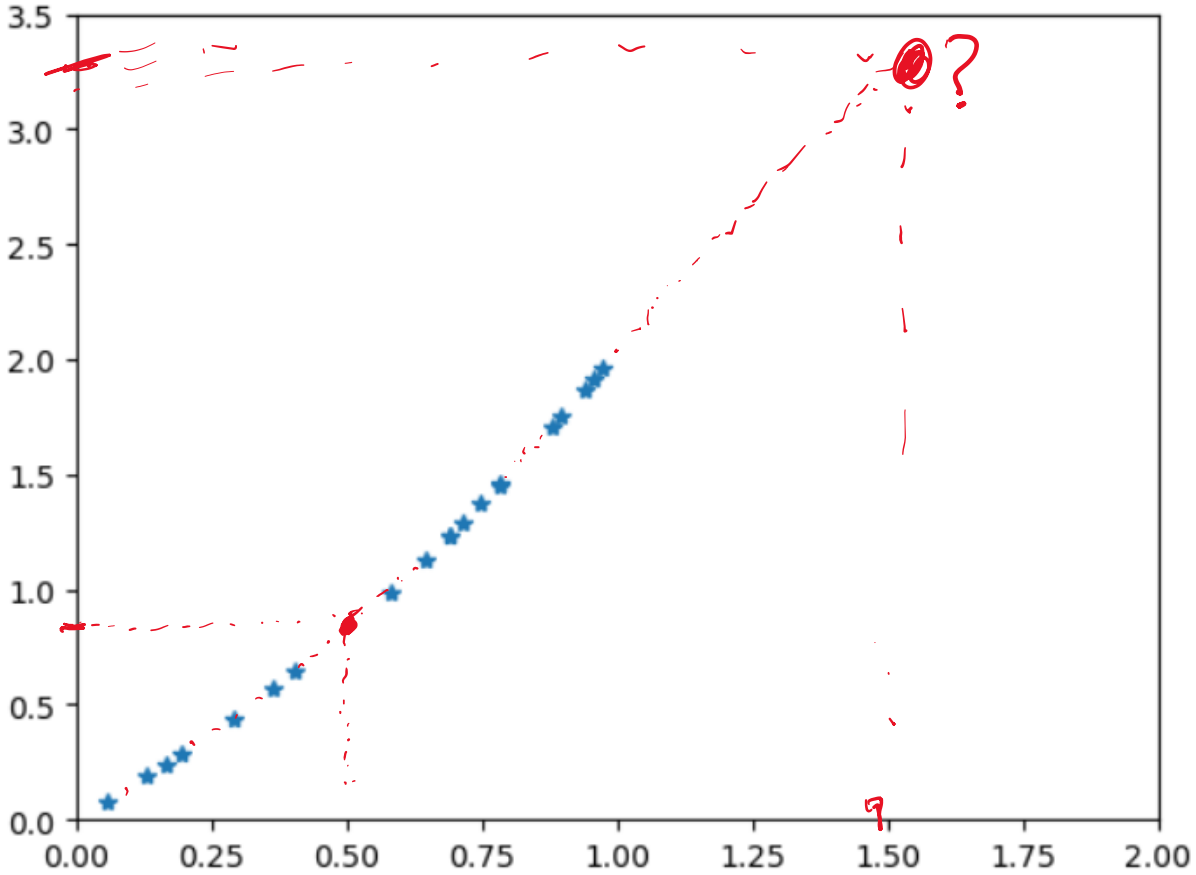
For instance, we could try to make a nonlinear regression to these data points using a quadratic or a cubic polynomial, to find out the plausible value at x=0,5

Interpolation



Here, the assumption is that the curve is not doing anything strange in the range in which I have some missing points.

Extrapolation

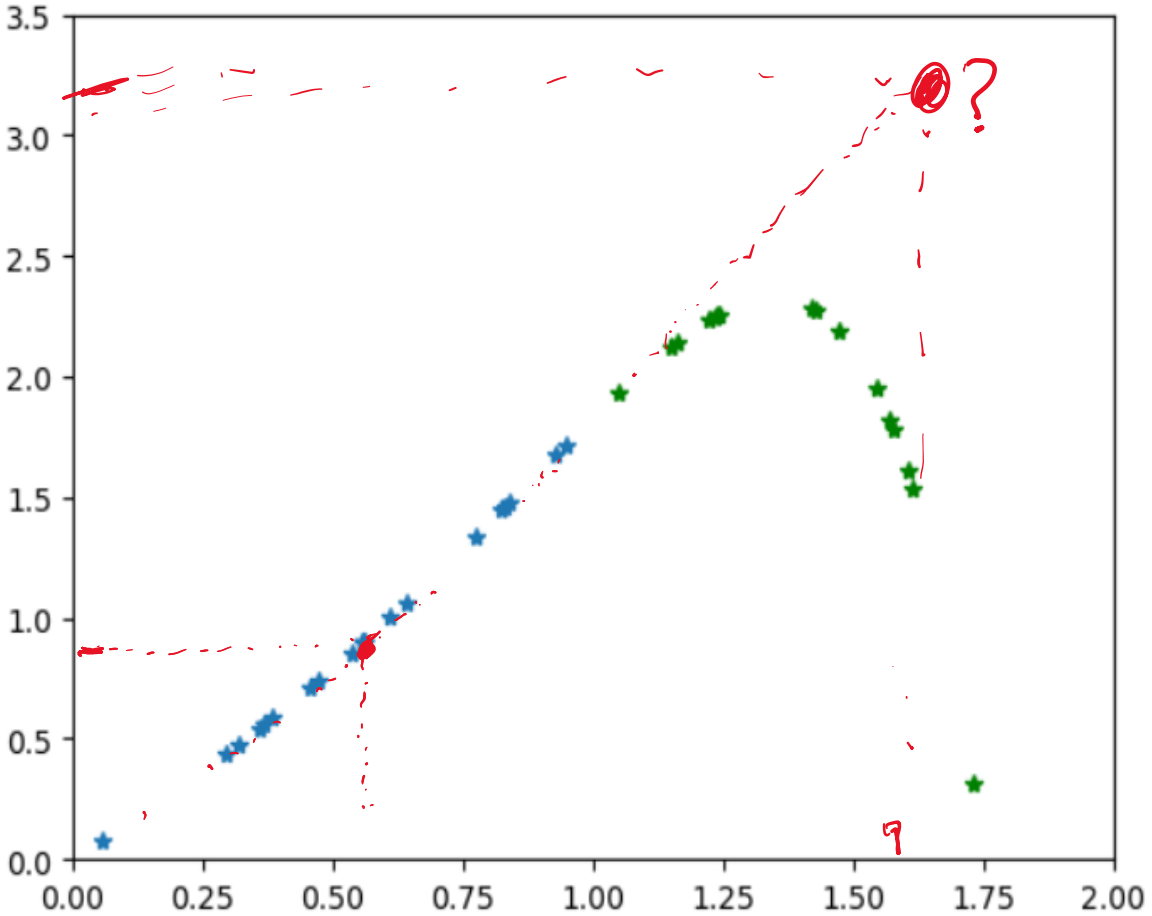


On the other hand, **extrapolation** consists in predicting the value of a data point way beyond the range of the curve. For instance, the point x=1.5

Here we must assume that the behaviour of my system will be the same, even in ranges outside measured data points.

The risk of being wrong is always higher outside the boundaries of my previous data.

Extrapolation



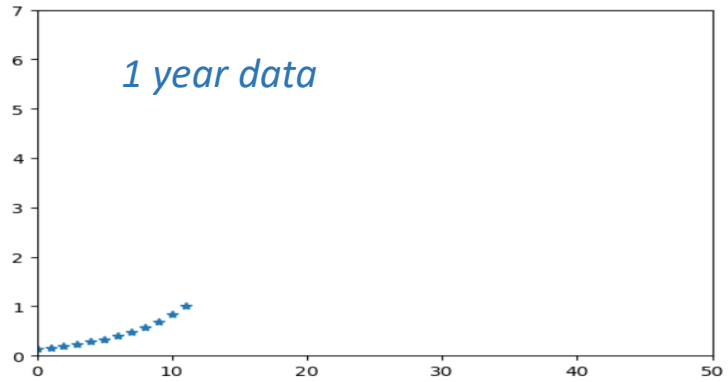
The green points are the real measured data.
Our extrapolation was completely false!

Things to consider

- Interpolation when we have points nearby is quite safe;
- Extrapolation is risky. We must assume that the model we used for regression was correct and will extend well (and this is not always reasonable);
- Predicting future data is always extrapolation;
- When you have many variables, you almost always extrapolate;
- Growth models never extrapolate indefinitely; they always ‘saturate’ at some point.

Example

A company doing social media has 1M followers, and it has been increasing its followers by 20 % every month for the last year. How many followers are they expecting to have in 4 years' time?



Naïve extrapolation will result in an estimate of 6 billion followers !