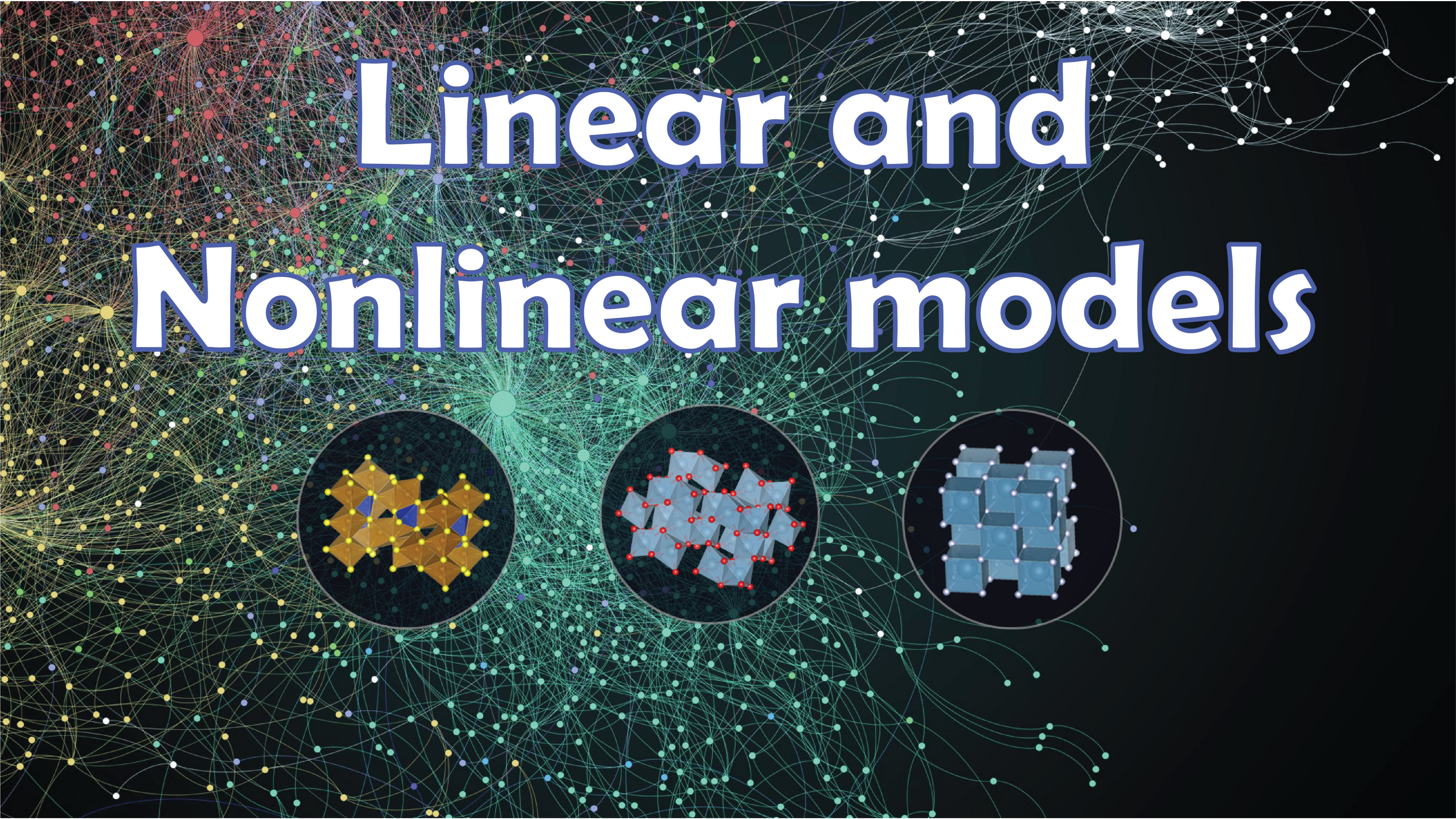
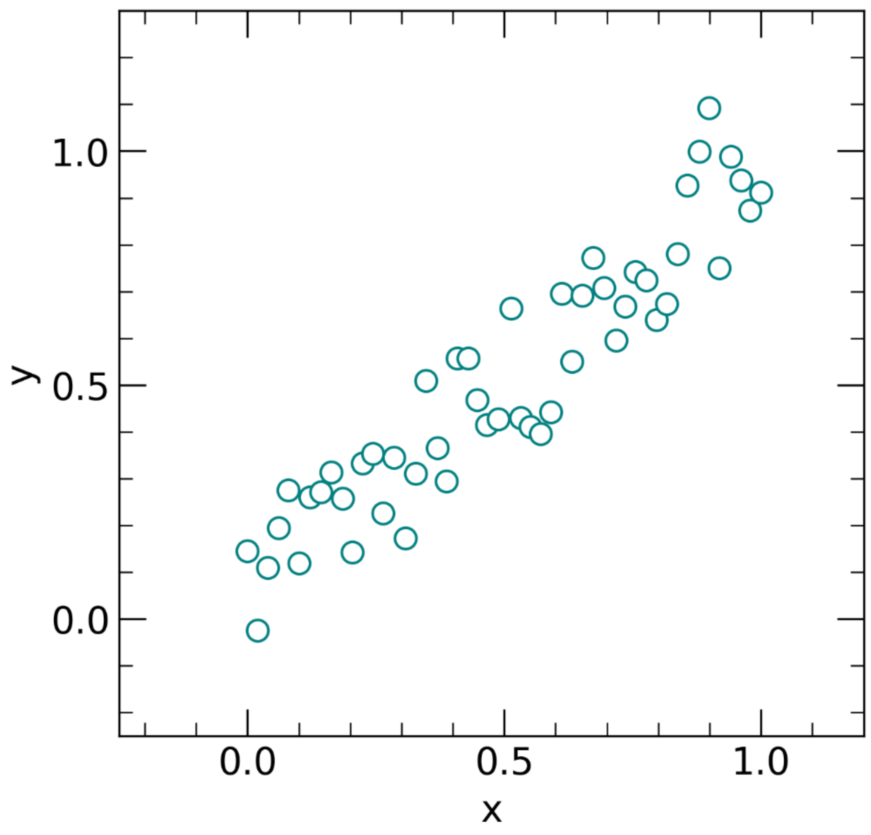
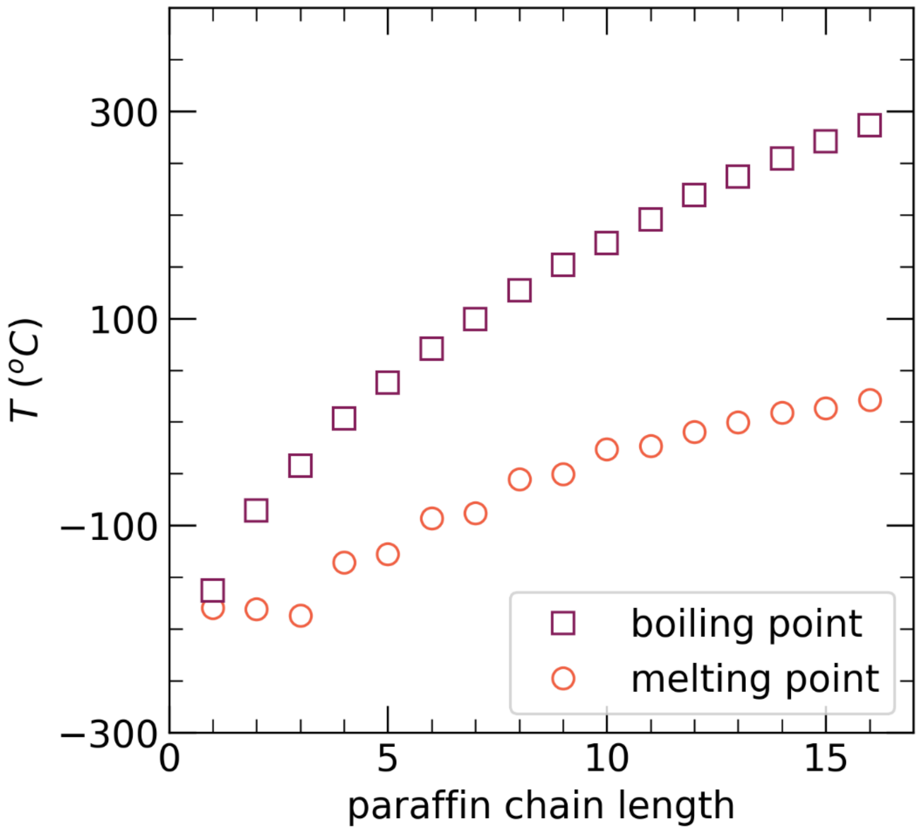

Linear and
Nonlinear models

Both linear and non-linear regression can fit curvature in data





Linear models follow a strict format

$$dependent\ variable = constant + parameter * IV + parameter * IV + \cdots$$

Example taken from https://statisticsbyjim.com/regression/choose-linear-nonlinear-regression/

Linear models follow a strict format

$$dependent\ variable = constant + parameter * IV + parameter * IV + \cdots$$

$$y = mx + b$$

$$\mathrm{x_{final}} = x_{initial} + v_0 t + 1/2at^2$$

Linear models follow a strict format

$$dependent\ variable = constant + parameter * IV + parameter * IV + \cdots$$

$$y = mx + b$$

$$\mathrm{x}_{\mathrm{final}} = x_{initial} + v_0 t + 1/2at^2$$

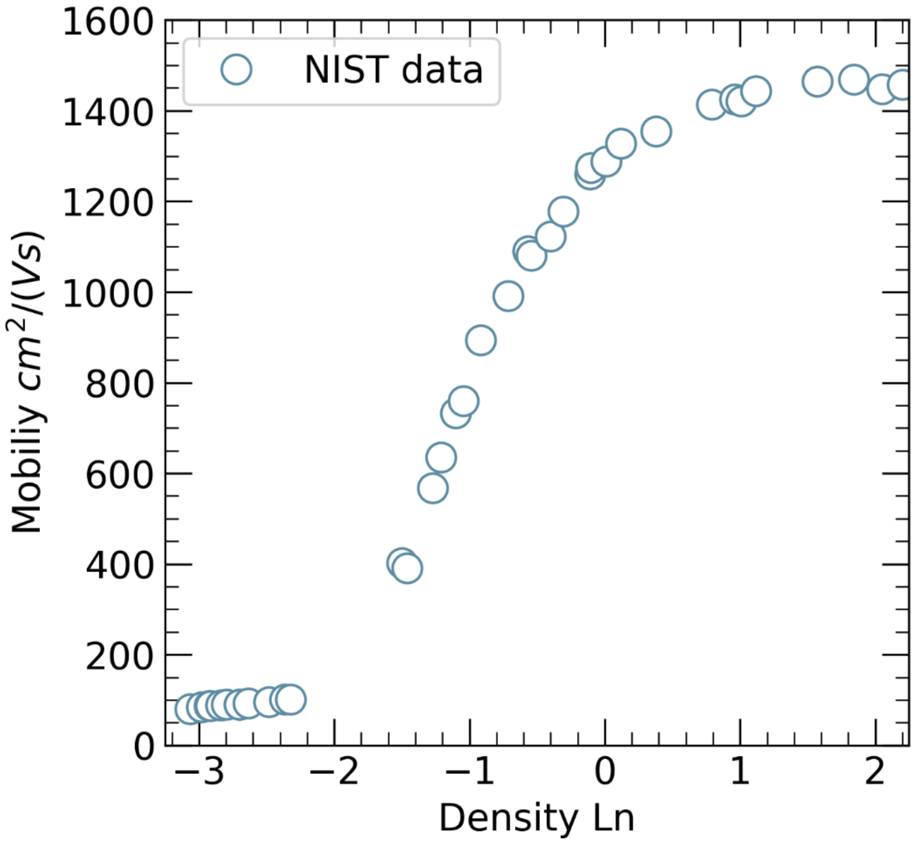
Linear models consist of sums of parameters multiplied by independent variables (IV)

Nonlinear models are anything that don’t follow this form

Linear models follow a strict format

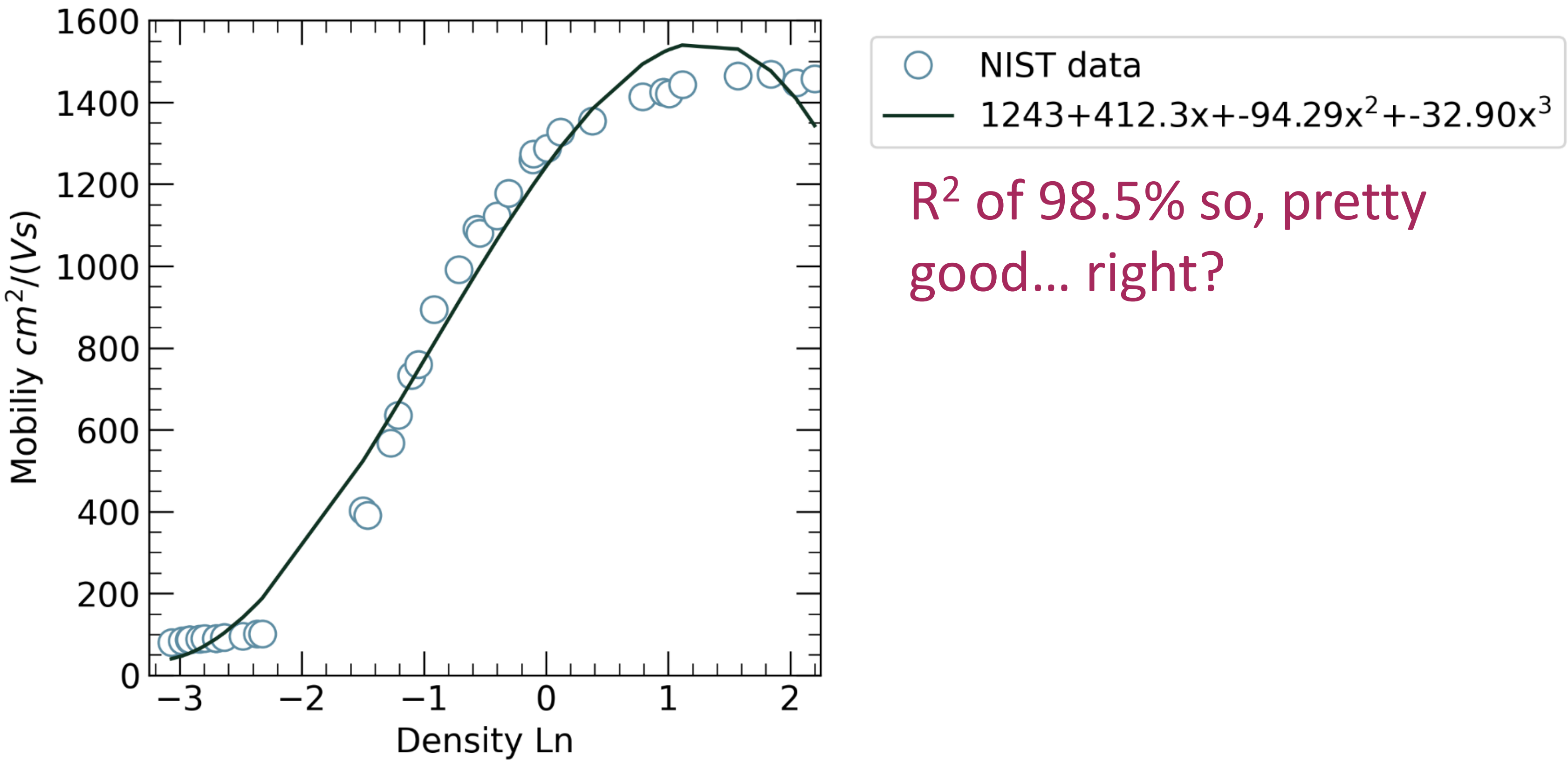
Benefits of linear models	Benefits of nonlinear models
Easy to implement	Harder to implement
Easy to interpret	Harder to interpret
Easily obtained statistics to assess model	Can fit more complicated trends in data (but this can be hard to accomplish)
	No R^2, no p-values possible for the parameter estimates

Let's look at an example from materials data



Data taken from NIST https://www.itl.nist.gov/div898/strd/nls/data/thurber.shtml

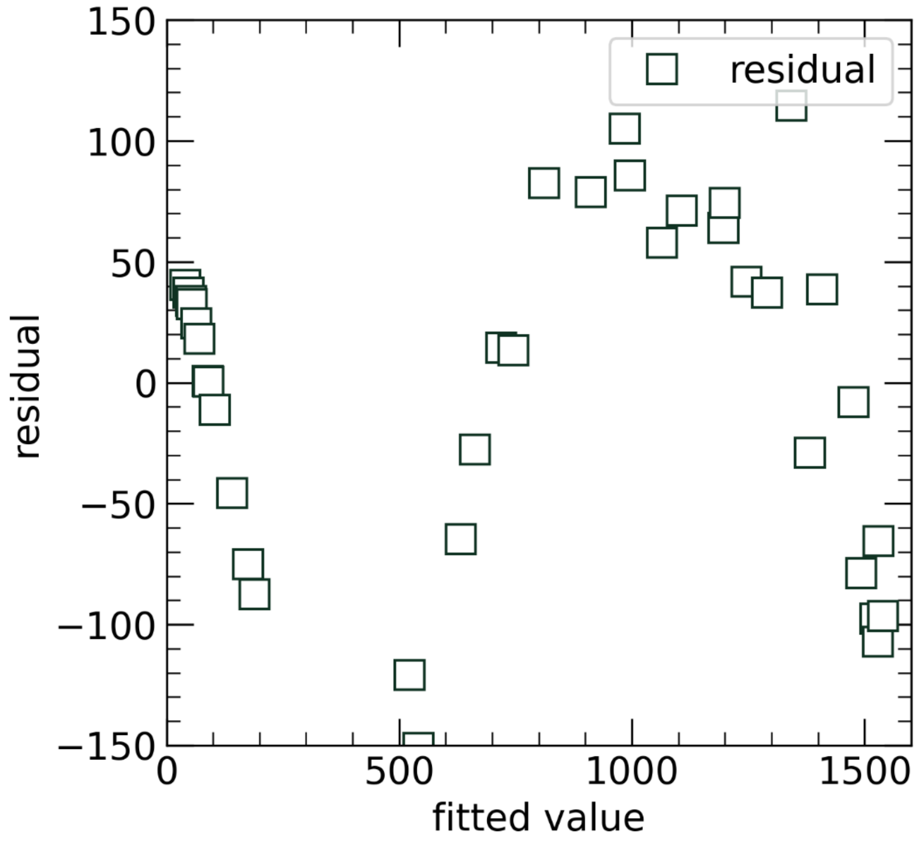
Data can be fit OK with a polynomial order 3 fit



R^2 of 98.5% so, pretty good… right?

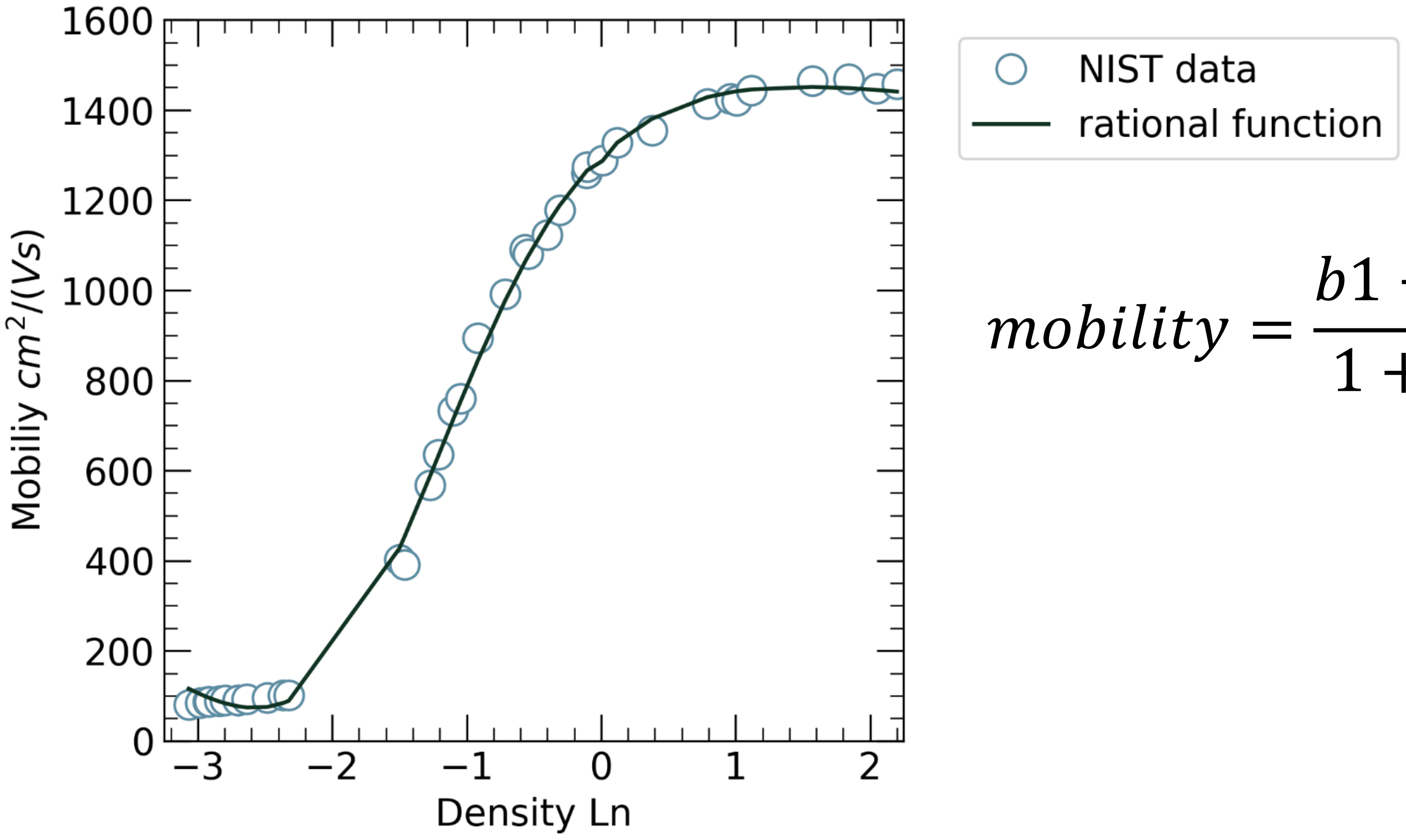
Clearly it overestimates and underestimates portions of the data in systematic (biased) ways

Examine the residual to see the bias



A model that is systematically incorrect is biased (R^2 doesn't tell us the whole story!)

Nonlinear fit is also possible, but more challenging

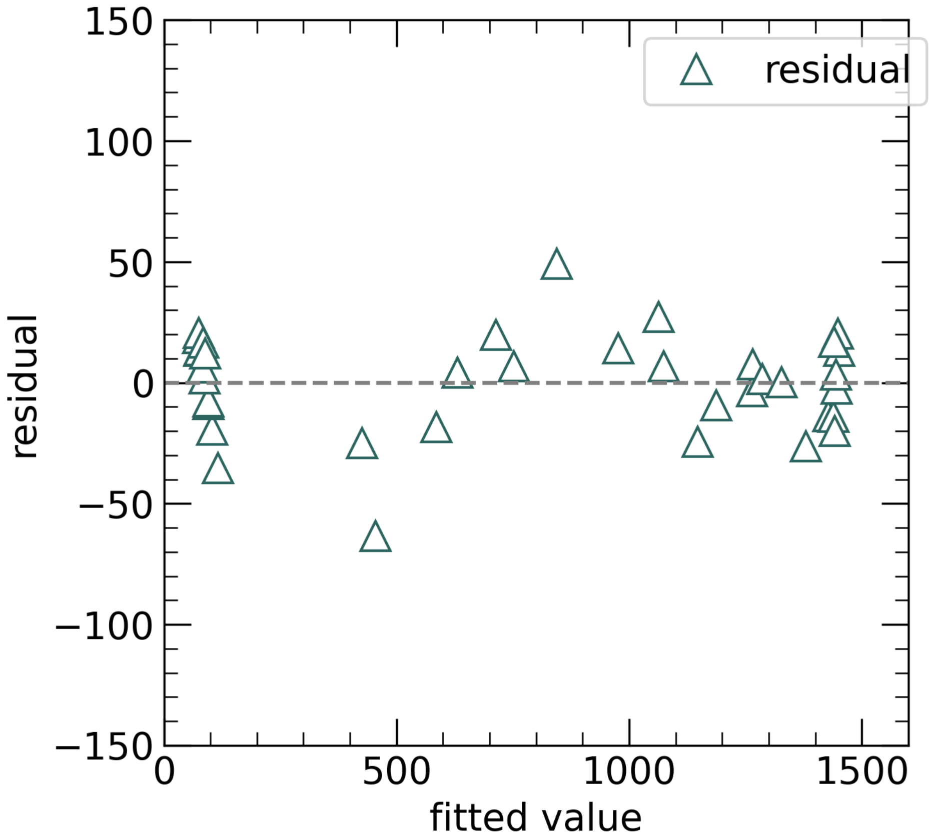


$$mobility = \frac{b1 + b2x + b3x^2 + b4x^3}{1 + b5x + b6x^2 + b7x^3}$$

A rational function is two polynomials divided by one another

Guessing the starting values for the fit is tricky!

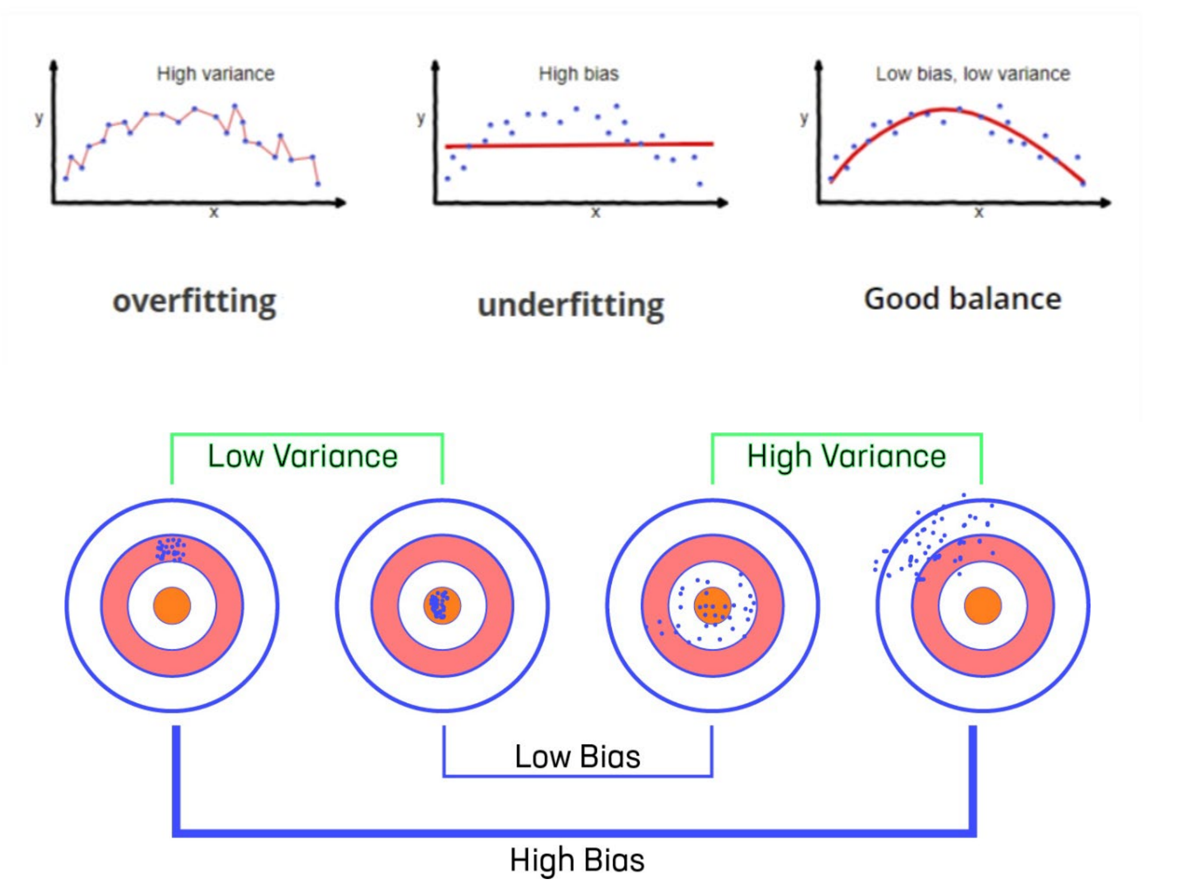
Examine the residual to see the improvement on the bias

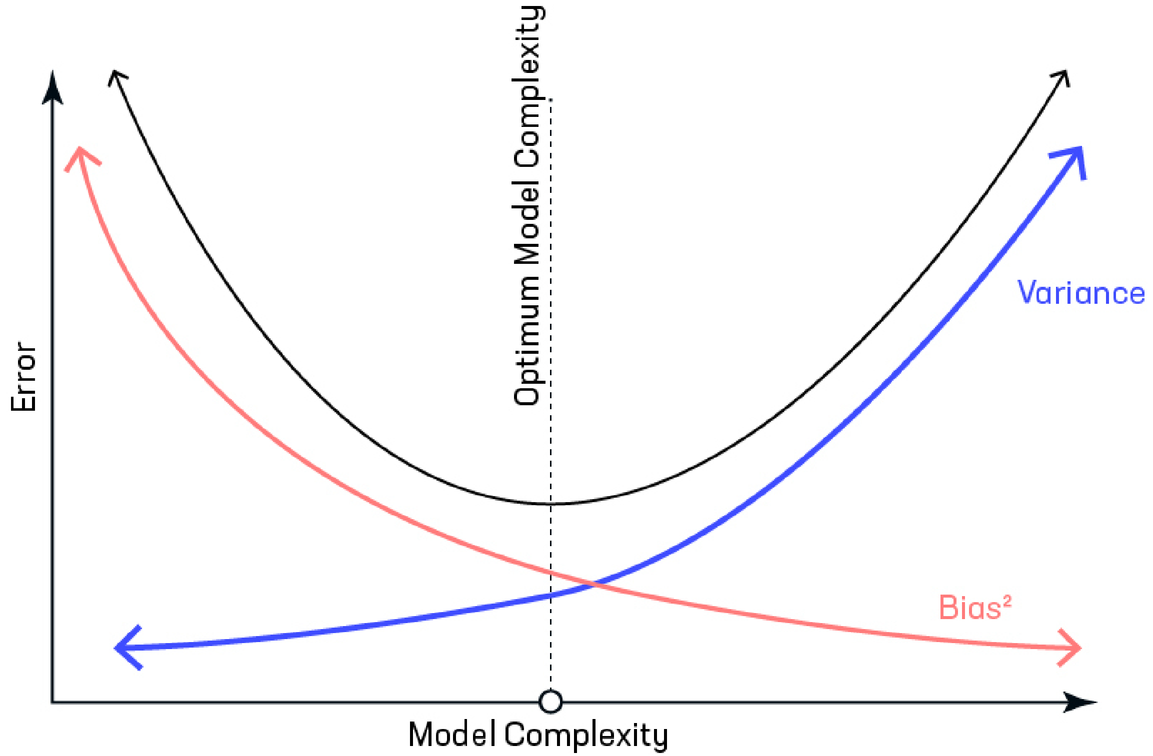


The residual from the rational function is improved and less biased

Curve fit parameters determined by Scipy.optimize curve_fit() method

Regularization can (and should) be used to prevent overfitting





https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/

Ridge, Lasso, and Elastic Net are all good options for regularization

Ridge regression introduces a penalty term based on the square of the coefficients plus a coefficient to control the penalty term. AKA L2 regularization

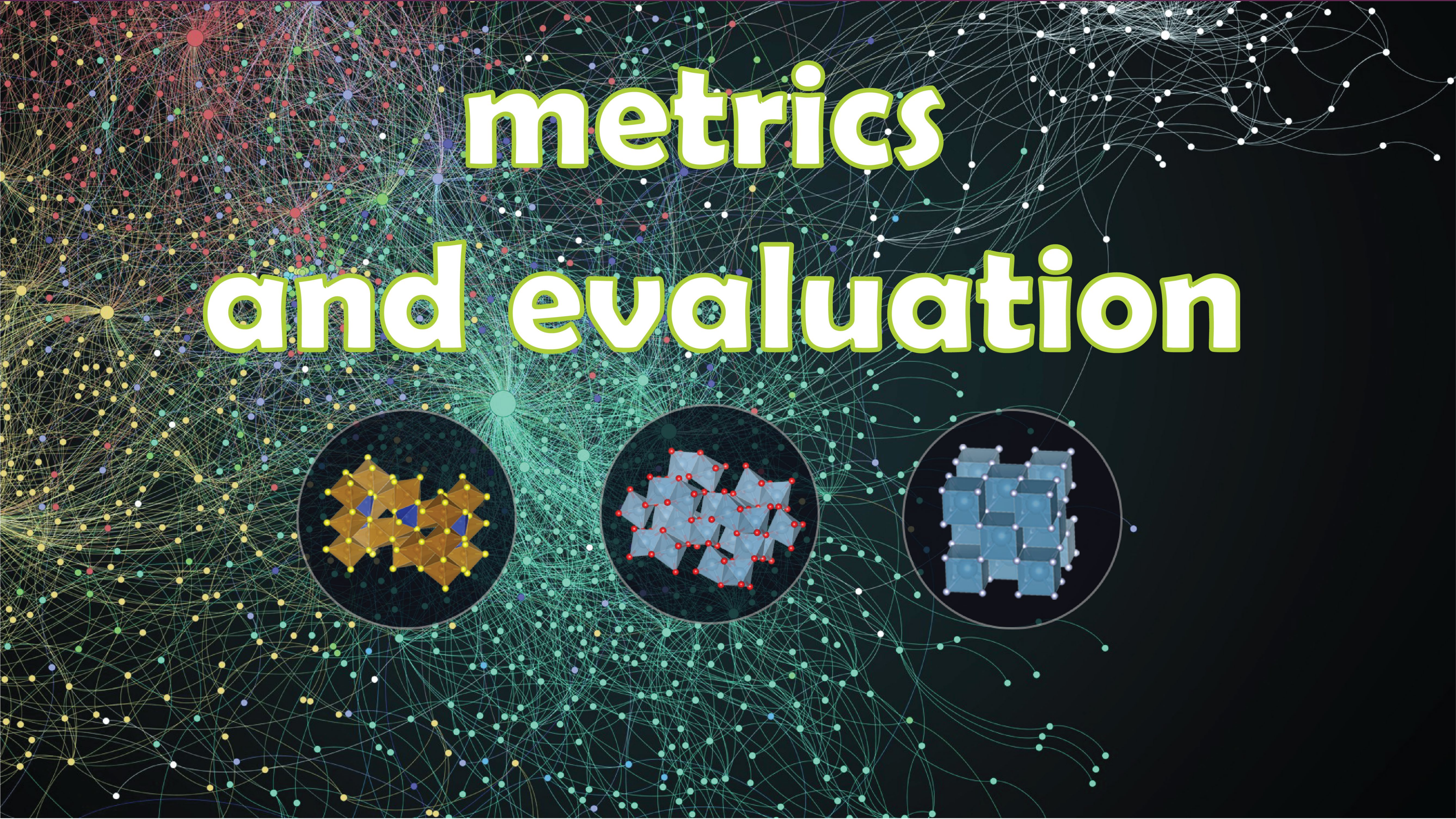
$$L_{ridge} = argmin_{\hat{\beta}}(\|Y - \beta * X\|^2 + \lambda\|\beta\|_2^2)$$

Lasso regression introduces a penalty term based on the sum of the coefficients plus a coefficient to control the penalty term. AKA L1 regularization

$$L_{lasso} = argmin_{\hat{\beta}}(\|Y - \beta * X\|^2 + \lambda\|\beta\|_1)$$

Elastic Net combines both.

https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/


metrics
and evaluation