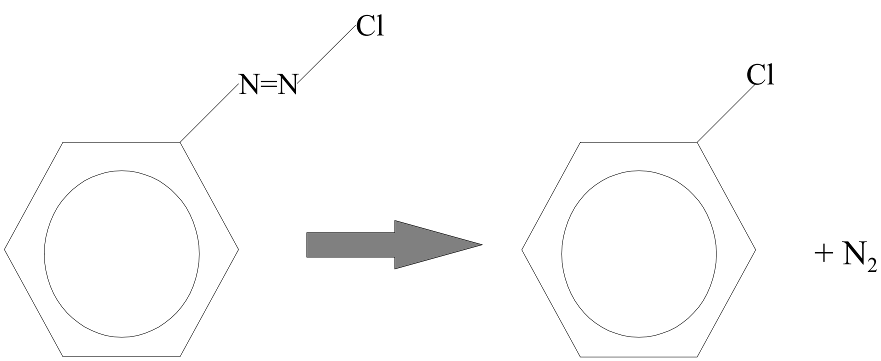
Regression & Curve Fitting

Regression Concepts

What is it and why use it?

Arrhenius Law: $k = Ae^{\frac{E_a}{RT}}$

- E_a: Activation Energy J/mol
- A: Pre-exponential factor (units vary)
- R: Gas constant (8.314 J/molK)
- T: Temperature (K)



Data:

k (s^{-1})	0.00041	0.00101	0.0021	0.0035	0.00717
T (K)	313	319	323	328	333

Here we have 5 equations and 2 unknowns!
How can we solve this???

Generalized Regression

Fitting Data to Arbitrary Functions

Rewrite equations in residual form:

$$k = Ae^{\frac{E_a}{RT}} \Rightarrow r_i = \left(k_i - Ae^{\frac{E_a}{RT_i}} \right)$$

Why r^2 ?

Now we want to *minimize*: $\sum_{i=1}^{n} r_i^2$

Use "*fminsearch*" or "*solver*" to minimize this

Note: This curve may not run through all of the data, but it will be the "best" fit of the given equation to the set of data.

Polynomial Regression

Fitting Data to Polynomials

Revisit the previous example:

$$k = Ae^{\frac{E_a}{RT}} \Rightarrow \ln(k) = \ln(A) + \frac{E_a}{RT}$$

Take natural log of both sides

$$y = mx + b$$

$$m = \frac{E_a}{R}, \quad b = \ln(A)$$

Now we can fit this to a polynomial (order 1)

Polynomial Regression

How to do it in MATLAB

Recall, we can write any polynomial as: $f(x) = \sum_{i=0}^{n} a_i x^i$

- a = polyfit(x, y, n)
  - x is the independent variable data
  - y is the dependent variable data
  - n is the desired order for the polynomial
    - If we have "m" data points, then what is the highest order polynomial that we can fit our data to?
    - coefs are the
  - Coefficients (a) are returned in ascending order
- y = polyval(a, x)
  - Evaluates the polynomial defined by the coefficients "a" at the point(s) given by "x"
  - Coefficients (a) are returned in ascending order