# NUMERICAL METHODS WEEK 3

# CURVE FITTING 1

We continue with Curve Fitting. This week polynomial and multiple linear regression.

Reading: Capra and Canale, introduction to part 5 and chapter 17.

Learning outcomes:

- Extend the work on Linear Regression to polynomial and multiple variables.
- Combine C++ and analytical or other platforms.
- Check your code works correctly, via an external reference.

## MATT WATKINS MWATKINS@LINCOLN.AC.UK

# LEAST SQUARES REGRESSION - SUMMARY

we saw in lecture 1 (https://mattatlincoln.github.io/teaching/numerical_methods/lecture_1/#/) that given our assumption of a straightline

$$y_i = a_0 + a_1 x_i + e_i$$

the error at each point is given by

$$e_i = y_i - a_0 - a_1 x_i$$

We take the sum of the squares of the errors

$$S_r = \sum_{i=0}^{n-1} e_i^2 = \sum_{i=0}^{n-1} (y_i - a_0 - a_1 x_i)^2$$

as our error criterion.

We can find an optimal $a_1$ and $a_0$.

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

and

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} = \bar{y} - a_1 \bar{x}$$

$\bar{y}$ and $\bar{x}$ are the means of the $x$ and $y$ values.

$$\bar{y} = \frac{1}{N}\sum_{i=0}^{N-1} y_i \qquad \bar{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

Correlation of the data: covariance of the data $(x, y)$ divided by the standard deviation of $x$ and $y$.

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

# EXERCISES

Remember in week 1 we looked at finding sums:

- What is $\sum_{n=1}^{100} n$
- What is $\sum_{n=2}^{200} 2n$
- What is $\sum_{n=0}^{99} 2n^2$

Now add an extra array into your code and calculate the following:

$$\text{let } x_i = i, i = 1, 2, \ldots, 10 \text{ and let } y_i = 2i + 0.3 \text{ for } i = 1, 2, \ldots, 10$$

- What is $\sum_i x_i$
- What is $\sum_i y_i$
- What is $\sum_i x_i y_i$
- What are $\bar{y}$ and $\bar{x}$
- Find the parameters $a_0$ and $a_1$ for a linear regression model of this data.

Check your results are correct! I'd suggest using both inspection and Excel.

# POLYNOMIAL LEAST-SQUARES REGRESSION

We can easily extend our method to deal with polynomicals:

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_n x_i^n + e_i$$

and an overall error function

$$S_r = \sum_{i=0}^{N-1} (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \ldots - a_n x_i^n)^2$$

We then take partial derivatives with respect to the parameters ($a_0 \ldots a_n$) to get a set of equations.

Setting these equations equal to zero, writing in matrix form, then solving, gives us the optimal set of parameters.

# FITTING A QUADRATIC FUNCTION

In the case that the largest power of $x$ is $x^2$ we have

$$y_i = a_0 x_i^0 + a_1 x_i^1 + a_2 x_i^2 + e_i$$

and an overall error function

$$S_r = \sum_{i=0}^{N-1} (y_i - a_0 x_i^0 - a_1 x_i^1 - a_2 x_i^2)^2$$

This leads to a set of equations

$$\left(\sum x_i^0\right) a_0 + \left(\sum x_i^1\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum x_i^0 y_i$$
$$\left(\sum x_i^1\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum x_i^1 y_i$$
$$\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum x_i^2 y_i$$

- Derive the above equations by differentiating $S_r$ with respect to each of the $a_i$ in turn.
- Write the equations in matrix form $\mathbf{A}x = b$, where $x$ is a column matrix with entries $a_0, a_1, a_2$ and $b$ is a column matrix with terms that do not depend on the fitting parameters, $a_0, a_1$ and $a_2$.
- using the data you can find on Blackboard for today's session under the assessments tab.
- Solve for $a_0, a_1$ and $a_2$.
- Plot your fitted parabola against the data to check the fit.
- Extend the derivation to fitting a cubic polynomial and fit to the data on Bb.

# MULTIPLE LINEAR REGRESSION

Instead of powers of a single variable, our model for $y_i$ could be that it is a function of several independent variables:

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i} + \ldots + a_n x_{ni}$$

- Derive the a set of equations for the case of two independent variables
- Write the equations in matrix form $Ax = b$, where $x$ is a column matrix with entries $a_0, a_1, a_2$

Using the following data

| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 0 | 0 | 5 |
| 2 | 1 | 10 |
| 2.5 | 2 | 9 |
| 1 | 3 | 0 |
| 4 | 6 | 3 |
| 7 | 2 | 27 |

- Solve for $a_0, a_1$ and $a_2$
- Plot your fitted function against the data to check the fit

# LINEARIZATION OF DATA SETS

Multiple Linear Regression is not just limited to obviously linear data.

- How would you apply multiple linear regression to data that you thought was related by $y = a_0 x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$ ?

# SUMMARY AND FURTHER READING

You should be reading additional material to provide a solid background to what we do in class

Reading: Capra and Canale, introduction to part 5 and chapter 17.

Lots of details in chapters 14 and 15 of Numerical Recipes (http://apps.nrbook.com/c/index.html).