Applied Mathematics 205

Unit I: Data Fitting

Lecturer: Dr. David Knezevic

Unit I: Data Fitting

Chapter I.1: Motivation

Motivation

Data fitting: Construct a continuous function that represents discrete data, fundamental topic in Scientific Computing

We will study two types of data fitting

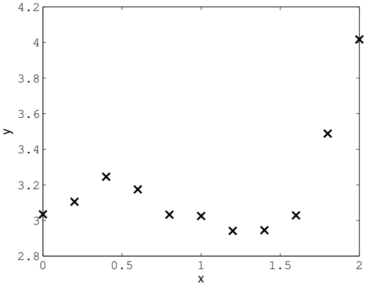
- interpolation: Fit the data points exactly
- least-squares: Minimize error in the fit (useful when there is experimental error, for example)

Data fitting helps us to

- interpret data: deduce hidden parameters, understand trends
- process data: reconstructed function can be differentiated, integrated, etc

Motivation

For example, suppose we are given the following data points



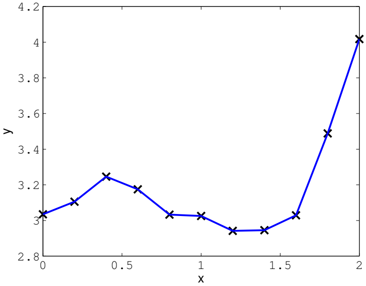
This data could represent

- Time series data (stock price, sales figures)
- Laboratory measurements (pressure, temperature)
- Astronomical observations (star light intensity)
- ...

Motivation

We often need values between the data points

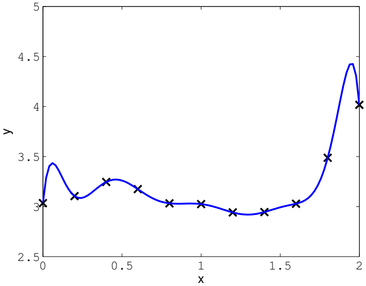
Easiest thing to do: "connect the dots" (piecewise linear interpolation)



Question: What if we want a smoother approximation?

Motivation

We have 11 data points, we can use a degree 10 polynomial



We will discuss how to construct this type of polynomial interpolant in I.2

Motivation

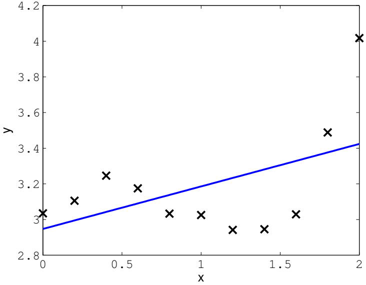
However, degree 10 interpolant is not aesthetically pleasing: too bumpy, doesn't seem to capture the "underlying pattern"

Maybe we can capture the data better with a lower order polynomial...

Motivation

Let's try linear regression (familiar from elementary statistics): minimize the error in a linear approximation of the data

Best linear fit: $y = 2.94 + 0.24x$

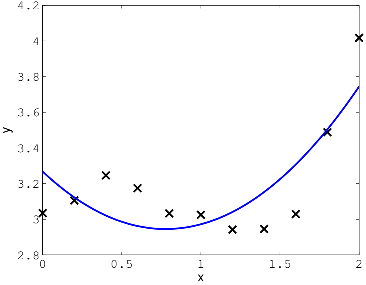


Clearly not a good fit!

Motivation

We can use least-squares fitting to generalize linear regression to higher order polynomials (see I.3)

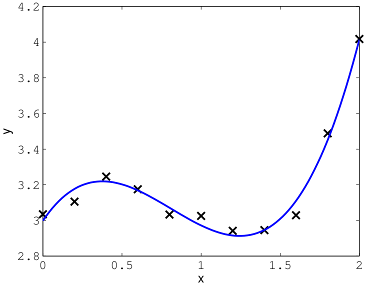
Best quadratic fit: $y = 3.27 - 0.83x + 0.53x^2$



Still not so good...

Motivation

Best cubic fit: $y = 3.00 + 1.31x - 2.27x^2 + 0.93x^3$



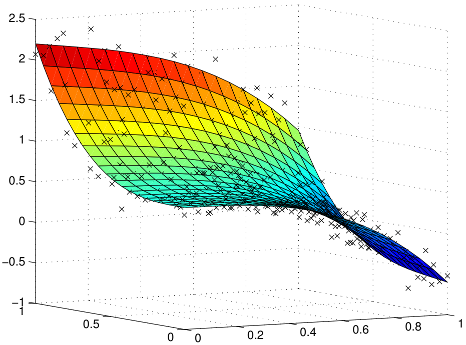
Looks good! A “cubic model” captures this data well

(In real-world problems it can be challenging to find the “right” model for experimental data)

Motivation

Data fitting is often performed with multi-dimensional data (find the best hypersurface in $\mathbb{R}^N$)

2D example:



Motivation: Summary

Interpolation is a fundamental tool in Scientific Computing, provides simple representation of discrete data

- Common to differentiate, integrate, optimize an interpolant

Least squares fitting is typically more useful for experimental data

- Smooths out noise using a lower-dimensional model

These kinds of data-fitting calculations are often performed with huge datasets in practice

- Efficient and stable algorithms are very important