# Lecture 5 - curve fitting

## Class Business

* prayer / spiritual thought
* unit of the day

$$1\,K = 1°C + 273.15$$

Kelvin (no "degree") — Degree Celsius

- Kelvin is an absolute temperature scale

## I. Curve fitting.

* We fit data to a curve when we believe that the data is described by a model plus random noise:

$$y_i = f(x_i) + \varepsilon_i$$

data : $\{x_i, y_i\}$

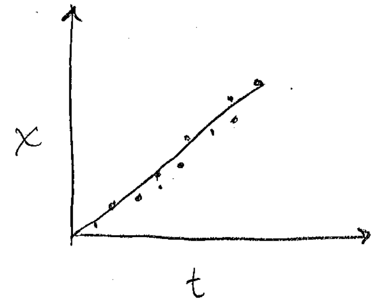
model : $f(x)$

noise : $\{\varepsilon_i\}$

* For example, imagine you are running around a track wearing a fitbit that measures your position and time.

- We can model your position with basic kinematics

$$x = v t$$




data: $\{x_i, t_i\}$

- we want some sort of line, $x = vt$, that "best fits" the data.
- The spread of the data about the model then represents our noise (e.g. measurement error)
- Our model could also be wrong (e.g. here we assumed a constant velocity.)

## II. Linear least-squares

* How do we find this best fit model?
* One procedure is called a "least-squares regression."

In a least-squares regression we minimize the sum of the squared error.

$$\varepsilon_i = y_i - f(x_i)$$

$$SSE = \sum_i (\varepsilon_i)^2 = \sum_i \left[y_i - f(x_i)\right]^2$$

↳ find model parameters that minimize SSE.

example: quadratic polynomial model

$$\left.\begin{aligned} f(x) &= ax^2 + bx + c \\ \varepsilon_i &= y_i - ax_i^2 - bx_i - c \end{aligned}\right\} \quad \text{parameters } a, b, c$$

$$SSE = \sum_i (y_i - a_i x_i^2 - b \cdot x_i - c)^2$$

$$\frac{\partial(SSE)}{\partial a} = 0 = \sum_i 2(y_i - a x_i^2 - b x_i - c)(-x_i^2)$$

$$\frac{\partial(SSE)}{\partial b} = 0 = \sum_i 2(y_i - a x_i^2 - b x_i - c)(-x_i)$$

$$\frac{\partial(SSE)}{\partial c} = 0 = \sum_i 2(y_i - a x_i^2 - b x_i - c)(-1)$$

* 3 linear equations & 3 unknowns

$$\begin{bmatrix} 2\sum_i x_i^4 & 2\sum_i x_i^3 & 2\sum_i x_i^2 \\ 2\sum_i x_i^3 & 2\sum_i x_i^2 & 2\sum_i x_i \\ 2\sum_i x_i^2 & 2\sum_i x_i & 2\sum_i 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i x_i^2 \\ \sum y_i x_i \\ \sum y_i \end{bmatrix}$$

$$A \quad x \quad = \quad b$$

* If the parameters appear linearly in the model, then these equations are linear. If the parameters appear non-linearly, one gets a system of non-linear equations.

* Can solve these equations in Excel (either linear or non-linear)!

* How measure the "goodness of fit"?

We use a value called the coefficient of determination or "R squared"

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

SST: total variance

$$SST = \sum_i (y_i - \bar{y})^2 \quad \bar{y} \leftarrow \text{average}$$

(total spread of the data)

## III. Curve fitting in Excel.

* There are four ways to do a least squares fit in excel.

  (1) Use a scatter plot to make a trend line.

  (2) Use the function LINEST.

  (3) Minimize SSE using solver.

  (4) Directly solve the linear least square equations

  * The first 3 are so much more convenient than #4, that we will skip it.

Activity

* Briefly demonstrate the first 2 techniques.
* Use the provided Excel sheet to practice them in class.

* what if we get a function not "on the list"?

  (a) Re-arrange to make it one on the list

  (b) Use solver.

  (c) Calculate $R^2$ using RSQ : = RSQ(data, model)

Activity

* Briefly demonstrate re-arrangement & solver.
* Use included Excel sheet to practice.