Nonlinear models

I don't think we're in Kansas anymore

Samuel Robinson, Ph.D.

Oct 13, 2023

Outline

- What are nonlinear models?
- Mechanistic models
  - Some common models
  - Strategies for fitting
- Empirical models
  - Some common models
  - GAMs

What are nonlinear models?

- All linear models take the form:

$$\hat{y} = X\beta = b_0 1 + b_1 x_1 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

- However, *nonlinear models* can't be reduced to this linear (matrix) form:

$$\hat{y}_t = \hat{y_{t-1}}(1 + r(1 - \frac{\hat{y_{t-1}}}{k}))$$
$$y \sim Normal(\hat{y}, \sigma)$$

Two common situations

1. “I have governing equations for this system, and I want to fit them to my data”

- e.g. Logistic growth equation, Michaelis-Menten kinematrics, Ricker model

2. “I don’t know what equations represent my system, but I need some kind of *smooth* process that describes them”

- e.g. Changes in organism population over growing season, changes in stock prices over time

Part 1: Mechanistic models

Governing equations

Some systems can be described by a set of governing equations, either in *discrete* or *continuous* time

- Exponential growth: *Discrete time*

$$n_t = n_{t-1} r$$

- Predator prey cycles: *Discrete time*

$$\begin{aligned} \text{prey}_t =& \text{prey}_{t-1}(r_1 - a_1 \text{pred}_{t-1}) \\ \text{pred}_t =& \text{pred}_{t-1}(a_2 \text{prey}_{t-1} - d) \end{aligned}$$

- Exponential growth: *Continuous time*

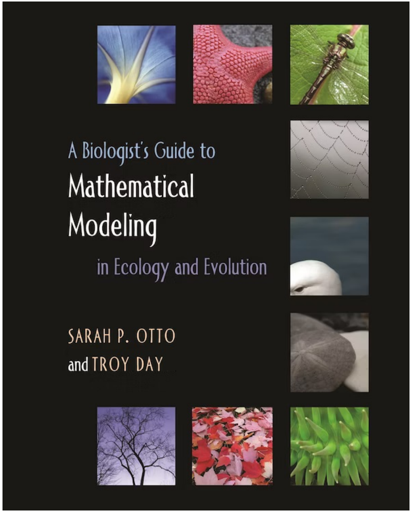
$$\frac{dn}{dt} = nr$$

- Predator prey cycles: *Continuous time*

$$\begin{aligned} \frac{d\text{prey}}{dt} =& r - a_1 \text{pred} \\ \frac{d\text{pred}}{dt} =& a_2 \text{prey} - d \end{aligned}$$

Where do these equations come from?

- Mostly from literature, or sometimes from your own derivations
- Can be derived from causal models, flow diagrams, organismal life cycles
- Math-heavy topic for another class! If you're interested, I might start with this book:

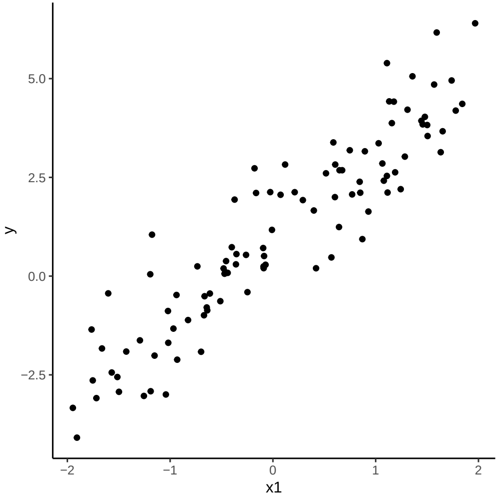


Fitting mechanistic models

- We have a *pretty good idea* what rules the system is following, and we want to figure out the parameters that it uses
- Let's start with a simple linear model, where we have *2 parameters* b_0 and b_1 that we're looking for

$$\hat{y} = X\beta = b_0 + b_1 x_1$$

- We're trying to find the parameters of a line that *most closely* fits our data:

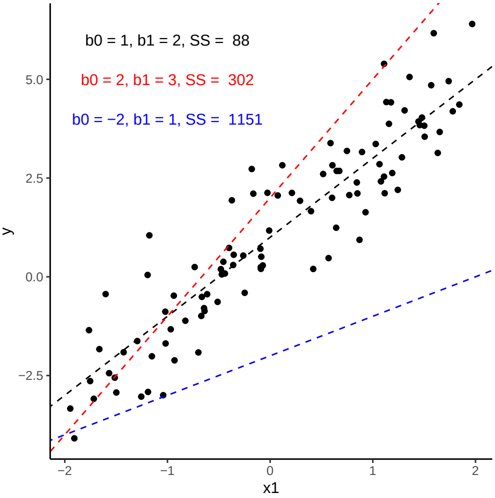


Fitting mechanistic models (cont.)

- How might we define "closest fit" in a mathematical sense?
- One common measure is *sum of squared distances*. This is just the difference between the data and the line:

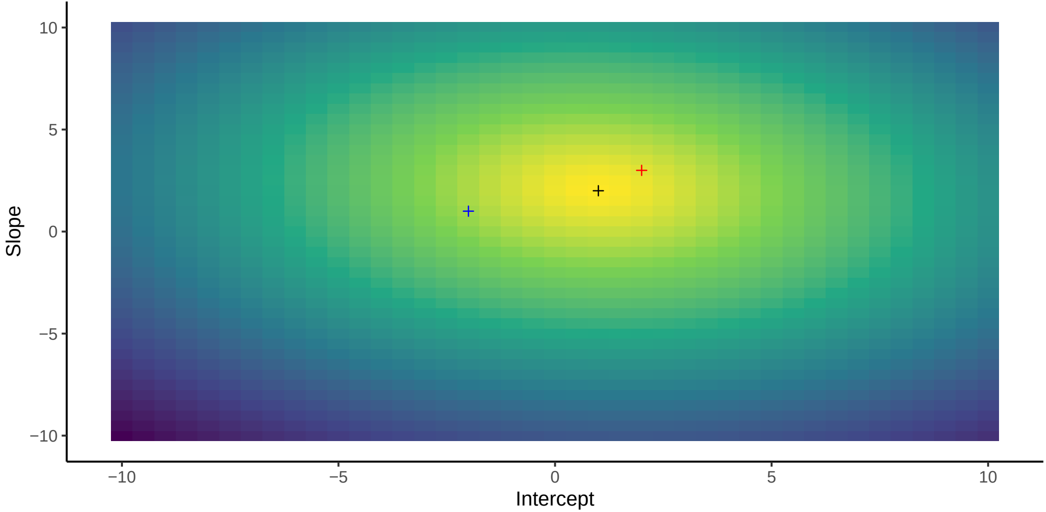
$$S = \sum_{i=1}^{N} (y_i - (b_0 + b_1 x_i))^2$$

- Here are three "guesses" at the slope and intercept, along with their SS scores. Which one looks to be the best?



Map of fitting surface

We can try this for a whole bunch of intercepts and slopes:



Getting R to do this

- It's pretty clear where the best intercept and slope is, but how do we get R to do this?
- First, we need a function that returns SS given a set of parameters:

```
#Function to calculate SS from b0, b1, and data (x & y)
ssFun <- function(B,xdat,ydat){
  sum((ydat - (B[1] + B[2]*xdat))^2)
}
```

- Next, we use the `optim` function to find the intercept and slope values that return the minimum value of SS. How did it do? (Actual values: b_0:1, b_1:2)

```
#Starts at 0,0 and "looks around" from there
optim(par = c(0,0) , fn = ssFun)
```

```
## $par
## [1] 0.9769795 2.0779779
##
## $value
## [1] 86.78819
##
## $counts
## function gradient
##       67       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

Part 2: Empirical models

Empirical smoothing

2-column slide

a

b