

02-712 Week 12

Biological Modeling and Simulation

Aidan Jan

November 13, 2025

Regression

Suppose we want to fit a straight line to our data (represented by points on the cartesian plane). We can do that by posing this as a gradient descent problem.

- Write the equation for our line (it's probably $y = mx + b$)
- We can calculate error by subtracting the inferred y -value from the line from the true y of the data point.
- Sum the squares of the error. We do squares to prevent negative errors from cancelling positive errors.

$$SE = (y_1 - (mx_1 + b))^2 + (y_m - (mx_m + b))^2$$

- From here, our goal is to find (m, b) such that it minimizes error. To do this, we take the derivative of the squared-error equation in terms of m and b .
- By expanding, then taking the derivative, we get the system

$$\begin{aligned}\frac{\partial SE}{\partial m} &= 0 \implies m\bar{x}^2 + b\bar{x} = \bar{x}\bar{y} \\ \frac{\partial SE}{\partial b} &= 0 \implies m\bar{x} + b = \bar{y}\end{aligned}$$

- What can we interpret from these equations? (the 'bar' symbol represents average.)
- The second equation tells us that the average point of all the points must be on the best fit line.
- The first equation is the same, but less intuitive, which is the point $\left(\frac{\bar{x}^2}{\bar{x}}, \frac{\bar{x}\bar{y}}{\bar{x}}\right)$ must be on the line.

Example: Maximum Likelihood Estimation (MLE)

Maximum Likelihood is a probabilistic method that determines values for the parameters of the model. The parameter values are found such that they maximize the likelihood that the process described by the model produces the observed data.

Suppose we have observations that are sampled (for this example, let's assume from a Gaussian). Therefore,

$$P(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Let our data points be 9, 9.5, 11. In this case, we want to maximize

$$P(\text{data points}, \mu, \sigma) = P_1 \cdot P_2 \cdot P_3$$

The goal is to find (μ, σ) that maximizes the probability above.

Example: Allele Densities

Suppose we have a population of animals with two different alleles of some trait, A and a . Let the population have N individuals. The probability that a individual is a mutant (allele = A) is p . How do we find p ?

- Well...we can always count the number of A 's in our sample and just use that as a guess for the population. But can we do better?

We can pose this as a maximum likelihood estimation problem. First, define a dummy variable.

$$x_i = \begin{cases} 1 & A \\ 0 & a \end{cases}$$

Now, let's define a function, f , that tells us the probability of getting our observations (under our observations.)

$$f(x_i|p) = p^{x_i} \cdot (1-p)^{1-x_i}$$

We use the binomial form here, since each variable follows a Bernoulli distribution (either mutant or not). This equation works for one individual, so now we can expand it to all individuals.

$$f(x_1, \dots, x_n|p) = p^{x_1}(1-p)^{1-x_1} \dots p^{x_n}(1-p)^{1-x_n}$$

This can be written as a product:

$$= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}$$

Now, we want to find $\frac{\partial L}{\partial p} = 0$ since we want to maximize the likelihood. It turns out that this function sucks to take the derivative of, so instead, we take the log of the function.

$$\ell = \log \mathcal{L} = \log\left(\prod_{i=1}^N p^{x_i}(1-p)^{(1-x_i)}\right)$$

This now converts into a sum of logs, and with some algebra to simplify, we will eventually get to

$$= N\bar{x} \log(p) + N(1-\bar{x}) \log(1-p)$$

We can now take the derivative:

$$\frac{\partial \log(\mathcal{L})}{\partial p} = \frac{\partial \ell}{\partial p} = \frac{N\bar{x}}{\hat{p}} - \frac{N(1-\bar{x})}{1-\hat{p}} = 0$$

Now, we can simplify:

$$\begin{aligned} 0 &= \frac{N\bar{x}}{\hat{p}} - \frac{N(1-\bar{x})}{1-\hat{p}} \\ \frac{N(1-\bar{x})}{1-\hat{p}} &= \frac{N\bar{x}}{\hat{p}} \\ \hat{p}(1-\bar{x}) &= \bar{x}(1-\hat{p}) \\ \hat{p} &= \bar{x} \end{aligned}$$

It turns out that our intuition at the beginning was correct.

Least Squares as Maximum Likelihood

$$y_i = \beta x_i + \varepsilon_i$$

We let (x_i, y_i) be our data point and ε_i be our error term, where $\varepsilon_i \sim N(0, \sigma^2)$. We can solve this equation in terms of our error, which gives:

$$\varepsilon_i = y_i - \beta x_i \sim N(0, \sigma^2)$$

Therefore,

$$p(x_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\beta x_i)^2}{2\sigma^2}}$$

The likelihood for points in our sample is

$$\mathcal{L} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\beta x_i)^2}{2\sigma^2}}$$

To solve this, we use the same method as before, taking the log of this, then the derivative.

Expectation Maximization

Expectation Maximization (EM) algorithms are parameter estimators in probabilistic models with **incomplete data**.

Suppose we have two unfair coins A and B , with bias θ_A and θ_B . Our goal is to be able to estimate θ_A and θ_B based off of some data. Consider the following 50 data points

1. HTTTHHTHTH
2. HHHHTHHHHH
3. HTHHHHTHH
4. HTHTTTHTT
5. THHHTHHHTH

Let $x = (x_1, \dots, x_5)$, where $x_i \in \{0, \dots, 10\}$, or the number of heads during the i -th toss. Let $Z = (z_1, \dots, z_5)$, during the i -th toss, $z_i \in \{A, B\}$, or the identity of the coins we toss at the i -th toss. The hard part is that we don't know which toss belongs to which coin, so for now, we make a guess. We can make a data table to keep track:

Toss	Coin A	Coin B
HTTTHHTHTH		5H, 5T
HHHHTHHHHH	9H, 1T	
HTHHHHHTHH	8H, 2T	
HTHTTTHTT		4H, 6T
THHHTHHHTH	7H, 3T	

Once we have this data, we can guess that coin A made 24H, 6T, and coin B made 9H, 11T. Once we have this, we can plug it into the maximum likelihood equation:

$$\max_{\theta} \log p(x, z, \theta) = \hat{\theta}$$

So what if we don't have an assumption on which toss belongs to which coin? We can start with our guess, and refine it.

The EM Algorithm

The EM Algorithm consists of two parts:

- The **E step** is the step of guessing the probability distribution over completions of the missing data given the current model
- The **M step** is the step of re-estimating model parameters using the assumptions

EM: Coin Flip Example

Suppose we start with $A \rightarrow \theta_A = 0.4$ and $B \rightarrow \theta_B = 0.7$. Now, suppose we have the first toss, with $E : (8H, 2T)$. In this case,

$$P(E|z_A) = P(HHHHHHHHTT|A) = \binom{10}{8} (0.4)^8 (0.6)^2$$

$$P(E|z_B) = P(HHHHHHHHTT|B) = \binom{10}{8} (0.7)^8 (0.3)^2$$

By solving these and using Bayes' Theorem, we can calculate $P(z_A|E)$ and $P(z_B|E)$. The formula for it essentially boils down to:

$$p(z_A|E) = \frac{\theta_A^h (1 - \theta_A)^t}{\theta_A^h (1 - \theta_A)^t + \theta_B^h (1 - \theta_B)^t}$$

$$p(z_B|E) = \frac{\theta_B^h (1 - \theta_B)^t}{\theta_A^h (1 - \theta_A)^t + \theta_B^h (1 - \theta_B)^t}$$

We can get

$$p(z_A|E) = 0.0435$$

$$p(z_B|E) = 0.0956$$

EM: Coin Flip Example (part 2)

Now, let's go back to the original problem with the five tosses. We can make a starting that $\hat{\theta}_A(0) = 0.6, \hat{\theta}_B(0) = 0.5$. For our E-step, we plug it into the formula for $P(z_A|E)$ and $P(z_B|E)$ we derived above. We get the first section of the following table. Now, additionally, we can multiply the probabilities for each coin with the number of heads and tails to generate weights.

Toss	Coin A	Coin B	Coin A	Coin B
HTTTHHTH	0.45	0.55	2.2H, 2.2T	2.8H, 2.8T
HHHHTHHHH	0.8	0.2	7.2H, 0.8T	1.8H, 0.2T
HTHHHHHTHH	0.73	0.27	5.9H, 1.5T	2.1H, 0.5T
HTHTTTTHHTT	0.35	0.65	1.4H, 2.1T	2.6H, 3.9T
THHHHTHHHTH	0.65	0.35	4.5H, 1.9T	2.5T, 1.1T

From here, we can total the coin A and coin B weights (21.2H, 8.5T) for coin A, (11.8H, 8.5T) for coin B. Finally, we can calculate the new θ values. (This is the M-step.)

$$\theta_A(1) = \frac{21.2}{21.2 + 8.5} = 0.7138$$

$$\theta_B(1) = \frac{11.8}{11.8 + 8.5} = 0.5812$$

Now, we can repeat the process again, for as many iterations as needed. Eventually, the θ values will converge to a number.