The EM (Expectation Maximization) Algorithm

Very general algorithm. Uses for:

- Learning Hidden Markov Models (Baum-Welsh)
- Maximum likelihood estimation
- Learning from incomplete examples

Useful for "chicken and egg" problems.

Warm-up

Want to find a solution of

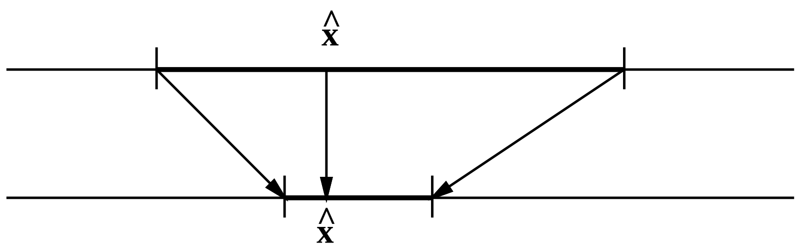
$$x^5 - 3x^2 + 2x - 17 = 0.$$

Idea: rewrite as

$$\begin{array}{rcl} x & = & \left(3x^2 - 2x + 17\right)^{1/5} \\ 1.7826 & \leftarrow & 1 \\ 1.8716 & \leftarrow & 1.7826 \\ 1.8845 & \leftarrow & 1.8716 \\ 1.8864 & \leftarrow & 1.8845 \\ 1.8866 & \leftarrow & 1.8864 \\ 1.8867 & \leftarrow & 1.8866 \end{array}$$

Why does this work?

Contraction mapping with solution as fixed point.



Waiting for a bus

- Suppose there's no schedule.
- Every second, 1/600 chance of bus coming.

(Waiting time given by exponential density $\frac{1}{\mu}e^{-t/\mu}$ with mean $\mu = 10$ minutes.)

Say we don't know μ. Want to estimate it from data. Sit and wait:

Day 1: 7 min

Day 2: 12 min

Day 3: 15 min

Day 4: 10 min

Maximum likelihood $\mu = \frac{7+12+15+10}{4} = 11$ min.

What if data looks like this...

Day 1: waited 7 min and caught bus.

Day 2: waited 12 min and gave up.

Day 3: waited 8 min and caught bus.

Day 4: waited 5 min and gave up.

Idea:

1. Start with some guess of μ. Say 8 min.

2. Fill in missing data with expected value as if guess were correct:

   $\Rightarrow$ just add μ to values for days 2&4

   New data: $(7, 20, 8, 13)$

3. Calculate ML $u = \frac{7+20+8+13}{4} = 12$.

   Go to 2.

Or, could calculate exactly:

$$\begin{aligned} [7 + (12 + \mu) + 8 + (5 + \mu)]/4 &= \mu \\ 32 + 2\mu &= 4\mu \\ \mu &= 16 \end{aligned}$$

More generally...

Start somewhere (guess missing info or guess hypothesis), then repeat:

E. Calculate expected value of missing info given h.

M. Calculate maximum likelihood h given our guess.

(More generally, in **E** step, calculate distribution of missing info, and in **M** step calculate h that maximizes expected (log) likelihood over this distribution.)

What do we need for this to converge?

Contraction + fixed point.