The Baum-Welch Algorithm

Given a set of sequences $X = x^1, \ldots, x^n$ generated by a hidden Markov model with fixed structure, our goal is to estimate the emission probabilities for each state as well as the transition probabilities between the states. To solve this problem, we employ an expectation-maximization algorithm known as the **Baum-Welch algorithm**.

Like all expectation-maximization strategies, we first define the observed data, hidden data, and parameters for the model:

- Observed data: $X = \{x^1, \ldots, x^n\}$ the set of observed sequences
- Hidden data: $Z = \{\pi^1, \ldots \pi^n\}$ the paths through the HMM that generated X
- Parameters: θ = the set of emission and transition probabilities

The idea is that the algorithm alternates between calculating the expected number of times each transition and emission is used to generate X. These expected counts are used to re-estimate the emission and transition probabilities (i.e. the parameters). These parameters are then used again for computing the expected counts of the emissions and transitions. The algorithm then converges on a local maxima of the likelihood surface for the observed sequences X.

Description

Since the Baum-Welch algorithm is an implementation of the EM algorithm, we break it down into the E (expectation) and M (maximization) steps:

E-Step

For each sequence, we run the forward and backward algorithms on x^j. This gives all forward values $f_k^j(i)$ and backward values $b_k^j(i)$.

We then use these values to calculate the expected number of times each transition a_{kl} was used for each sequence. To do this, we first note that the probability that a_{kl} was used at position i in sequence x is calculated as:

$$P(\pi_i = k, \pi_{i+1} = l, x \mid \theta) = P(\pi_i = k, \pi_{i+1} = l, \mid x, \theta)\; P(x)$$

$$= f_k(i)\; a_{kl}\; e_l(x_{i+1})\; b_l(i+1)$$

Therefore,

$$P(\pi_i = k, \pi_{i+1} = l, x \mid \theta) = \frac{f_k(i)\ a_{kl}\ e_l(x_{i+1}\ b_l(i+1)}{P(x)}$$

$$= \frac{f_k(i)\ a_{kl}\ e_l(x_{i+1})\ b_l(i+1)}{f_N(L)}$$