METL – TP1

Alexandre Kabbach
alexandre.kabbach@unige.ch

23.02.2017

Deadline: 09.03.2017

1 Exercise 1 – 10 points

Translate the following sentences from Arcturan to Centauri:[1]

1. iat lat pippat eneat hilat oloat at-yurp
2. totat nnat forat arrat mat bat
3. wat dat quat cat uskrat at-drubel

To help you out, we provide a sample Arcturan document translated into Centauri, and a sample of monolingual Centauri text (see Table 1).

You can use any approach for translation, including your intuition, analytic or automatic calculations. Explain in two or three sentences the procedure you devised.

2 Exercise 2 – 10 points

For the two coins example detailed in Appendix, implement a program which calculates the MLE estimation of coin parameters using the EM algorithm. The input can be fixed to the 5 sets of observations, each of 10 tosses, as in the example. The output should be the computed estimations of parameters of the two coins, that is, their probabilities to land on the head. The details of the two coins example and of the EM algorithm can be found in the supplementary material in the folder of this TP.

The exercise will be evaluated based on the correctness of the implementation of the EM algorithm (particularly, its convergence), as well as the clarity of the code.

[1]This exercise was created by Professor Kevin Knight.

Translated document	**Monolingual Centauri text**
1. **C:** ok-voon ororok sprok **A:** at-voon bichat dat 2. **C:** ok-drubel ok-voon anok plok sprok **A:** at-drubel at-voon pippat rrat dat 3. **C:** erok sprok izok hihok ghirok **A:** totat dat arrat vat hilat 4. **C:** ok-voon anok drok brok jok **A:** at-voon krat pippat sat lat 5. **C:** wiwok farok izok stok **A:** totat jjat quat cat 6. **C:** lalok sprok izok jok stok **A:** wat dat krat quat cat 7. **C:** lalok farok ororok lalok sprok izok enemok **A:** wat jjat bichat wat dat vat eneat 8. **C:** lalok brok anok plok nok **A:** iat lat pippat rrat nnat 9. **C:** wiwok nok izok kantok ok-yurp **A:** totat nnat quat oloat at-yurp 10. **C:** lalok mok nok yorok ghirok clok **A:** wat nnat gat mat bat hilat 11. **C:** lalok nok crrrok hihok yorok zanzanok **A:** wat nnat arrat mat zanzanat 12. **C:** lalok rarok nok izok hihok mok **A:** wat nnat forat arrat vat ga	1. ok-drubel anok ghirok farok 2. wiwok rarok nok zerok ghirok enemok 3. ok-drubel ziplok stok vok erok enemok kantok ok-yurp 4. zinok jok yorok clok 5. lalok clok izok vok ok-drubel 6. ok-voon ororok sprok 7. ok-drubel ok-voon anok plok sprok 8. erok sprok izok hihok ghirok 9. ok-voon anok drok brok jok 10. wiwok farok izok stok 11. lalok sprok izok jok stok 12. lalok brok anok plok nok 13. lalok farok ororok lalok sprok izok enemok 14. wiwok nok izok kantok ok-yurp 15. lalok mok nok yorok ghirok clok 16. lalok nok crrrok hihok yorok zanzanok 17. lalok rarok nok izok hihok mok

Table 1: Sample Arcturan to Centauri and monolingual Centauri sentences

Appendix

2.1 Example of EM for two coins

1. Consider 5 sets (X_i for $i = 1 \ldots 5$) of 10 tosses each. For each set we know which coin was used (A or B). Calculation of the Maximum Likelihood Extimation (MLE) of coin parameters theta is straightforward.

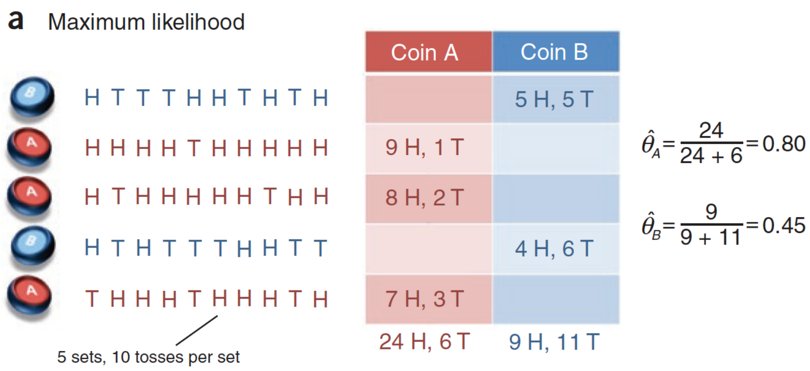

Figure 1: Calculating MLE for a simple example with known parameters values

2. Consider the same 5 sets of 10 tosses each but this time we don't know which coin was used. Application of EM for MLE of theta parameters is given by:

   (a) (Randomly) initializing theta parameters, in this case by 0.6 and 0.5

   (b) Calculating $P(X_i|A)$ and $P(X_i|B)$ for all $i = 1 \ldots 5$. This is done by calculating $P(A|X_i)$, $P(B|X_i)$ and corresponding expected counts for heads, that is $C_i \cdot P(A|X_i)$ and $C_i \cdot P(B|X_i)$, where C_i is the number of heads in the X_i set.

   (c) (Re-) Calculate theta parameters

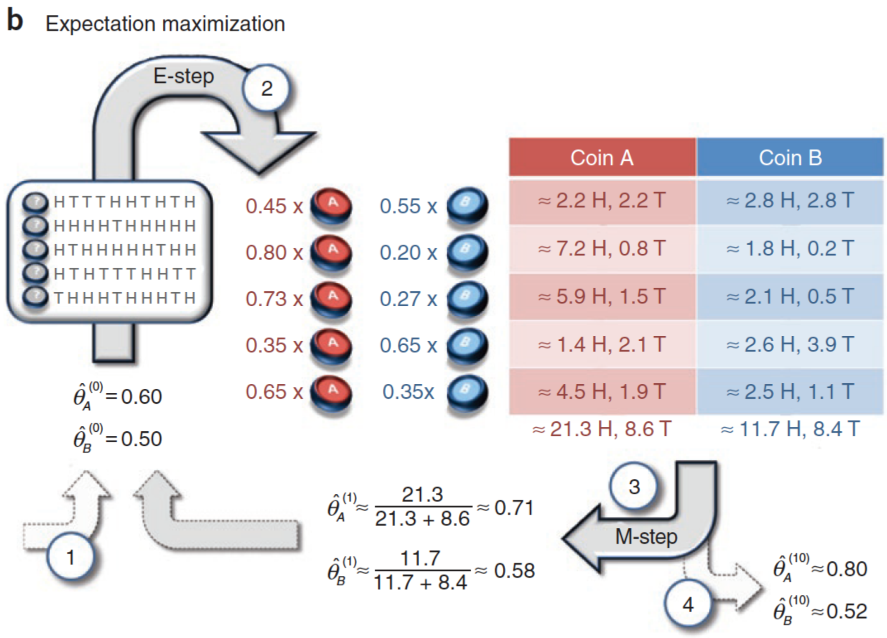

Figure 2: Calculating MLE with EM for a complex example with unknown parameters values

2.2 Detailed explanations

Recall the definition of likelihood:

$$L[\theta|X] = P[X|\theta] = \sum_Z P[X, Z|\theta] \tag{1}$$

where in our case $\theta = (\theta_A, \theta_B)$ are the estimators for the probability that coins A and B respectively land heads, $X = (X_1, \ldots, X_5)$ being the outcomes of our experiments, each X_i consisting of 10 flips, and $Z = (Z_1, \ldots, Z_5)$ being the coin used in each experiment.

We want to find the maximum likelihood estimator $\hat{\theta}$. The Expectation-Maximization (EM) algorithm is one such method to find (at least local) $\hat{\theta}$. It works by finding the conditional expectation, which is then used to maximize θ. The idea is that by continually finding a more likely (i.e. a more probable) θ in each iteration, we will continually increase $P[X, Z|\theta]$ which, in turn, will increase the likelihood function. There are three things that need to be done before going forward with an EM-based algorithm:

1. Construct the model;
2. Compute the Conditional Expectation under the model (E-Step);
3. Maximize the likelihood by updating the current estimate of θ (M-Step).

2.2.1 The model

Before going further with EM, we need to figure out what exactly we are computing. In the E-Step we are computing the expected value of $log(P[X, Z|\theta])$. So what is this value? Observe that:

$$\begin{aligned}
\log P[X, Z|\theta] &= \sum_{i=1}^{5} \log \sum_{C\in\{A,B\}} P[X_i, Z_i = C|\theta] && (2)\\
&= \sum_{i=1}^{5} \log \sum_{C\in\{A,B\}} P[Z_i = C|X_i, \theta] \cdot \frac{P[X_i, Z_i = C|\theta]}{P[Z_i = C|X_i, \theta]} && (3)\\
&\geq \sum_{i=1}^{5} \sum_{C\in\{A,B\}} P[Z_i = C|X_i, \theta] \cdot \log \frac{P[X_i, Z_i = C|\theta]}{P[Z_i = C|X_i, \theta]} && (4)
\end{aligned}$$

The reason is that we have 5 experiments to account for, and we don't know which coin was used in each. The inequality comes from the Jensen's inequality and log being concave. We need to lower bound the equations as we cannot curectly compute the *argmax* of the original equation.

$P[Z_i = C, |X_i, \theta]$ is the probability to see coin C given experiment X_i and θ. Using conditional probabilities we have:

$$P[Z_i = C|X_i, \theta] = \frac{P[X_i, Z_i = C|\theta]}{P[X_i|\theta]} \tag{5}$$

While we have made some progress we are not done designing the model just yet. One question remains: what is the probability that a given coin flipped the sequence X_i? Letting $h_i = \#heads$ in X_i we get:

$$P[X_i, Z_i = C|\theta] = \frac{1}{2} \cdot \theta_C^{h_i}(1-\theta_C)^{10-h_i}, \quad for \quad C \in \{A, B\} \tag{6}$$

Now $P[Z_i = A] = P[Z_i = B] = \frac{1}{2}$ gives us:

$$P[X_i|\theta] = \frac{1}{2} \cdot (P[X_i|Z_i = A, \theta] + P[X_i|Z_i = B, \theta]) \tag{7}$$

2.2.2 E-Step

The EM algorithm starts by making some random guess for θ. In this example we have $\theta^0 = (0.6, 0.5)$. We compute:

$$P[Z_1 = A|X_1, \theta] = \frac{\frac{1}{2} \cdot (0.6^5 \cdot 0.4^5)}{\frac{1}{2} \cdot (0.6^5 \cdot 0.4^5 + 0.5^5 \cdot 0.5^5)} \approx 0.45 \tag{8}$$

This value lines up with what is in the paper. Now we can compute the expected number of heads in $X_1 = (H, T, T, T, H, H, T, H, T, H)$ from coin A:

$$E[\#heads\_by\_coin\_A|X_1, \theta] = h_1 \cdot P[Z_1 = A|X_1, \theta] = 5 \cdot 0.45 \approx 2.2 \tag{9}$$

Doing the same for B we get:

$$E[\#heads\_by\_coin\_B|X_1, \theta] = h_1 \cdot P[Z_1 = B|X_1, \theta] = 5 \cdot 0.55 \approx 2.8 \tag{10}$$

We can compute the same for the number of tails by substituting h_1 by $10 - h_1$. This continues for all other values of X_1 and h_1 with $1 \leq i \leq 5$. Linearity of expectation gives us:

$$E[\#heads\_by\_coin\_A|X, \theta] = \sum_{i=1}^{5} E[\#heads\_by\_coin\_A|X_i, \theta] \tag{11}$$

2.2.3 M-Step

With our expected values in hand, we want to maximize θ. This is done by simple normalization:

$$\theta_A^1 = \frac{E[\#heads\_over\_X\_by\_coin\_A|X, \theta]}{E[\#heads\_and\_tails\_over\_X\_by\_coin\_A|X, \theta]} = \frac{21.3}{21.3 + 9.6} \approx 0.71 \tag{12}$$

Likewise for B. This process starts again with the E-Step and θ^1 and continues until the values for θ converge (or to some allowable threshold). In this example we have 10 iterations and $\hat{\theta} = \theta^{10} = (0.8, 0.52)$. In each iteration the value of $P[X, Z|\theta]$ increases, due to the improved estimate of θ.

Now in this case the model was fairly simplistic. Things can get much more complicated pretty quickly, however the EM algorithm will always converge, and will always produce a maximum likelihood estimator $\hat{\theta}$. It may be a *local* estimator, but to get around this we can just restart the EM process with a different initialization. We can do this a constant amount of times and retain the best results (i.e. those with the highest final likelihood).