Homework Assignment N°2

AML3
Thibault Douzon
Georgios Lioutas

December 15th, 2018

Contents

1 Exercise 1 **3**
1.1 Part a 3
1.2 Part b 3

2 Exercise 2 **4**
2.1 Part a 4
2.2 Part b 5
2.3 Part c 6

1 Exercise 1

1.1 Part a

We want to compute for each sample in the data the responsability $\gamma(z_k)$ of each Bernoulli distribution.
What we call responsability is the probability knowing the data occured that model z_k is involved. We can rewrite it in term of probabilities:

$$\gamma(z_k) = P(z_k|x) = \frac{p(x|z_k)p(z_k)}{p(x)}$$

Where $p(z_k)$ is the prior of each model, $p(x|z_k)$ is given by the distribution z_k (here Bernoulli) and $p(x)$ is the sum over all models inthe mixture of the probability they produced x.
For example with the first point:

- The prior is given by our assumptions: $p(z_k) = [0.5, 0.5]$,
- Each model is a Bernoulli distribution,
  hence $p(x|z_k) = \begin{pmatrix} x_0 + x_1 \\ x_0 \end{pmatrix} {\mu_k}^{x_0}(1 - \mu_k)^{x_1}$
- And the total probability of x is just the marginalization over all models: $p(x) = \sum_{z_k} p(x|z_k)p(z_k)$

We finally obtain the following values:

$$p((1, 4)) \approx 0.243$$

and the responsabilities:

$$\gamma(z_1) = 0.678$$

$$\gamma(z_2) = 0.322$$

We repeat the same procedure for all given inputs and obtain the following results:

x	$\gamma(z_{n1})$	$\gamma(z_{n2})$
$(1, 4)$	0.678	0.322
$(3, 2)$	0.345	0.655
$(4, 1)$	0.208	0.792
$(2, 3)$	0.513	0.487

1.2 Part b

We now want to improve our mixture of model in order to better explain the data we saw. We must compute a new value for the prior and the parameter of each Bernoulli distribution.

Using the responsabities we just computed in the previous step, the EM algorithm gives us formulas to updates thoses values:
The new prior of model z_k:

$$\pi_k^2 = \frac{\gamma(z_k)}{\sum_i \gamma(z_i)}$$

The new proportion of head of model z_k:

$$\mu_k^2 = \frac{\gamma(z_k)\mu_k}{\sum_i \gamma(z_i)\mu_i}$$

But because we now operate in batch after seeing multiple data, we need to modify those equations by summing over all seen points:

$$\pi_k^2 = \frac{\sum_n \gamma(z_{nk})}{\sum_n \sum_i \gamma(z_{ni})}$$

$$\mu_k^2 = \frac{\sum_n \gamma(z_{nk})\mu_k}{\sum_n \sum_i \gamma(z_{ni})\mu_i}$$

With the responsabilities from the previous exercise and the parameters given, we obtain the following results:

$$\pi^2 = [0.436, 0.564]$$

$$\mu^2 = [0.409, 0.570]$$

This means that we have two models with priors π^2. Each one of them is still a Bernoulli model. The first one give a probability of 0.409 to H and the second a probability of 0.570 to H.

2 Exercise 2

2.1 Part a

We are now working with Hidden Makov Models. To compute α and β, we must do a forward pass and then a backward pass.
Starting with α and the forward pass:
The general formula for α is the folowing: $\alpha(z_n) = p(x_1, \ldots, x_n, z_n)$.
But it can be computed with this other recursive definition:

$$\alpha(z_1) = p(x_1, z_1) = p(x_1|z_1)p(z_1)$$

$$\alpha(z_n) = p(x_n|z_n) \sum_{z_{n-1}} \alpha(z_{n-1})p(z_n|z_{n-1})$$

Where the likelihoods $p(x|z)$ are given by the matrix ϕ and the transition probabilities $p(z_n|z_{n-1})$ are given by the matrix A

For $t = 1$, we simply have to multiply the prior of each state by the emission probability that corresponds to the data:

$$\alpha(z_1) = [0.5 \cdot 0.7, 0.5 \cdot 0.2]$$

For the next iteration, we now must use the second part of the recusrsive definition.
For the sake of simplicity, let's only compute the first model value of α:

$$\alpha(z_{2,1}) = p(x_2|z_{2,1}) \sum_{z_1} \alpha(z_1)p(z_{2,1}|z_1)$$

Because the second character in the observation is T, $p(x_2|z_{2,1}) = 0.7$.

$$\alpha(z_{2,1}) = 0.7(0.35 \cdot 0.6 + 0.1 \cdot 0.4) = 0.14$$

We apply this for both models and each time step and we obtain the following table:

	$t=1$	$t=2$	$t=3$	$t=4$
$\alpha_{t,1}$	0.35	0.14	0.026	0.018
$\alpha_{t,2}$	0.1	0.05	0.083	0.039

Now we deal with the backward pass and compute β.
The general formula for β is the folowing: $\beta(z_n) = p(x_{n+1}, \ldots, x_N|z_n)$.
Again we can use a recursive definition to compute it:

$$\beta(z_N) = 1$$

$$\beta(z_n) = \sum_{z_{n+1}} \beta(z_{n+1})p(x_{n+1}|z_{n+1})p(z_{n+1}|z_n)$$

For example we can compute the value of β for the first model at $t = 3$:

$$\beta(z_{3,1}) = \sum_{z_4} \beta(z_4)p(x_4|z_4)p(z_4|z_{3,1})$$

$$\beta(z_{3,1}) = 1 \cdot 0.3 \cdot 0.4 + 1 \cdot 0.8 \cdot 0.6 = 0.6$$

By applying the same procedure for the whole observation, we get the following:

	$t=1$	$t=2$	$t=3$	$t=4$
$\beta_{t,1}$	0.120	0.312	0.6	1
$\beta t, 2$	0.152	0.268	0.5	1

2.2 Part b

Once we have computed α and β for all positions in the observation, it is easy to compute the overall probability of the observation:

$$p(x_n) = \sum_{z_n} \alpha(z_n)\beta(z_n)$$

And because $\beta(z_N) = 1$ we get

$$p(x_N) = \sum_{z_N} \alpha(z_N)$$

In our case:

$$P(O) = p(x_N) = 0.018 + 0.039 = 0.057$$

This is the probability that this observation has been generated by one of the models in the mixture.

2.3 Part c

We can use the formulas from the book to update the prior, the probability of H for each model and the transition matrix:
Our new prior is the following:

$$\pi = [0.774, 0.226]$$

The new transition matrix is:

$$\begin{bmatrix} 0.505 & 0.495 \\ 0.627 & 0.374 \end{bmatrix}$$

And finally the new μ is:

$$\mu = [0.736, 0.234]$$