Homework 2 Advanced Machine Learning

For the deadline see Canvas

Version: Mon 3rd Dec, 2018 at 18:54.

Introduction

1. Each group has to submit a **pdf** with their answers and explanation. **Please put your names and group number at the top of the hand in**.
2. For questions about this homework assignment use the Discussion Board on Canvas.
3. Of course you may use a calculator or a programming environment such as Matlab or Python. **But your report should not contain any code. Explain your computations and results in English!**
4. It is allowed to incorporate handwritten notes or derivations or drawings in your submission as long as these are readable!
5. Explain your answers!

Exercise 1: Mixture Models

In this part we consider the mixture of two Bernoulli distributions. To be more specific: each observation $x = (h, t)$ is the outcome of 5 coin tosses with a specific coin, with h the number of heads and t the number of tails, so $h + t = 5$. But there are two coins in play with possible different probabilities for heads. Hence it can be modeled as a mixture of two probability distributions:

$$P(|\mu, \pi) = \pi_1 P(x|\mu_1) + \pi_2 P(x|u_2)$$

With

$$P((h, t)|\mu) = \binom{h+t}{h} \mu^h (1-\mu)^t$$

in which μ is the probability of head. Assume the following initialization for the parameters of the model: $\pi = (1/2, 1/2)$ and $\mu = (1/3, 1/2)$.

Part a

Consider the data points/observations in the table below, compute for each observation x_n the posterior probabilities, or responsibilities, for each component: i.e. compute $\gamma(z_{n_1})$ and $\gamma(z_{n_2})$.

x	$\gamma(z_{n_1})$	$\gamma(z_{n_2})$
$(1, 4)$		
$(3, 2)$		
$(4, 1)$		
$(2, 3)$		

Part b

Compute the new values for π and μ given the observations above, using a batch approach. That is, calculate the update to the parameters once after seeing all the data. Clearly explain your computations.

Exercise 2: Hidden Markov Models

Consider the simple 2 state Hidden Markov Model M based on the tossing of two coins. The parameters of M are:

1. Initial probabilities $\pi = [0.5, 0.5]$.
2. Transition probabilities $A = [[0.4, 0.6], [0.6, 0.4]]$. So the state transition probability from state 0 to state 1 is $A(0, 1) = 0.6$.
3. Emission probabilities $\phi_1 = [0.3, 0.7]$ and $\phi_2 = [0.8, 0.2]$. The first number is the probability of head H.

Part a

Compute the forward variable α and backward variable β for the observation sequence $O = TTHH$. Explain your computation and fill the following tables:

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$\alpha_{t,1}$				
$\alpha_{t,2}$				

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$\beta_{t,1}$				
$\beta_{t,2}$				

Part b

Compute the probability $P(O|M)$, i.e. the probability that the observation sequence O above is generated by the HMM M. This probability can be computed from the forward variable α.

Hint

Let us denote $P(O|M)$ by $P(O)$ because M is fixed. Show by marginalization that $P(O) = P(O, z_{4,0}) + P(O, z_{4,1})$ (z_4 is the latent variable and $z_{4,i}$ denotes that the state of M is i after observing from O the sequence of first 4 observations, which is O itself). Next link both components in the above decomposition of $P(O)$ to α.

Part c

Given the observation sequences:

1. $O_1 = TTHH$
2. $O_2 = THHT$

Compute the new parameters for the model M after seeing both observations (batch learning) by applying one step of the EM algorithm. You may use the results of exercise 13.12 of the course book.