**Submission Instruction**

1. Report

~~(1) the EM derivation~~

~~(2) the topic-word vectors of the two topics~~

~~(3) the document-topic vector of the first 20 documents in your report~~

2. Upload

~~(1) your implementation for the EM algorithm~~

~~(2) the complete document-topic vectors of the 10000 documents~~

1

SEE PDF

2

Python [0.0385 0.0419 0. 0. 0. 0.1591 0.0794 0.0711]

Java [0.0592 0.0614 0.2441 0.1158 0.1295 0. 0. 0. ]

3  
p(Java|document100) = 0.624950, topic = Java

p(Java|document101) = 0.466044, topic = Python

p(Java|document102) = 0.393987, topic = Python

p(Java|document103) = 0.679373, topic = Java

p(Java|document104) = 0.314074, topic = Python

p(Java|document105) = 0.649435, topic = Java

p(Java|document106) = 0.961393, topic = Java

p(Java|document107) = 0.710673, topic = Java

p(Java|document108) = 0.473344, topic = Python

p(Java|document109) = 0.472750, topic = Python

p(Java|document110) = 0.498325, topic = Python

p(Java|document111) = 0.212979, topic = Python

p(Java|document112) = 0.284598, topic = Python

p(Java|document113) = 0.309914, topic = Python

p(Java|document114) = 0.273899, topic = Python

p(Java|document115) = 0.771862, topic = Java

p(Java|document116) = 0.877277, topic = Java

p(Java|document117) = 0.282659, topic = Python

p(Java|document118) = 0.798260, topic = Java

p(Java|document119) = 0.151169, topic = Python