

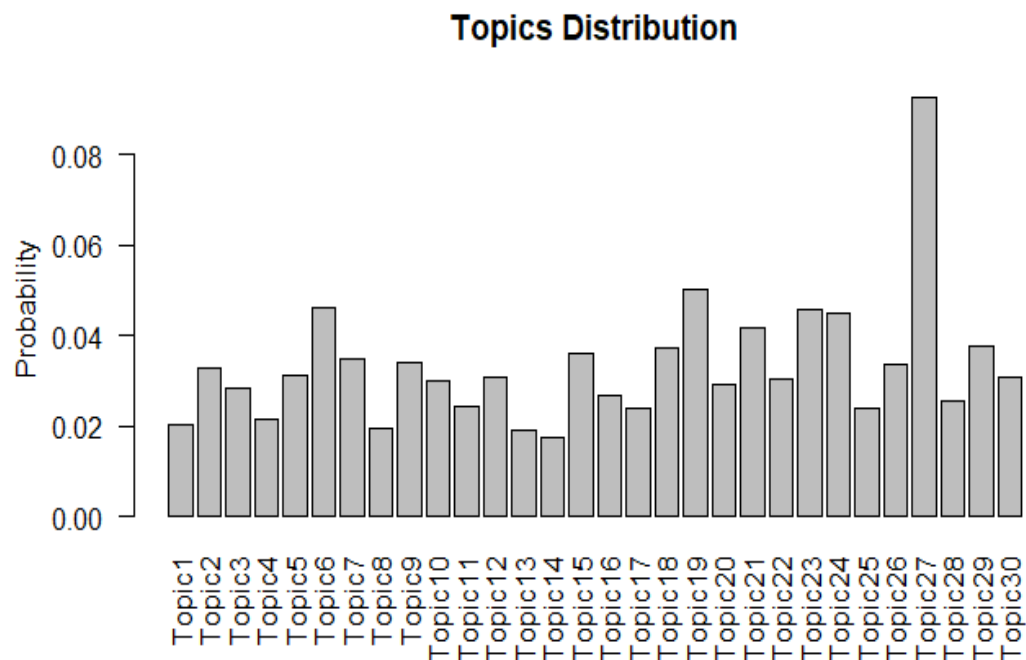
Problem 1

Topics was initialized randomly by assigning random documents to them. The probability distribution for the topics words was calculated according to the documents each topic has, and the same for the topic probabilities (π 's).

Additional probability was added for each topic word (0.001), and then the topic word distribution was normalized to sum up to 1.

E step was implemented in the log space, and the convergence depended on the biggest difference between the weights (deltas) to be smaller a certain value ($\exp(-7)$) with maximum number of iterations equal to 50.

Here's the probability distribution (π 's) for the topics



This table shows the top 10 highest probability words for each topic, sorted from the lowest to the highest probability:

| | | | | | | | | | |
|-----------|-----------|-----------|-----------|-----------|----------|-----------|----------|-----------|---------|
| algorithm | training | set | function | neural | data | input | model | learning | network |
| problem | weight | output | unit | function | system | model | input | learning | network |
| data | set | input | neural | function | learning | algorithm | system | network | model |
| neural | data | training | set | function | input | algorithm | learning | model | network |
| set | training | system | neural | function | data | input | learning | model | network |
| set | error | function | algorithm | input | neural | learning | model | data | network |
| output | algorithm | neuron | system | input | neural | learning | function | model | network |
| input | system | algorithm | set | result | neural | learning | function | model | network |
| data | system | unit | algorithm | model | input | neural | function | learning | network |
| training | data | input | unit | set | learning | system | model | function | network |
| data | training | system | set | input | neural | function | learning | model | network |
| set | training | data | neural | function | learning | input | system | model | network |
| training | neural | input | algorithm | function | data | set | model | learning | network |
| training | neural | function | set | system | data | learning | input | model | network |
| learning | number | neural | pattern | neuron | function | unit | input | model | network |
| output | set | algorithm | data | function | input | neural | learning | model | network |
| training | system | neural | set | cell | input | learning | function | model | network |
| input | error | system | training | set | neural | function | model | learning | network |
| output | function | algorithm | training | system | learning | input | neural | model | network |
| neural | data | input | set | system | function | model | learning | algorithm | network |
| set | unit | neural | error | system | function | input | learning | network | model |
| data | set | algorithm | input | function | unit | neural | learning | model | network |
| set | unit | neural | output | data | learning | function | input | model | network |
| training | set | unit | function | input | neural | algorithm | model | learning | network |
| algorithm | set | data | neural | unit | input | function | learning | model | network |
| weight | system | set | error | neural | learning | input | model | function | network |
| algorithm | training | method | neural | data | input | set | model | function | network |
| system | result | training | set | algorithm | function | neural | model | learning | network |
| output | neural | data | set | algorithm | function | learning | model | input | network |
| word | training | neural | input | set | model | algorithm | function | learning | network |

Problem 2

Here are the different EM segmented images for the first image:



Image1 with 10 clusters



Image1 with 20 clusters



Image1 with 50 clusters

Here are the different EM segmented images for the second image:



Image2 with 10 clusters



Image2 with 20 clusters



Image2 with 50 clusters

Here are the different EM segmented images for the third image:



Image3 with 10 clusters



Image3 with 20 clusters



Image3 with 50 clusters

Part 2

Below is the sunset image segmented into 20 segments with different random initialization each time (KMeans with a different seed)

Conclusion: different initialization leads to a bit different results, and that's because each run can converge into a different local minimum.

Although the differences are subtle, but I will focus on one area. The color and size of the light around the sun at the right of the images is different.





