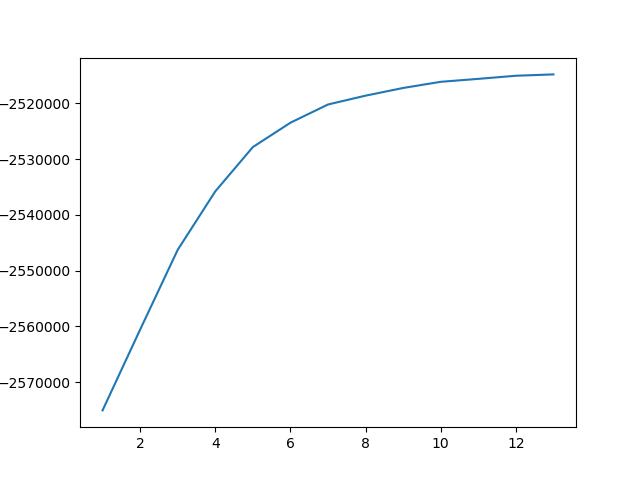
Alon Zanbar \*\*\*\*\*\*\*\*\* , Osnat Drein 302673298

**Applied Probability Models for CS - Ex. 3**

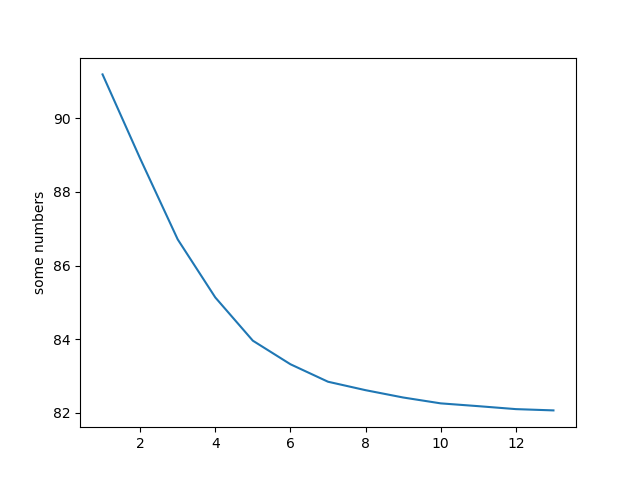
1. The thresholds we use is:  
   | ( current\_likelihood - last\_likelihood ) / last\_likelihood | > 0.0001
2. The log likelihood:

|  |  |
| --- | --- |
| iteration | likelihood |
| 1 | -2575050.9302638522 |
| 2 | -2560565.8508624635 |
| 3 | -2546262.617349307 |
| 4 | -2535777.4529035944 |
| 5 | -2527835.0599140967 |
| 6 | -2523464.7528216336 |
| 7 | -2520198.060784034 |
| 8 | -2518612.6076569105 |
| 9 | -2517235.7717371495 |
| 10 | -2516129.5191852646 |
| 11 | -2515602.298663719 |
| 12 | -2515042.981353067 |
| 13 | -2514801.090842582 |



The perplexity:

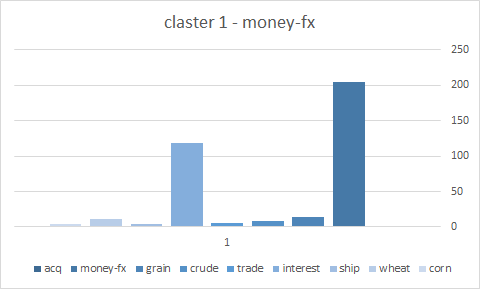
|  |  |
| --- | --- |
| iteration | perplexity |
| 1 | 91.19864699304254 |
| 2 | 88.91256566347317 |
| 3 | 86.71142048715527 |
| 4 | 85.13253491766595 |
| 5 | 83.95571190908825 |
| 6 | 83.31511710599895 |
| 7 | 82.83948326385352 |
| 8 | 82.60961928855835 |
| 9 | 82.41051881497629 |
| 10 | 82.2508943855779 |
| 11 | 82.17492901324736 |
| 12 | 82.09441562185354 |
| 13 | 82.0596200675998 |

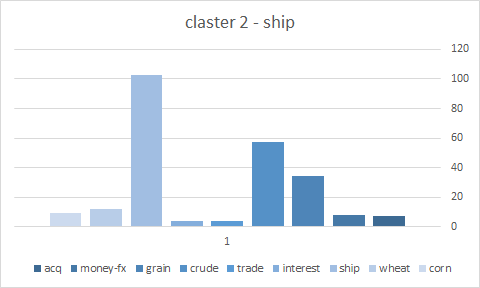


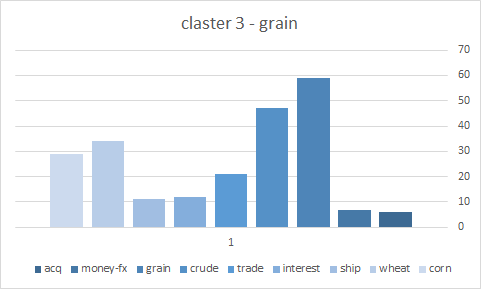
1. Confusion matrix:

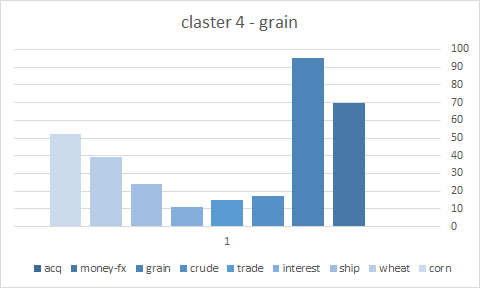
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **acq** | **money-fx** | **grain** | **crude** | **trade** | **interest** | **ship** | **wheat** | **corn** | **sum** |
| claster 1 | 0 | **205** | 14 | 9 | 5 | 119 | 4 | 11 | 4 | 371 |
| claster 2 | 7 | 8 | 34 | 57 | 4 | 4 | **103** | 12 | 9 | 238 |
| claster 3 | 6 | 7 | **59** | 47 | 21 | 12 | 11 | 34 | 29 | 226 |
| claster 4 | 0 | 70 | **95** | 17 | 15 | 11 | 24 | 39 | 52 | 323 |
| claster 5 | 18 | 10 | 2 | **131** | 1 | 3 | 6 | 0 | 1 | 172 |
| claster 6 | 5 | 83 | 3 | 26 | 78 | **115** | 3 | 1 | 1 | 315 |
| claster 7 | **354** | 10 | 2 | 27 | 5 | 1 | 24 | 0 | 2 | 425 |
| claster 8 | 0 | 4 | **165** | 22 | 13 | 1 | 11 | 97 | 57 | 370 |
| claster 9 | 10 | 54 | 14 | 13 | **195** | 9 | 0 | 3 | 5 | 303 |

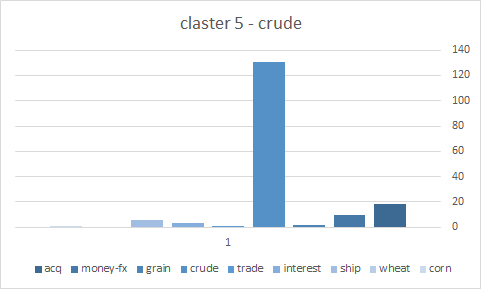
1. histograms from 3 in excel

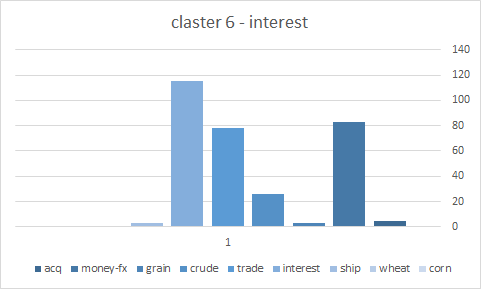


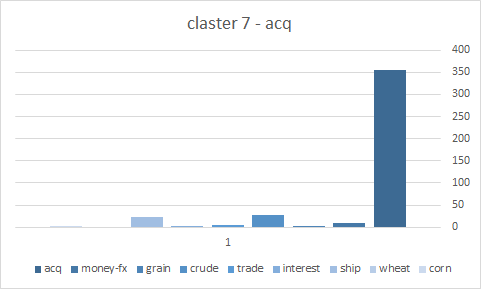


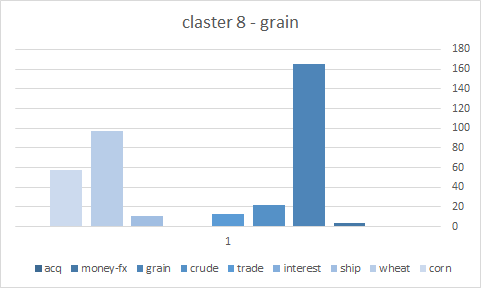


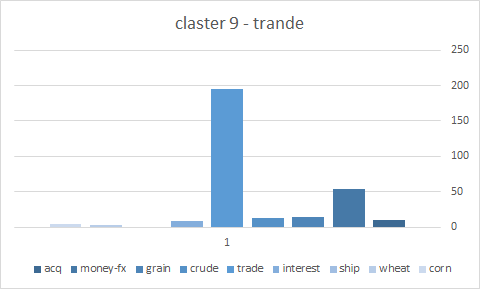












1. The accuracy of the classification is: 0.66949
2. K = 10  
   Vocablary size = 6800  
    λ = 0.3