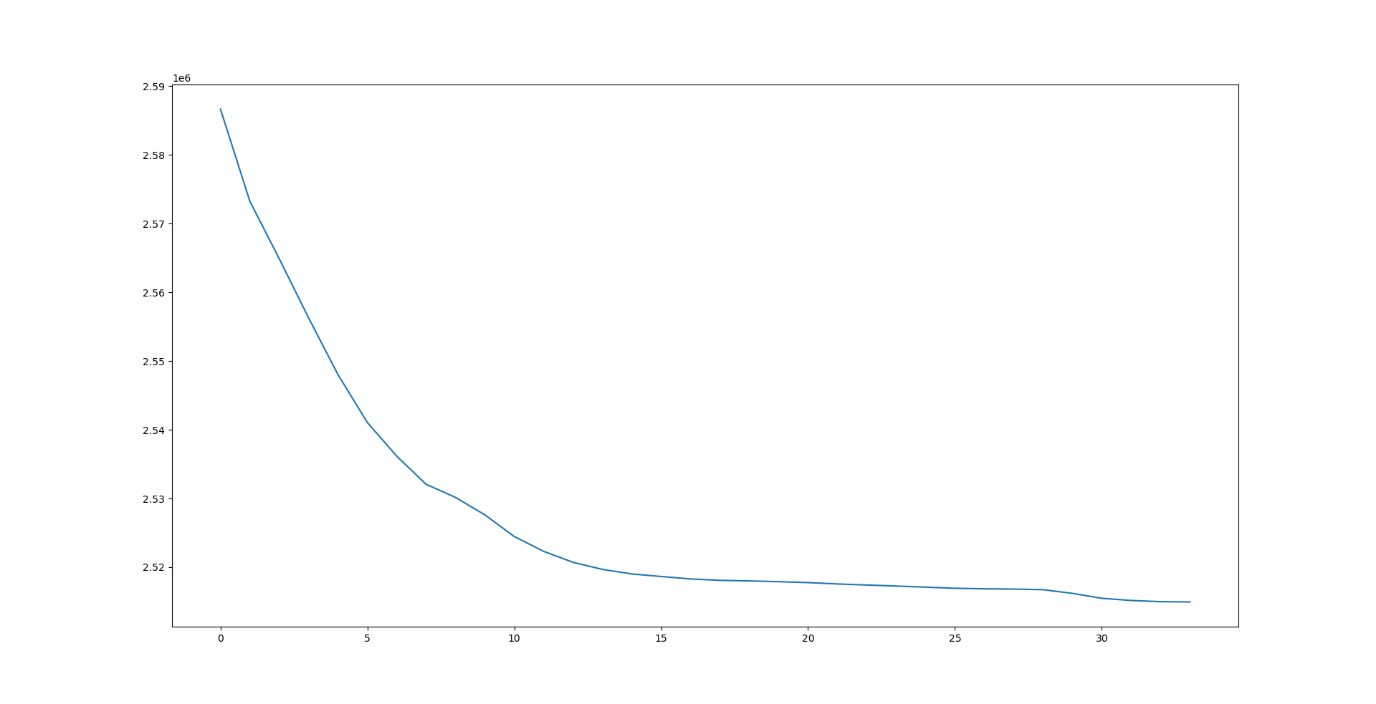
תרגיל 3 מודלים הסתברותיים

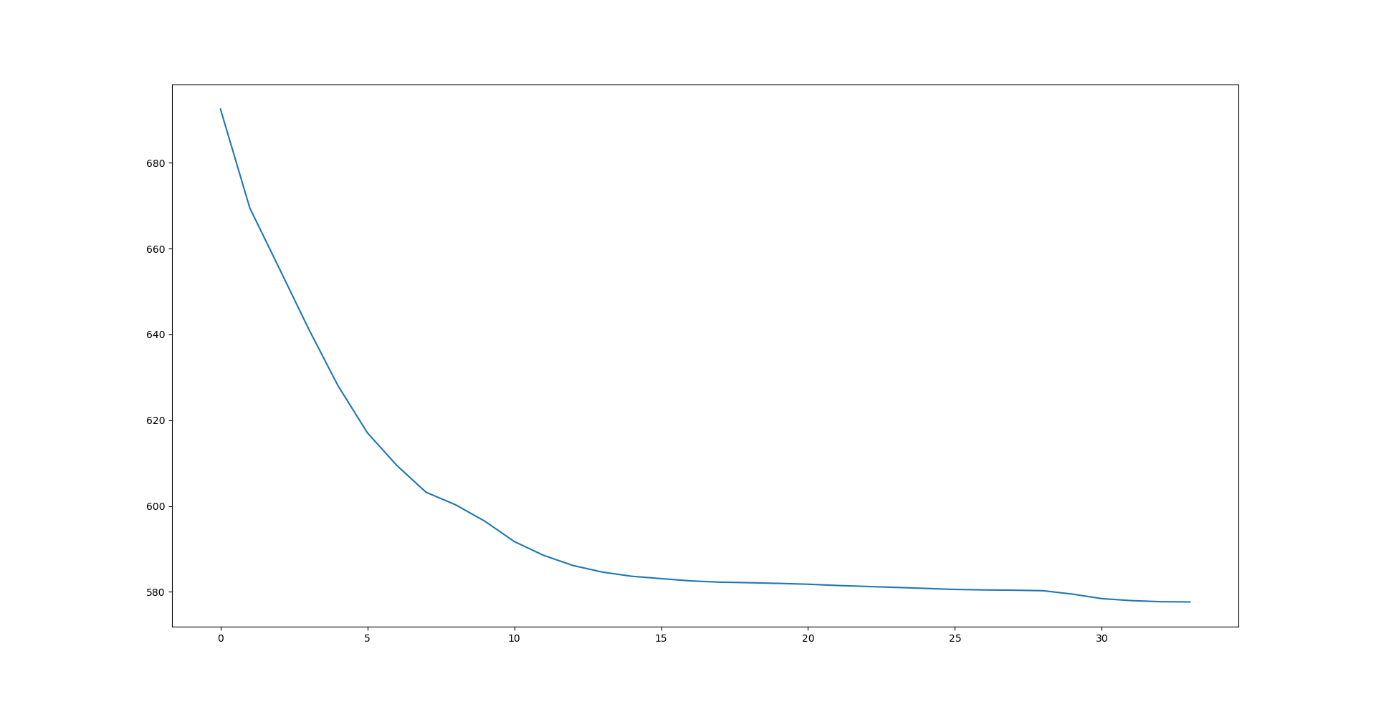
יהונתן שאקי 204920367, אור שחר 209493709

1. ה-threshold הנבחר היה 20 (לאחר שיפור של פחות מ-20 ב-log likelihood הריצה נעצרה)
2. הגרפים הם כדלהלן:

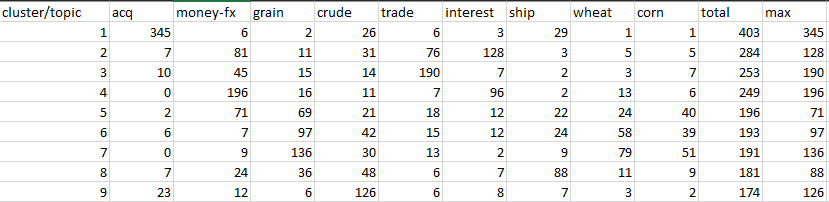
Minus log likelihood:



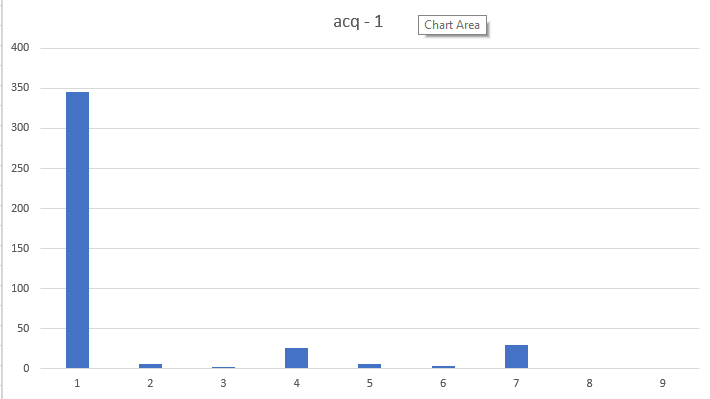
Perplexity:



1. המטריצה היא:



1. ההיסטוגרמות הן:

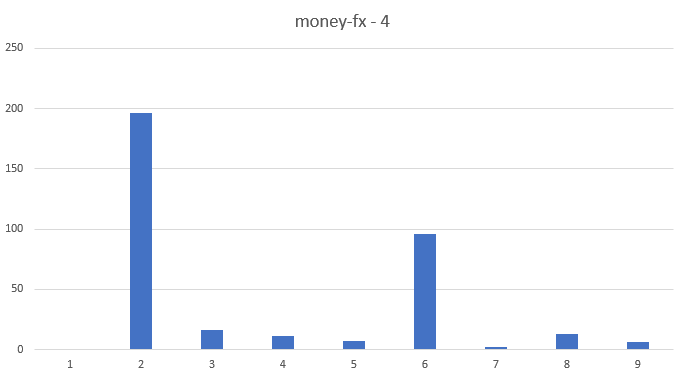


Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated





Chart

Description automatically generated

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

1. הדיוק הוא (מספיק פשוט לסכום את כל המקסימומים ולחלק במספר המסמכים הכולל):

שגדול מ-0.5 :)

1. בחרנו ב-k=-15, lambda=0.08, epsilon=0.005

כפי שהיה אמור להיות, גודל המילון אכן היה 6800.