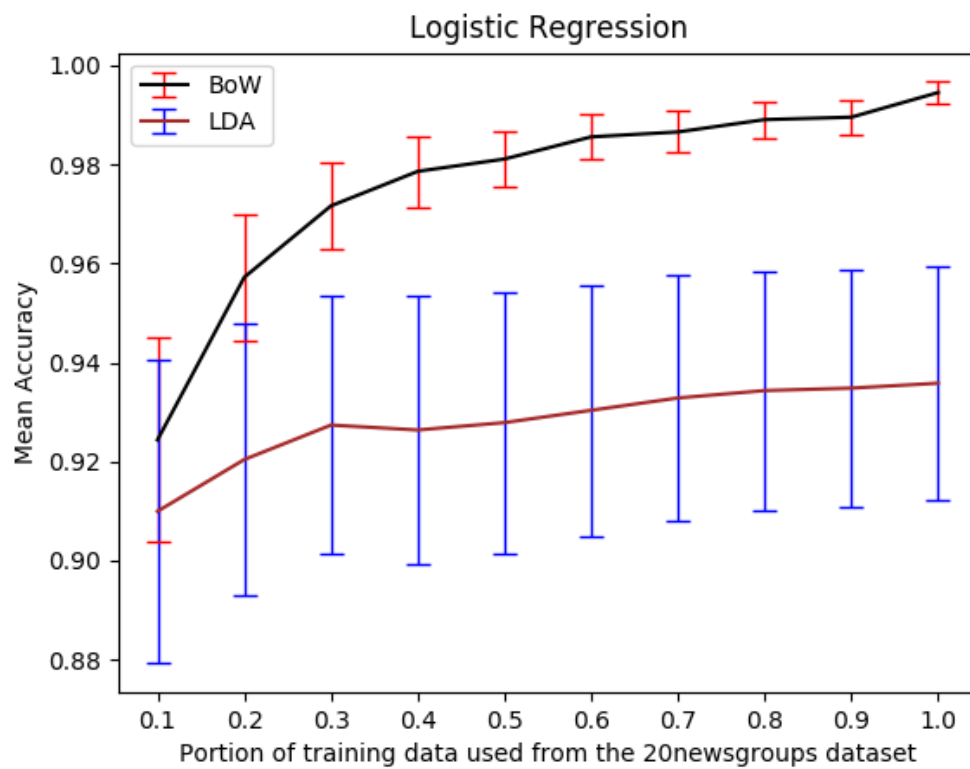
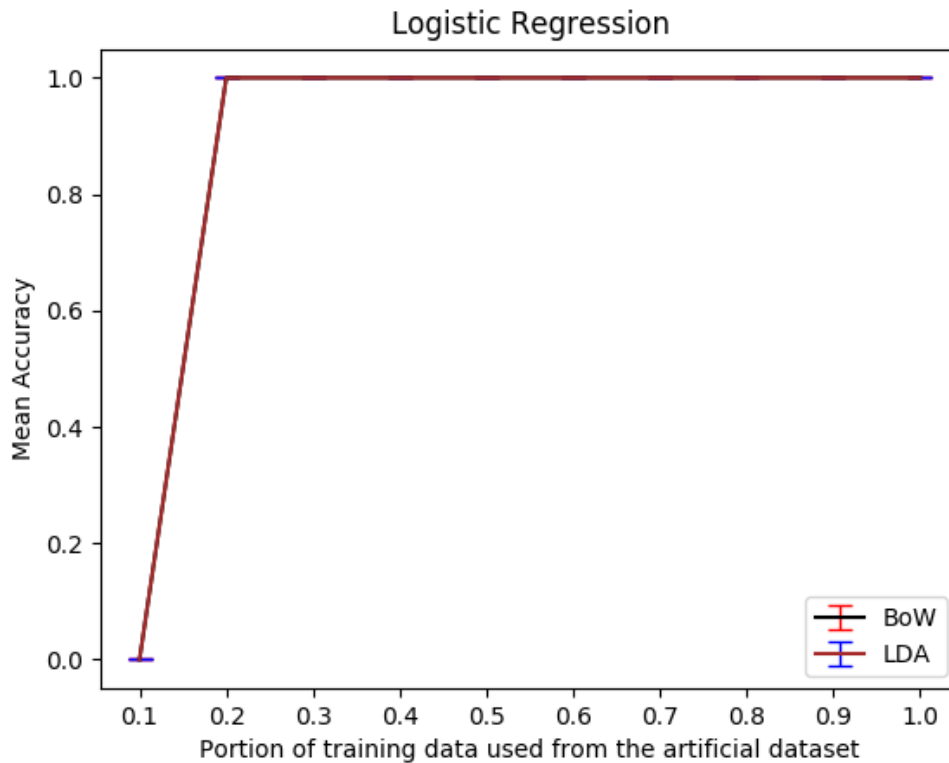


CSCI B-555: Machine Learning Programming Project 4 Report

Aniruddha Patil (2000578987)

12/11/2019





Discussion

The topics unraveled by the LDA seem correlated with respect to the domain of the document.

There are multiple reasons for lesser accuracy of the LDA versus the Bag of Words (BoW) representations.

- Insufficient number of hidden topics
- Insufficient number of iterations
- Improper guess of α and β
- BoW providing a better representation overall due to no coupling, whereas LDA sacrificing some accuracy for more coupling.

Frequent Words

american,isn,gauge,speed,big
gauge,note,ics,top,today
advertising,won,throw,years,give
throw,good,years,big,start
automatic,day,come,university,ford
second,taurus,big,sci,shift
life,moon,read,isn,good
big,good,isn>manual,light
taurus,writes,won,ford,gets
taurus,good,boost,edu,such
thing,sky,good,second,dark
sci,come,space,quite,boost
big,great>manual,note,head
quite,throw,isn,sci,great

throw,sky,great,another,thing
head,boost,dark,sci,redesign
gauge,note,ics,american,quite
thing,ford,years,great,boost
taurus,big,find,ics,isn
taurus,come,american,sci,throw

Conclusion

LDA provides more coupling and might unravel representations, sacrificing some accuracy for explainability. BoW is more independent and looks at the data as it is presented, instead of trying to find any representation in it.