

SDS 385 Exercise Set

Kevin Song

November 3, 2016

I chose a sparsity value of 70, which seemed to allow for about 4-6 topics per singular vector. According to the code, the top 5 singular vectors are:

Singular Vector #1

```
[ 0.57544165] travel
[ 0.54449698] politics
[ 0.05497661] news
[ 0.60775613] computers
```

This singular vector is a bit tricky to interpret. I initially suspected it corresponded to the tech news sector (which seems to really like using Twitter), but I don't understand why travel would appear here in that case.

Singular Vector #2

```
[ 0.44995508] sports_fandom
[ 0.2930393] food
[ 0.06892398] family
[ 0.5557378] religion
[ 0.45737669] parenting
[ 0.4346032] school
```

This seems to correspond to those from a traditional upbringing: religion, football, and sports.

Singular Vector #3

```
[ 0.16808094] photo_sharing
[ 0.58160072] cooking
[ 0.54317472] beauty
[ 0.5817651] fashion
```

This quite clearly corresponds to vanity

Singular Vector #4

```
[ 0.37928222] sports_fandom
[ 0.6493284] food
[ 0.188782] family
[ 0.48125843] religion
[ 0.40283475] parenting
[ 0.07066444] school
```

Curious. Why does this vector show up again?

Singular Vector #5

```
[ 0.1351572] chatter
[ 0.05330012] uncategorized
[ 0.0164934] home_and_garden
[ 0.02286321] sports_playing
[ 0.05141225] business
[ 0.02938269] outdoors
[ 0.04253007] crafts
[ 0.0293551] beauty
```

```
[ 0.96291035] dating  
[ 0.20007547] school  
[ 0.06738907] fashion  
[ 0.01232688] small_business
```

And I have no idea what this corresponds to.