

## 1. What are the hyperparameters?

### a. test run result

AD: DescribeResult(nobs=10L, minmax=(0.44460834284922141, **0.62586512599901789**), mean=0.56490534338011489, variance=0.0030233054508866999, skewness=-0.9767006052297202, kurtosis=0.20046594092452885)

QMR: DescribeResult(nobs=10L, minmax=(0.30590245565208152, **0.37219657000830131**), mean=0.33969487138409327, variance=0.00049676602616625823, skewness=-0.31100711595283453, kurtosis=-1.0541450035569122)

### b. Paper's claimed result

AD: **0.634**

QMR: **0.346**

Number of components / maximum iteration / tolerance of the PLSR regression are the hyperparameters

## 2. The Random Baseline Result

AD: DescribeResult(nobs=10L, minmax=(0.33995878794885581, **0.54826880171893977**), mean=0.44629504965251671, variance=0.0043395970240757767, skewness=-0.22439645643612394, kurtosis=-0.9982319377043707)

QMR: DescribeResult(nobs=10L, minmax=(0.014677641781583527, **0.13395372369906819**), mean=0.078930851874245159, variance=0.0014519521330691237, skewness=-0.029673797877119244, kurtosis=-0.9241926820339597)

## 3. Mode Baseline, Nearest-Neighbor Baseline

### a. Mode Predictor

AD: DescribeResult(nobs=10L, minmax=(0.53121282146346915, **0.62060017607809903**), mean=0.56896583736431317, variance=0.0008205748323736094, skewness=0.3536796865755825, kurtosis=-0.9173923689132963)

QMR: DescribeResult(nobs=10L, minmax=(0.59684451841116948, **0.64821071566121147**), mean=0.62792253765743766, variance=0.00033114888551118198, skewness=-0.5274049949484996, kurtosis=-1.1937367809935386)

b. Nearest-Neighbor Predictor

AD: DescribeResult(nobs=10L, minmax=(0.44838487203753619, **0.58497044001510357**), mean=0.50922798217591381, variance=0.0015713691488781218, skewness=0.25501270345992066, kurtosis=-0.33495588878533233)

QMR: DescribeResult(nobs=10L, minmax=(0.12897023888906747, **0.27752344299156373**), mean=0.20661650068563447, variance=0.0017673698148160373, skewness=-0.13206771784116014, kurtosis=-0.29365050592854924)

#### 4. How good are the paper's results?

On the AD dataset, the google word vectors only slightly outperform the random baseline. Also, the paper's results are only slightly better than the Nearest-Neighbor Predictor. Also, the Mode Predictor is almost as good as the Paper's result when it's run on the AD dataset, and actually *outperforms* when it's run on the QMR dataset. It can be concluded that the paper's results are not so great since they do not add significant value to the straw baseline cases.

#### 5. Mode Baseline, Nearest-Neighbor Baseline

AD, framenet

DescribeResult(nobs=10L, minmax=(0.47014630014947112, **0.67852214644813935**), mean=0.5772293202259029, variance=0.0031501209126493131, skewness=-0.1422946452526227, kurtosis=0.08070360966281864)

QMR, framenet

DescribeResult(nobs=10L, minmax=(0.29164902313462476, **0.35953848248569853**), mean=0.32451931838782438, variance=0.00051342804772366336, skewness=-0.049569823869686766, kurtosis=-1.1761754578666828)

AD, PPDB

DescribeResult(nobs=10L, minmax=(0.5120344522215583, **0.6585714192744142**), mean=0.58100302475425347, variance=0.0021771927190770684, skewness=-0.03394004687283423, kurtosis=-0.6778066885981935)

QMR, PPDB

DescribeResult(nobs=10L, minmax=(0.30938359234302282, **0.39368317643628009**),  
mean=0.34359094114733002, variance=0.00090872617309850385, skewness=0.4062048510421312,  
kurtosis=-1.1060224162396473)

AD, Wordnet

DescribeResult(nobs=10L, minmax=(0.54539551526006458, **0.64407151714421251**),  
mean=0.59572237195965627, variance=0.00068907928502519181, skewness=-0.06637842299699248,  
kurtosis=0.19375682722716148)

QMR, Wordnet

DescribeResult(nobs=10L, minmax=(0.29605604318358081, **0.37396044366785247**),  
mean=0.33384183688769464, variance=0.00078983480198914415, skewness=0.17090028838776578,  
kurtosis=-1.395704697214117)

AD, Wordnet+

DescribeResult(nobs=10L, minmax=(0.53809260579922258, **0.687601318047804**),  
mean=0.61356962910761292, variance=0.0025828791139659224, skewness=-0.21356343577718317,  
kurtosis=-1.3677205122021419)

QMR, Wordnet

DescribeResult(nobs=10L, minmax=(0.29221248584186876, **0.35675807765236162**),  
mean=0.33494672839311618, variance=0.00049407327265666269, skewness=-1.0894657322877284,  
kurtosis=-0.22389824337179354)

We compared the above results against the result we found in 1(a). For all Lexicons, retrofitting improves performance when run on AD data. When run on QMR data, retrofitting improves performance for Wordnet and PPDB lexicons.