

Assignment 2

Topic clustering on **20newsgroup** dataset.

Gamma = 0.01 was chosen based on silhouette score

For **kernelized k-means**, the values are tabulated below

	Homogeneity	Completeness	V-measure	Adjusted Rand Index	AMI	NMI	FMI
Naive TF-IDF	0.302	0.387	0.340	0.009	0.299	0.342	0.194
Sub-Linear	0.380	0.464	0.416	0.174	0.376	0.420	0.255
Max TF-normalization	0.192	0.312	0.238	0.045	0.188	0.245	0.175
Bi-gram	0.01	0.286	0.001	0.01	0.001	0.009	0.224