ELL409 Assignment 3 Report
Saksham Jain
2017MT10747

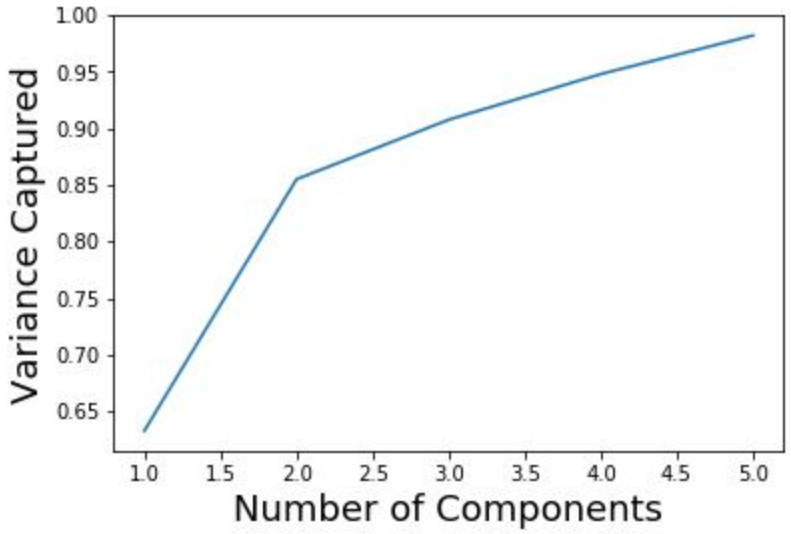
Dataset Used: **Breast Cancer Dataset** (As given in Assignment 2)
It has **569 rows and 30 features** per data point, all **continuous** valued.

The following table shows the principal components, along with corresponding eigenvalues.

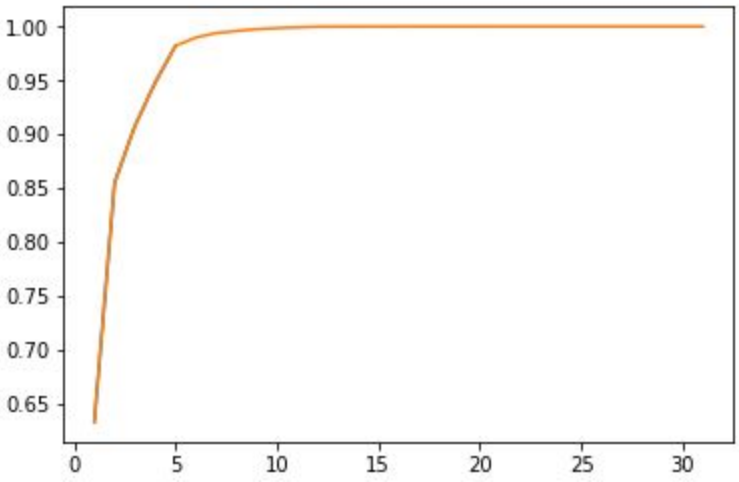
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	EigenValues
-6.21E-05	-0.00023	-0.00128	3.74E-05	-0.00039	0.002533	-0.00152	-0.00016	0.001493	-0.00361	0.007333	0.001108	-0.00041	-0.00021	-0.00154	-0.01088	0.010029	0.003966	-0.00209	-0.00847	-0.00508	0.036	0.603444	0.579503	-0.16602	-0.50119	-0.03361	-0.02175	-0.13587	0.005858	210492.1
2.50E-05	-4.49E-05	-0.00034	0.000101	-7.82E-05	0.000551	-0.00198	-8.42E-05	-9.38E-05	-0.00109	0.002137	0.000606	0.003866	-0.00247	-0.00398	-0.01005	-0.01251	-0.07074	-0.0083	-0.00968	-0.01706	0.875124	-0.02431	-0.0395	-0.4551	0.138048	0.008421	-0.00027	-0.00034	-0.00011	73909.6
-5.49E-06	-3.50E-05	-0.00019	-6.82E-06	-5.83E-05	0.000364	-0.0002	-2.35E-05	0.000242	-0.00053	0.001074	0.000171	-2.46E-05	-0.00016	-0.00027	-0.00151	0.001263	0.0006	-0.00041	-0.00113	-0.00067	0.004937	0.08601	0.075769	-0.02325	-0.06952	0.01883	0.186428	0.971944	0.004041	17502.45
-3.82E-07	-1.87E-06	-1.22E-05	4.72E-07	-3.51E-06	2.49E-05	-1.28E-05	-1.59E-06	1.64E-05	-3.49E-05	8.20E-05	1.17E-05	6.08E-06	2.68E-05	-3.32E-05	-0.00018	9.36E-05	6.35E-05	-1.83E-05	-3.15E-05	-1.25E-05	0.000332	0.006391	0.005189	-0.00128	-0.00456	0.031559	0.011306	0.034289	-0.49272	13425.74
0.063089	0.020434	-0.05206	-0.43328	-0.17685	0.599781	0.451576	-0.16647	0.24015	-0.01939	-0.1746	-0.29384	-0.08075	0.012978	0.023924	-0.01013	-0.01532	-0.00016	-0.00477	0.002561	-0.00076	0.001476	0.000453	-8.99E-05	0.000186	0.000138	-1.51E-05	-1.79E-07	3.74E-06	-1.50E-07	11328.1
0.020472	-0.02203	-0.04356	-0.08477	-0.03232	0.082373	-0.13021	-0.00161	0.165664	-0.16909	0.220918	0.084332	0.01798	-0.89354	-0.03612	-0.00308	0.172736	-0.03157	-0.12642	-0.07925	0.005918	-0.0062	-0.00594	-0.00335	0.000234	5.99E-05	0.000224	-5.80E-05	-0.00014	-1.87E-06	2543.595
0.011696	-0.01404	-0.04411	-0.03408	-0.0205	0.071748	-0.08811	-0.00092	0.091513	0.016342	0.329598	-0.02464	0.071395	-0.02147	-0.17799	-0.01234	-0.80099	0.170441	-0.26619	0.292324	-0.02971	0.000526	0.000882	0.011004	0.000392	0.001973	-0.00059	-0.00014	-0.00011	3.04E-06	1395.17
0.014401	-0.02109	-0.10657	-0.07341	-0.05106	0.206082	-0.10673	0.014523	0.14398	-0.2903	0.685988	0.109136	-0.00248	0.360165	-0.24929	0.180334	0.28416	-0.05049	0.161596	-0.06577	0.013906	0.000296	-0.00515	-0.00891	0.000636	1.11E-05	0.000238	1.05E-05	-2.48E-06	-6.55E-07	642.3484
0.03021	-0.01335	-0.04488	-0.17428	0.052571	0.175469	-0.10151	0.927625	0.068602	0.042599	-0.10862	-0.01845	0.201656	0.025368	0.042929	0.012783	0.007609	0.009112	0.007295	0.002406	-0.00116	0.000771	0.000729	0.000389	0.000386	1.72E-05	5.28E-06	-3.53E-06	9.96E-07	-6.23E-07	566.4135
0.221562	-0.08511	0.347538	-0.67746	-0.11483	-0.55167	0.117573	0.017116	0.08368	0.005135	0.102677	0.110594	-0.02732	0.037419	0.011057	-0.0078	-0.00636	-0.00125	0.010572	-0.00096	-0.00084	0.000643	0.0006	0.000936	-4.59E-06	-0.00014	-5.14E-06	3.55E-06	1.87E-06	-6.21E-08	313.2564
0.002099	-0.00046	-0.01383	-0.00142	0.001087	0.022214	-0.00044	0.004829	0.018468	-0.04049	0.080722	0.008179	0.085664	0.063881	-0.18057	-0.95722	0.083135	0.05419	0.019491	-0.05357	0.029346	-0.00496	-0.10507	0.071057	-0.00305	-0.01804	-0.00402	0.002334	0.000228	4.64E-05	256.0071
0.001489	0.00126	-0.00111	0.001072	0.004464	-0.00151	0.004658	1.92E-05	0.005338	-0.00546	0.002894	-0.00506	0.066472	0.033747	-0.04092	-0.05137	-0.14547	-0.96693	-0.15085	-0.01267	0.035468	-0.09487	0.00288	0.009729	-0.0235	0.006783	0.000923	5.50E-05	4.76E-06	-1.57E-05	179.0813
0.000302	-0.00012	-0.00201	-0.00037	0.000214	0.002327	-0.0006	0.000186	0.003241	-0.00632	0.010392	0.001473	0.012478	0.005401	-0.02384	-0.1219	0.004341	0.005928	0.000566	-0.00266	0.008237	-0.02143	0.712649	-0.686	0.002262	0.067621	-0.01447	-0.01921	-0.00083	0.000401	38.4443
7.17E-06	-3.57E-06	-7.92E-05	-3.49E-06	-3.24E-06	0.000148	-3.47E-05	1.12E-06	0.000145	-0.00023	0.000542	3.97E-05	0.000479	0.00046	-0.00096	-0.00554	0.000352	0.001072	-0.00046	0.000686	0.000666	-0.00048	0.021022	-0.00239	0.004754	-0.039	0.997445	0.016215	-0.02853	-0.01272	17.78488
0.532425	0.832415	0.024122	0.096061	-0.09337	-0.01019	-0.0521	0.020817	0.004363	-0.0167	0.016503	-0.03036	-0.01734	-0.00392	0.004561	-0.00065	0.00027	9.55E-05	0.000183	-0.00022	0.000115	-2.32E-05	4.92E-05	-2.00E-05	3.92E-05	3.67E-06	-1.65E-06	-7.19E-07	2.95E-07	2.50E-08	8.361987
0.095804	-0.06772	-0.11069	-0.04131	0.007205	0.000153	-0.50839	-0.16633	0.849405	-0.03821	-0.36868	0.199986	0.002529	0.15656	-0.01162	-5.46E-05	-0.0011	0.012374	-0.04101	-0.00875	-2.96E-05	0.000399	-0.00027	0.000737	-0.00016	2.69E-05	-4.48E-05	9.97E-06	-1.98E-06	-3.01E-08	4.772099
0.04942	-0.04139	-0.0916	-0.00334	-0.00978	-0.04434	-0.14362	-0.04275	0.168939	0.826106	0.296977	-0.35789	0.018044	-0.03088	0.034197	-0.00381	0.177364	-0.03244	0.017354	0.018797	-0.00475	0.000201	0.000628	-0.00187	-0.00016	0.000116	-1.57E-05	8.07E-06	1.65E-05	1.07E-06	3.918537
0.227929	-0.12073	-0.87885	-0.13198	-0.17333	-0.26458	0.090052	0.001214	-0.16254	-0.05377	-0.07121	0.003904	-0.00091	-0.00698	-0.02908	0.00432	-0.01349	0.002342	0.00432	-0.00066	0.000368	0.000116	6.63E-06	0.000511	0.000104	-4.87E-06	2.41E-05	3.48E-06	-1.61E-08	-5.28E-07	2.188407
0.127541	0.061393	-0.12706	-0.26973	0.923678	0.078787	-0.01828	-0.13446	-0.09411	-0.00523	0.016618	0.001989	0.061501	-0.0018	0.014548	0.007734	0.003908	0.005868	0.002056	0.000964	-0.00036	0.000458	-1.84E-05	0.000202	0.000229	2.90E-05	5.13E-06	2.25E-06	-5.32E-07	-1.46E-07	1.331023
0.759965	-0.51532	0.188448	0.2962	0.028833	0.151593	0.081246	0.017915	-0.04334	-0.00735	0.007464	0.026798	-0.00778	-0.01001	0.004159	0.000209	-0.00451	-0.00032	0.003345	-0.00041	0.00013	-3.10E-06	2.75E-05	-0.00019	4.89E-06	-3.28E-07	4.76E-06	-2.31E-06	-2.84E-07	-1.20E-08	0.598432
-4.38E-05	-0.00018	-0.00086	-9.57E-05	-0.00037	0.001895	-0.00121	-3.18E-05	0.000534	-0.0024	0.005236	0.001386	-0.00062	0.001006	-0.00122	-0.01872	0.007768	0.000385	-0.00011	-0.01018	-0.00395	0.039559	0.324434	0.421365	0.263305	0.791331	0.026382	-0.13435	0.026155	-0.00802	0.453929
-8.71E-06	-5.22E-05	-0.00015	-7.39E-05	-0.00018	0.000594	-0.00167	-0.00016	-0.00061	-8.45E-05	0.001079	0.001004	0.001555	-9.34E-05	-0.00212	-0.00689	-0.01065	-0.06514	-0.01095	-0.00441	-0.00897	0.470174	0.00205	-0.03665	0.83278	-0.28144	-0.01483	0.000811	0.000253	0.000261	0.039657
-3.92E-06	-2.77E-05	-0.00012	-2.24E-05	-5.33E-05	0.000261	-0.0002	-7.44E-06	0.000107	-0.00035	0.000748	0.000204	-6.68E-05	-2.86E-06	-0.00014	-0.00257	0.000807	7.45E-05	0.000171	-0.00114	-0.00043	0.004764	0.055672	0.042865	0.03623	0.11362	-0.0186	0.972422	-0.18591	-0.01255	0.009281
-2.63E-07	-1.26E-06	-6.81E-06	-6.08E-07	-2.99E-06	1.54E-05	-9.64E-06	-1.53E-08	5.55E-06	-1.94E-05	4.81E-05	1.25E-05	1.22E-06	2.54E-05	-1.94E-05	-0.00021	5.99E-05	1.47E-05	2.42E-06	-6.31E-05	2.36E-07	0.000324	0.00293	0.003469	0.003215	0.009548	0.032584	0.018719	0.012939	0.869939	0.008487
0.028731	0.02424	0.011883	-0.23869	-0.18106	0.349716	-0.17613	-0.15067	-0.38232	0.351628	-0.07235	0.633171	0.227631	-9.81E-05	-0.06565	0.018755	0.013041	0.004609	0.00614	0.000772	0.000745	-0.00127	0.00044	0.000265	-0.00073	-0.00012	3.89E-06	3.79E-06	-5.08E-06	4.92E-07	0.001235
0.004404	-0.00969	-0.00833	-0.02554	-0.01407	0.026794	-0.09548	-0.01843	0.019993	-0.03275	0.005556	-0.0041	-0.03416	-0.1604	0.136908	-0.06499	-0.17055	-0.12479	0.853526	0.402947	-0.04262	-0.00051	0.004256	0.006477	0.001069	-0.00027	9.60E-05	-0.00026	0.00012	1.87E-06	0.000704
0.003223	-0.00725	-0.01199	-0.01505	-0.01147	0.021839	-0.06209	-0.01206	0.009015	0.039876	0.06784	-0.03264	0.004191	-0.02621	0.12356	0.003932	-0.39237	0.022125	0.297642	-0.84635	0.110439	-0.00746	-0.00216	-0.00411	0.000421	-0.00329	0.000778	2.17E-05	-1.60E-05	-9.46E-06	2.57E-05
0.005819	-0.01844	-0.06164	-0.05058	-0.04546	0.099172	-0.12651	-0.0125	-0.03149	-0.09585	0.242901	0.092352	-0.15486	0.103175	0.884551	-0.15429	0.004328	-0.00684	-0.2016	0.076789	-0.00197	-0.00078	-0.0026	-0.00333	-0.00219	-0.00059	-0.00026	-1.85E-05	2.16E-05	6.86E-06	2.68E-06
0.003228	-0.01297	-0.00642	-0.08646	0.037558	0.094383	-0.17597	0.157185	-0.11938	0.097252	-0.04305	0.083243	-0.91592	-0.01615	-0.20317	-0.05684	-0.04244	-0.0424	-0.03243	-0.01492	0.008118	-0.00343	-4.48E-05	-0.00154	-0.00231	-0.00027	2.35E-05	-1.49E-06	3.60E-06	2.51E-06	4.30E-07
0.062167	-0.0817	0.098139	-0.22905	-0.13844	0.081866	-0.5899	-0.12202	-0.3975	-0.22619	-0.10883	-0.54237	0.119705	0.080467	-0.05502	0.015138	0.046361	0.009598	-0.05176	0.002352	0.001375	-0.00058	-9.09E-05	-9.57E-05	1.37E-05	0.000162	-2.54E-05	2.04E-05	-1.53E-06	4.10E-07	7.10E-09
-7.22E-05	0.000294	0.001659	0.000978	0.000883	-0.00396	0.003955	-0.00014	0.001942	0.004202	-0.00855	-0.00144	0.002309	-0.0073	0.001212	0.024824	0.010968	0.029512	-0.00136	0.123939	0.991048	0.024281	0.001696	0.008166	0.000824	0.000584	-0.00062	-3.80E-05	1.13E-05	-4.84E-06	8.55E-08

This table can be accessed from ‘Principal_Components.csv’

The following figure shows the variance captured vs the number of components for the first 5 components



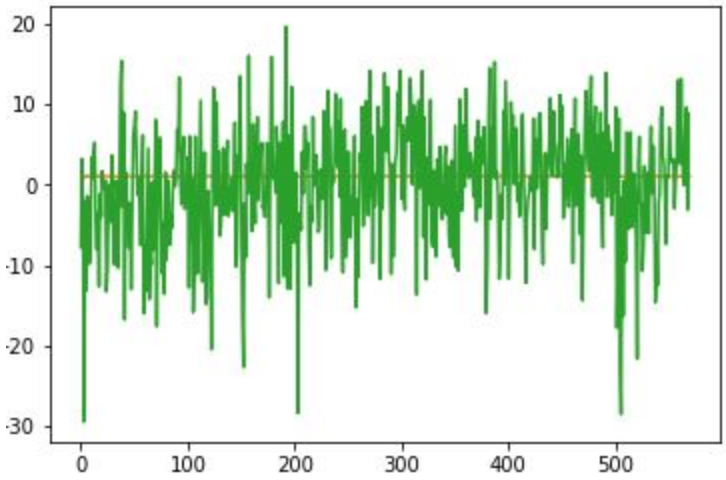
The following figure shows the variance captured vs the number of components for all the components



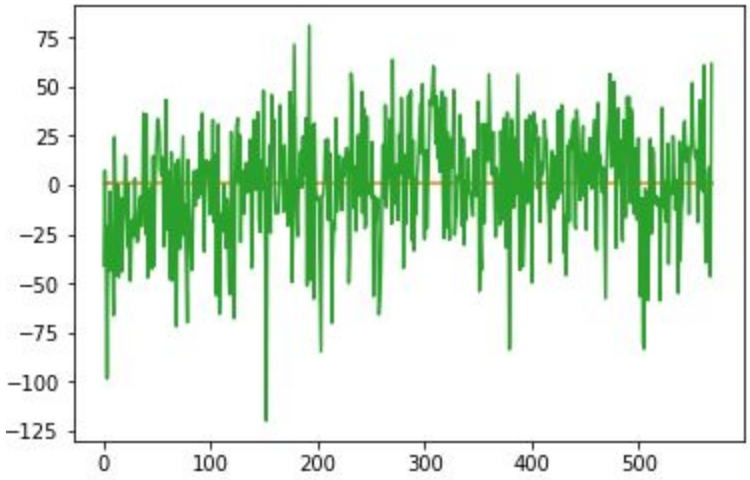
The following table shows numerical variation of variance captured across number of components:

0	float64	1	0.6328081214228533
1	float64	1	0.8550045793364838
2	float64	1	0.907622686548168
3	float64	1	0.9479848720727337
4	float64	1	0.9820408460421969
5	float64	1	0.9896877247152803
6	float64	1	0.9938820638950729
7	float64	1	0.9958131735565928
8	float64	1	0.997515998184408
9	float64	1	0.9984577496256195
10	float64	1	0.9992273908734283
11	float64	1	0.9997657680334285
12	float64	1	0.9998813441825941

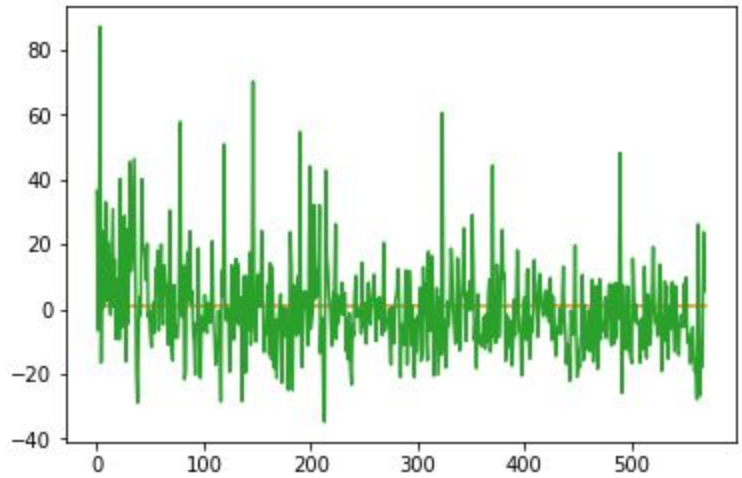
The variables f1, f2, f3, f4, f5 in the code file attached correspond to the projection of the data onto the eigenvectors in decreasing order of eigenvalues.
Plots for the same are as follows:
Plot of f1:



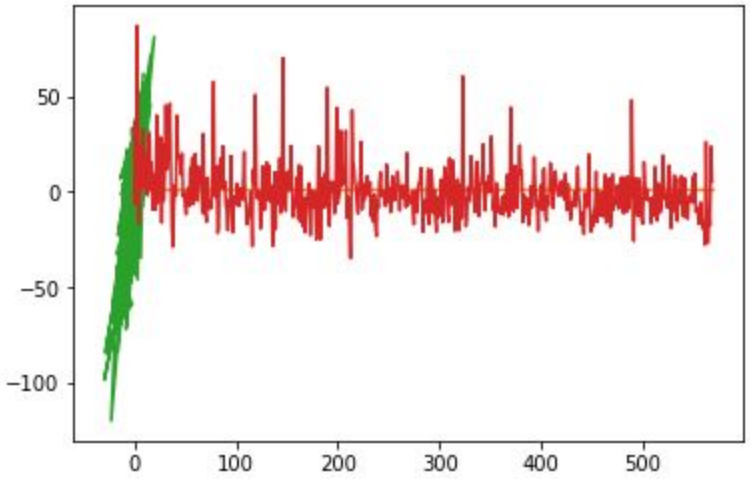
Plot of f2:



Plot of f3:

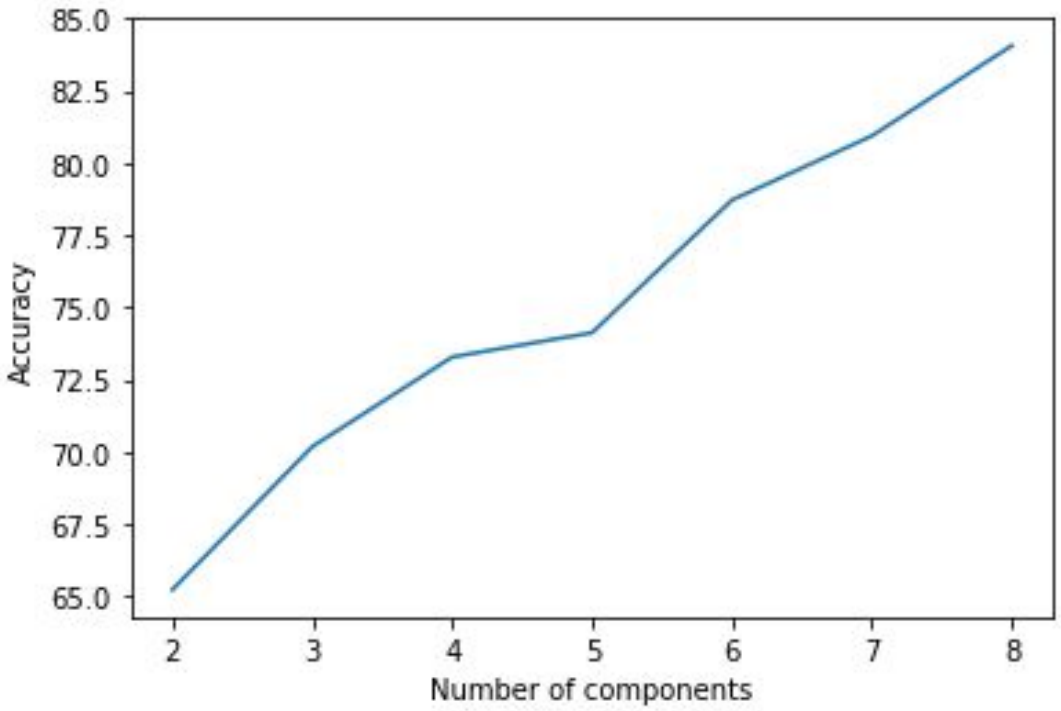


Plot of (f1, f2, f3):



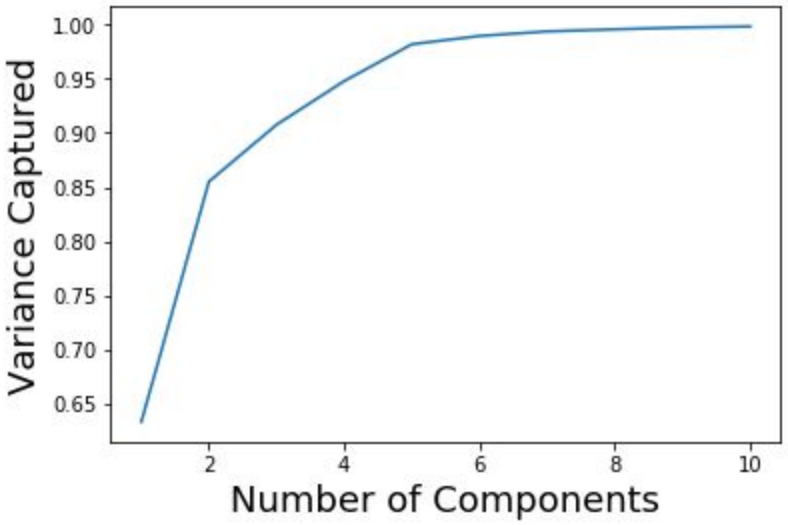
To capture 75% variance , we require 2 components.
To capture 85% variance , we require 2 components.
To capture 95% variance , we require 5 components.

The following plot shows Test Set Accuracy vs Number of Components used for a **Gaussian Naive Bayes Classifier:**

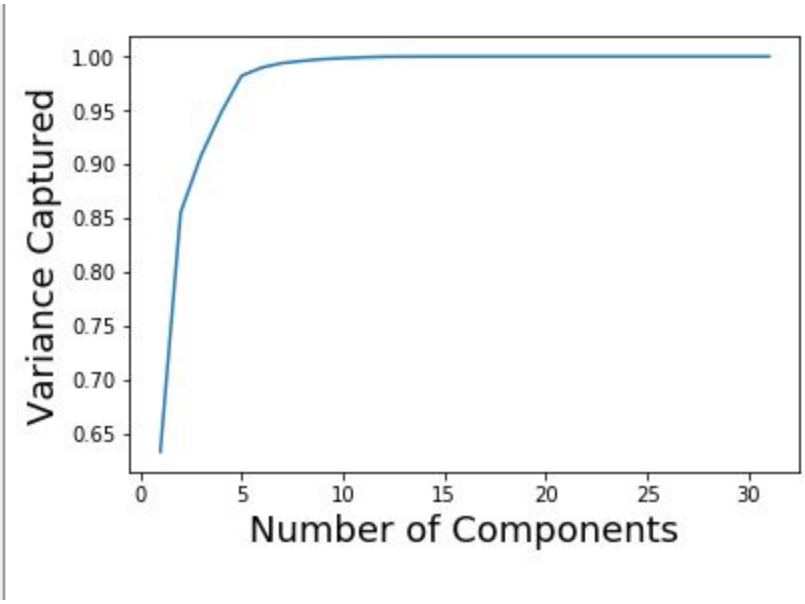


Sanger's M-Unit Rule:

The following figure shows the variance captured vs the number of components for the first 10 components :



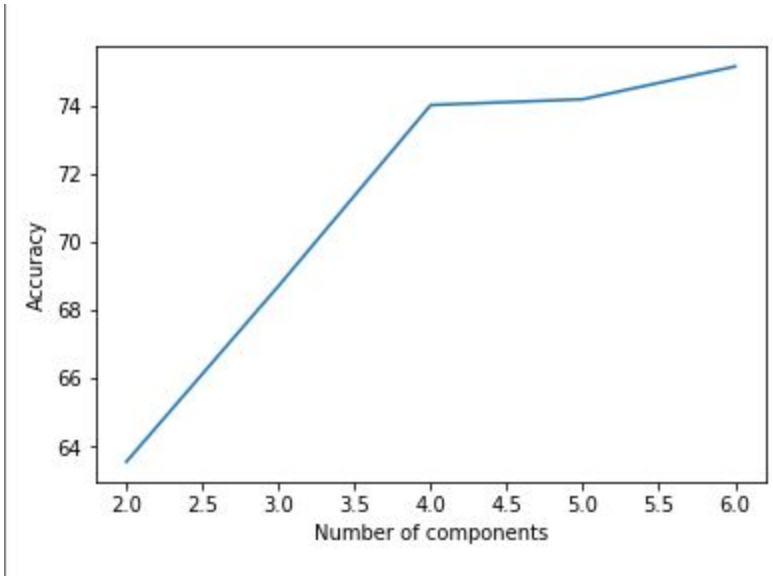
The following figure shows the variance captured vs the number of components for all the components :



The table below shows variance captured vs number of components:

0	float64	1	0.6328088703175576
1	float64	1	0.8550055911887603
2	float64	1	0.9076237606679151
3	float64	1	0.9479859939578462
4	float64	1	0.9820420082298187
5	float64	1	0.9896888959339843
6	float64	1	0.9938832400590477
7	float64	1	0.9958143520059111
8	float64	1	0.9975171786444669
9	float64	1	0.9984589311793188
10	float64	1	0.9992285732515929
11	float64	1	0.9997669510462895
12	float64	1	0.999882527325989
13	float64	1	0.9999359945182329
14	float64	1	0.9999611334178717
15	float64	1	0.9999754792582308
16	float64	1	0.9999872595507094
17	float64	1	0.9999938377873149
18	float64	1	0.9999978392807746
19	float64	1	0.9999996334608926

The following plot shows Test Set Accuracy vs Number of Components used for a Gaussian Naive Bayes Classifier:

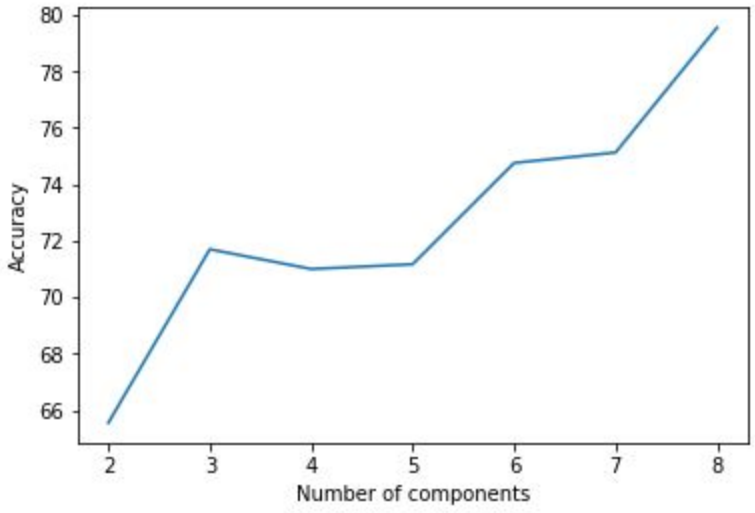


LDA

Components Captured can be seen from ‘LDA.csv’
Here is a glimplse of the same file:

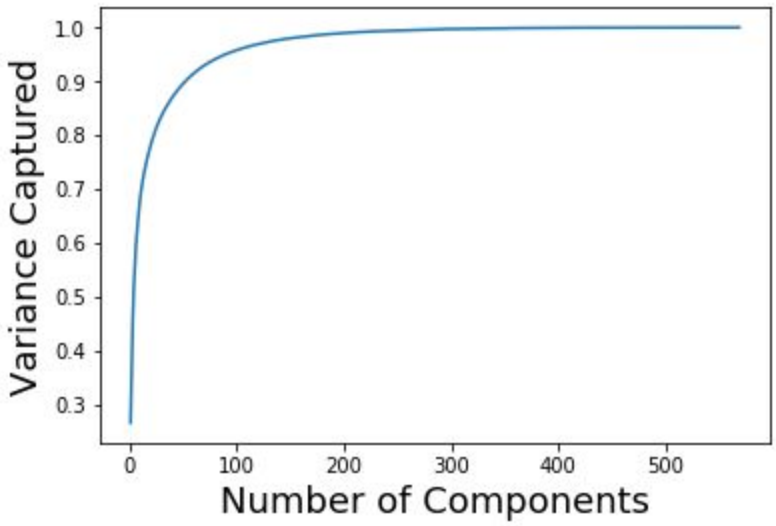
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	
0	(-0.00738	(-0.00418	(0.017822	(-0.01750	(0.027377	(0.022180	(-0.44280	(-0.44077	(0.487620	(0.486884	(-0.20934	(0.149268	(-0.21068	(0.147737	(-0.017
1	(-0.00056	(0.000926	(0.000502	(0.000287	(0.001695	(-0.00115	(-0.00137	(-0.00757	(-0.00698	(-0.01018	(-0.00112	(0.012465	(0.005527	(0.012966	(-0.005
2	(0.029751	(0.030223	(-0.12245	(0.122164	(-0.68627	(-0.68514	(-0.10569	(-0.10511	(-0.03949	(-0.03966	(0.035205	(-0.00293	(0.035133	(-0.00307	(0.002
3	(0.084195	(0.084225	(-0.68894	(0.688952	(0.130949	(0.130619	(-0.02888	(-0.02891	(-0.00996	(-0.00975	(-0.00783	(-0.00134	(-0.00780	(-0.00143	(-0.000
4	(-0.19712	(0.197119	(0.121581	(0.121526	(-0.02237	(0.022454	(0.113606	(-0.11342	(-0.21558	(0.214123	(-0.15867	(0.106785	(0.159930	(-0.11049	(-0.438
5	(-0.08735	(0.087307	(-0.13144	(-0.13150	(-0.05702	(0.057313	(-0.28824	(0.288957	(0.007787	(-0.00759	(0.332446	(-0.12535	(-0.32888	(0.125165	(0.028
6	(-0.06752	(0.067557	(-0.00374	(-0.00383	(-0.09915	(0.099188	(0.182025	(-0.18317	(-0.16497	(0.165794	(0.018638	(-0.03150	(-0.02569	(0.032651	(0.207
7	(-0.15474	(0.154754	(-0.13441	(-0.13442	(-0.44958	(0.450793	(-0.04950	(0.048658	(0.150664	(-0.14724	(0.065798	(0.379774	(-0.06517	(-0.37962	(-0.077
8	(-0.09321	(0.093212	(-0.14030	(-0.14029	(-0.19687	(0.197393	(-0.23093	(0.230537	(-0.14874	(0.149937	(0.071935	(-0.08916	(-0.07427	(0.088809	(0.126
9	(0.014058	(-0.01405	(0.028736	(0.028727	(-0.01552	(0.015571	(0.031694	(-0.03185	(0.066544	(-0.06622	(0.035410	(0.017353	(-0.03507	(-0.01728	(0.079
10	(-0.01574	(0.015883	(-0.00123	(-0.00144	(-0.00717	(0.006802	(0.001412	(-0.00275	(0.038281	(-0.03949	(0.031067	(0.051589	(-0.03728	(-0.04428	(0.083
11	(0.000111	(-0.00012	(0.000189	(0.000237	(5.077746	(-5.10830	(4.061523	(-0.00010	(-6.51236	(0.000304	(-9.87350	(-0.00060	(-9.88871	(-0.00035	(0.001
12	(-0.00232	(0.001929	(0.001451	(0.001688	(0.006953	(-0.00640	(-0.00101	(-0.00113	(0.011058	(-0.00341	(-0.04597	(-0.00852	(-0.04328	(-0.00969	(0.701
13	(-0.00418	(-0.00399	(-0.00377	(0.003647	(0.002770	(0.002235	(-0.02430	(-0.02371	(0.257129	(0.257274	(0.200869	(-0.62377	(0.204037	(-0.62496	(0.009
14	(0.172576	(-0.17257	(0.117662	(0.117862	(-0.27333	(0.273305	(0.015045	(-0.01514	(-0.40017	(0.403265	(-0.21381	(0.057141	(0.214095	(-0.05680	(0.272
15	(-0.12650	(0.126501	(-0.03485	(-0.03484	(0.245713	(-0.24603	(0.219262	(-0.21836	(-0.00498	(0.006347	(0.129204	(0.405932	(-0.12549	(-0.40742	(0.246
16	(-0.06499	(0.064983	(-0.06736	(-0.06736	(-0.01715	(0.017339	(-0.13740	(0.137680	(-0.07828	(0.078454	(-0.07248	(0.144369	(0.073779	(-0.14505	(-0.133
17	(-0.51131	(0.511307	(0.383203	(0.383282	(0.045655	(-0.04568	(-0.15318	(0.155406	(-0.01286	(0.011324	(-0.03147	(-0.08305	(0.035144	(0.084509	(0.133
18	(0.006100	(-0.00609	(0.011324	(0.011317	(-0.00176	(0.001760	(0.005477	(-0.00555	(0.022364	(-0.02240	(0.010191	(-0.00198	(-0.01011	(0.002211	(0.013
19	(-0.22783	(0.227835	(-0.44673	(-0.44666	(0.153750	(-0.15352	(-0.07989	(0.081282	(0.000494	(-0.00226	(-0.40900	(-0.05214	(0.412597	(0.054613	(0.120
20	(0.008512	(0.010996	(0.004153	(-0.00504	(-0.02691	(-0.02814	(0.030396	(0.034921	(-0.15952	(-0.15830	(-0.63250	(-0.27356	(-0.62777	(-0.27075	(-0.041
21	(3.837165	(0.001153	(0.000946	(-0.00028	(-0.00153	(0.000491	(-0.00088	(4.573040	(-0.00802	(-0.00771	(-0.02261	(0.042427	(-0.02573	(0.035292	(0.068
22	(-0.00606	(-0.00570	(0.046363	(-0.04680	(0.104221	(0.105179	(-0.53912	(-0.53961	(-0.41225	(-0.41093	(0.117653	(-0.10977	(0.118524	(-0.10942	(0.008
23	(-0.70132	(-0.70129	(-0.08841	(0.088399	(-0.01512	(-0.01507	(0.000770	(0.000798	(-0.00718	(-0.00714	(-0.00866	(-0.00071	(-0.00863	(-0.00068	(-0.000
24	(-0.14272	(0.142722	(-0.22546	(-0.22544	(0.030873	(-0.03117	(0.226296	(-0.22454	(-0.23112	(0.231197	(0.235637	(-0.03467	(-0.23432	(0.034267	(-0.095
25	(-0.02903	(0.029046	(-0.03255	(-0.03252	(-0.02826	(0.028204	(0.012007	(-0.01207	(-0.02517	(0.024607	(0.029639	(-0.05868	(-0.03053	(0.057391	(0.029
26	(-0.02441	(0.024481	(0.050563	(0.050553	(-0.02100	(0.021043	(0.033630	(-0.03299	(0.014606	(-0.01514	(0.083159	(0.020235	(-0.08625	(-0.01550	(0.105

The following plot shows Test Set Accuracy vs Number of Components used for a Gaussian Naive Bayes Classifier:



Kernel PCA

The following graph shows the number of components vs variance captured:



The following table shows numerical variation of variance captured across number of components:

Index	Type	Size	
0	float64	1	0.2659096934831364
1	float64	1	0.3633720675101518
2	float64	1	0.45389777231866757
3	float64	1	0.5158437282003095
4	float64	1	0.5593667079995778
5	float64	1	0.5953879499769009
6	float64	1	0.6215951525396239
7	float64	1	0.6438031837617548
8	float64	1	0.6651540188612318
9	float64	1	0.6845495659266562
10	float64	1	0.699898824681299
11	float64	1	0.7122056196705261
12	float64	1	0.7238057157985499
13	float64	1	0.7341853368837816
14	float64	1	0.7435763023837851
15	float64	1	0.7524353708903211
16	float64	1	0.761151847944809
17	float64	1	0.7688777341503742
18	float64	1	0.7763758210958462
19	float64	1	0.7834807881530946

To capture **80% of the variance, 24 components** were required