

Problem Setup

The dataset has 1086 samples, each sample has features

	GSM316183	GSM316184	GSM316185	GSM316186	GSM316187	GSM316188	GSM316189	GSM316190	GSM316191
1007_s_at	9.466410	8.785954	8.945697	8.800308	8.371134	8.009487	9.024723	8.234867	7.936837
1053_at	9.300222	9.296332	9.374943	10.073341	10.351786	10.249215	9.663270	9.879482	9.670637
117_at	3.442133	5.841318	5.008332	6.574541	5.918695	5.600266	6.492298	6.414826	6.341468
121_at	9.688976	9.448451	9.467896	9.421499	8.804552	9.185173	9.364890	9.087581	8.753255
1255_g_at	4.515020	3.649851	4.812977	6.122662	5.177852	5.842581	4.533285	4.084906	4.528694
1294_at	8.531066	8.422788	8.826361	8.142186	7.993740	8.280510	8.231663	8.061548	8.343496
1316_at	5.965259	5.951902	6.001788	6.077552	6.311912	5.623541	5.886500	5.794314	5.625299
1320_at	2.892429	2.705467	1.732685	1.707659	2.935347	2.812667	2.708900	4.710675	2.580294
1405_i_at	9.280956	9.411652	9.129943	8.668567	8.804932	8.593733	9.345318	9.587427	9.674255
1431_at	5.488501	5.189835	5.046043	3.812368	5.706839	4.727902	4.401646	4.755690	4.720619
1438_at	3.670293	4.947504	4.827182	5.364968	4.248668	6.139010	4.635325	3.170527	3.170036
1487_at	8.893249	8.464358	8.460420	8.497818	8.497466	8.433403	8.639447	8.754044	8.712932
1494_f_at	6.759217	6.740798	6.065692	7.126267	6.248790	6.018455	6.538063	6.647617	6.226009
1598_g_at	8.356571	8.288346	8.185832	8.219428	8.242359	7.960347	8.507478	7.973756	7.234547
160020_at	7.366809	7.208398	7.134906	7.833785	6.834711	6.702703	7.793795	6.661017	7.072390
1729_at	8.515638	8.173193	8.670208	8.056884	7.785230	7.997558	7.832485	7.701622	8.192453

The goal is to build a Bayesian networks-based classifier

Model: $R^{22277} \rightarrow R^3$

Feature Selection

Assume the optimal number of features is between 10 ~ 100

Selection criteria: (1) ANOVA F-value; (2) chi-squared stats

For i in range(MC):

 For num in range(10, 100):

 Score[num] = criteria

Rank(num) / Frequency(num)

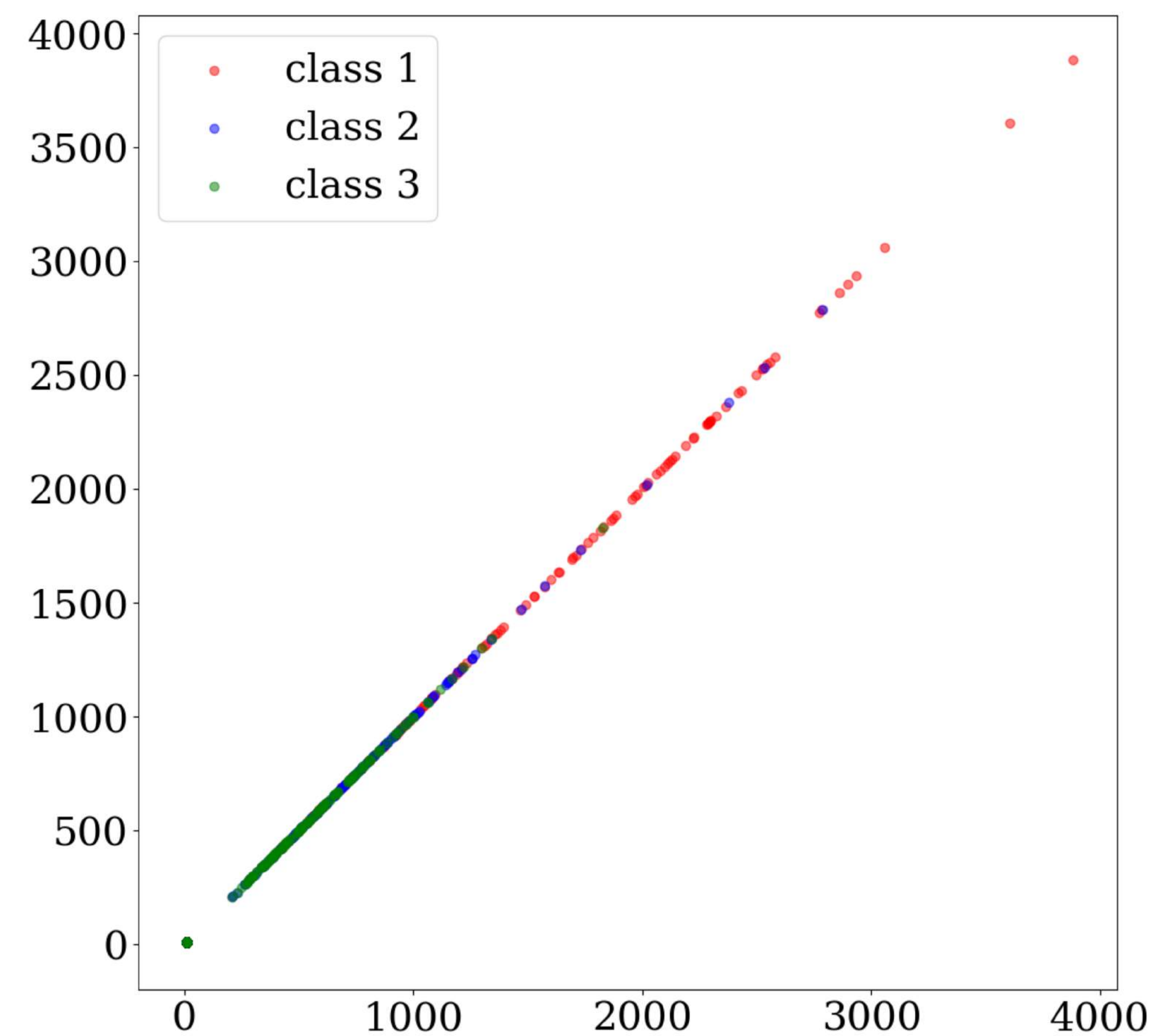
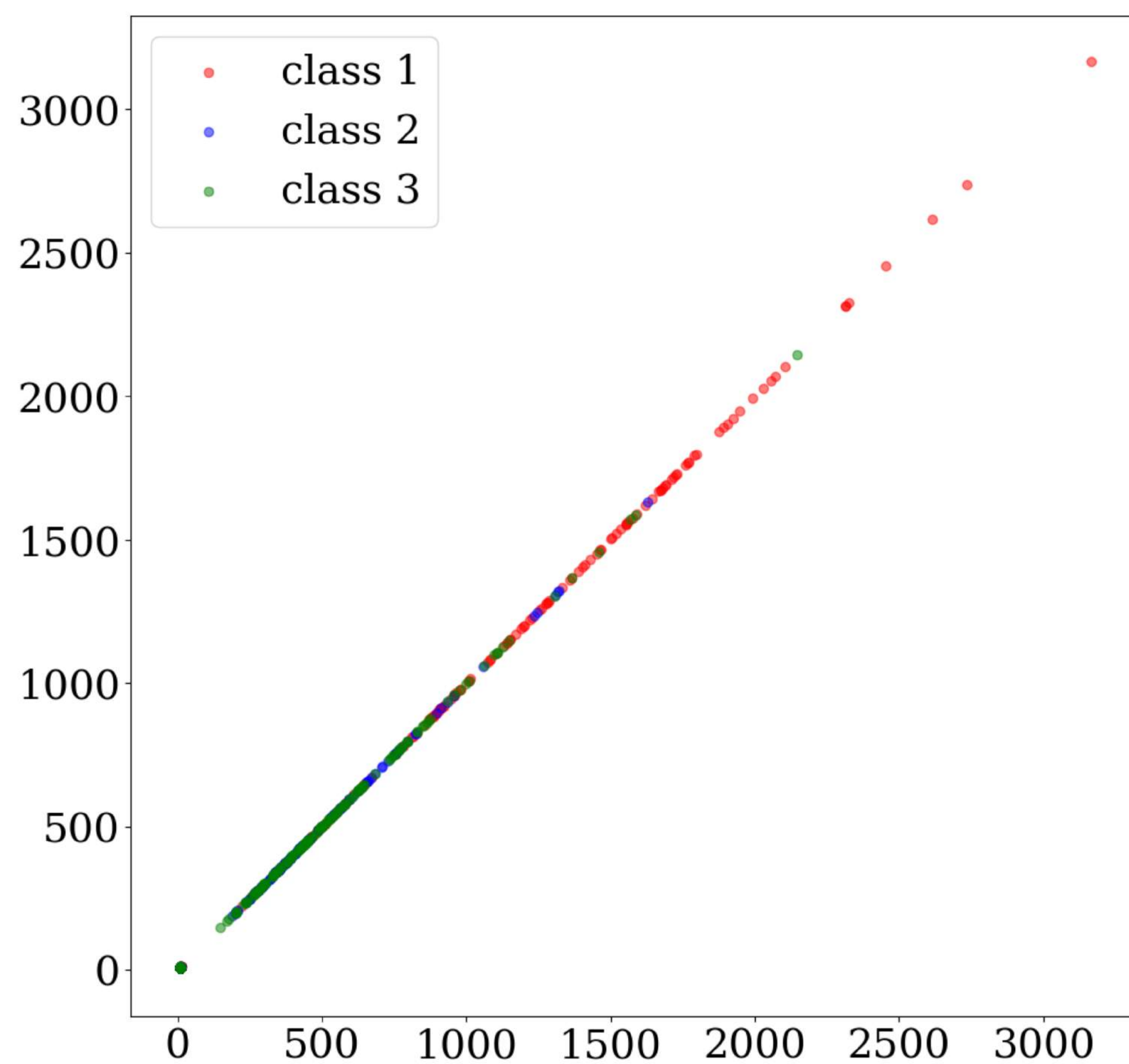
Num = [56, 70, 39, ...]

Model: $R^{56} \rightarrow R^3$

Data Transform

Model: $\mathbb{R}^{56} \rightarrow \mathbb{R}^3$

Continuous variable to categorical variable



Compute the mean, larger $\rightarrow 1$ and smaller $\rightarrow 0$

Data Transform

Model: $R^{56} \rightarrow R^3$

Continuous variable to categorical variable

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	0	0
7	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
8	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	1	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1

Showing 1 to 17 of 1,086 entries, 58 total columns

Data Transform

Model: $\mathbb{R}^{56} \rightarrow \mathbb{R}^3$

