AMMI_BootCamp_project

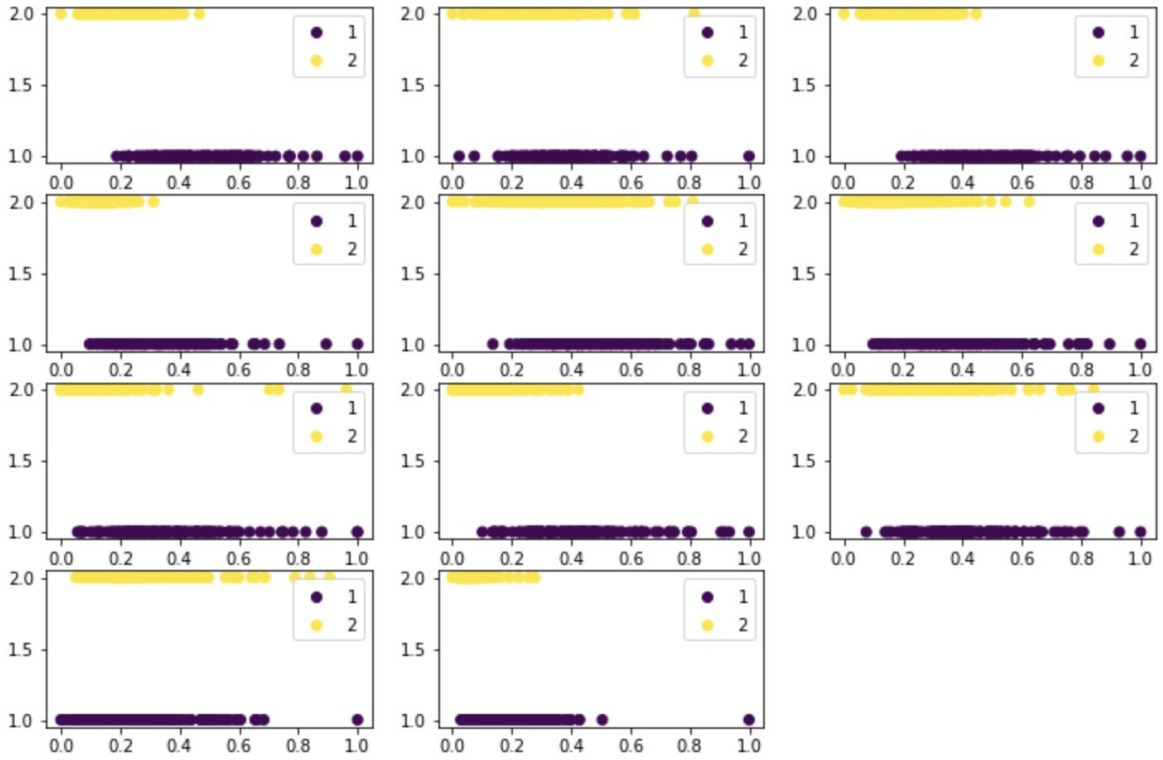
Group 1:

After observing the feature data against each class (using correlation matrix and plotting), here are our inferences:
Correlation matrix:

```
data.corr()
```

	ID	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	class
ID	1.000000	-0.054848	-0.064085	-0.070918	-0.048961	-0.289301	-0.263939	-0.210081	-0.218828	-0.229391	-0.209549	-0.108090	0.253638
f1	-0.054848	1.000000	0.419081	0.997498	0.988365	0.140992	0.497840	0.640096	0.799991	0.163715	-0.298526	0.679442	-0.718062
f2	-0.064085	0.419081	1.000000	0.426658	0.408253	0.062307	0.331583	0.360760	0.380248	0.134569	-0.029980	0.292952	-0.498645
f3	-0.070918	0.997498	0.426658	1.000000	0.986882	0.182424	0.553201	0.683526	0.832748	0.202813	-0.243096	0.695119	-0.731083
f4	-0.048961	0.988365	0.408253	0.986882	1.000000	0.147717	0.487429	0.647696	0.796726	0.160449	-0.271021	0.720414	-0.687940
f5	-0.289301	0.140992	0.062307	0.182424	0.147717	1.000000	0.648751	0.550790	0.571164	0.551946	0.585594	0.317299	-0.367730
f6	-0.263939	0.497840	0.331583	0.553201	0.487429	0.648751	1.000000	0.890362	0.844251	0.627954	0.580874	0.528627	-0.592944
f7	-0.210081	0.640096	0.360760	0.683526	0.647696	0.550790	0.890362	1.000000	0.913620	0.542710	0.406648	0.646363	-0.649691
f8	-0.218828	0.799991	0.380248	0.832748	0.796726	0.571164	0.844251	0.913620	1.000000	0.509739	0.212158	0.707365	-0.753920
f9	-0.229391	0.163715	0.134569	0.202813	0.160449	0.551946	0.627954	0.542710	0.509739	1.000000	0.490449	0.325885	-0.328024
f10	-0.209549	-0.298526	-0.029980	-0.243096	-0.271021	0.585594	0.580874	0.406648	0.212158	0.490449	1.000000	0.043404	-0.005336
f11	-0.108090	0.679442	0.292952	0.695119	0.720414	0.317299	0.528627	0.646363	0.707365	0.325885	0.043404	1.000000	-0.533952
class	0.253638	-0.718062	-0.498645	-0.731083	-0.687940	-0.367730	-0.592944	-0.649691	-0.753920	-0.328024	-0.005336	-0.533952	1.000000

Plots of features vs class:



Conclusions:

1. From the correlation matrix some features are highly correlated (f3,f4,f8,f6,f7)
2. Plotting features against each class, we found that some features are the most informative & can be used to deduce a threshold value and do a simple rule-based model
3. We used 2 features f1 & f3 and a threshold value of 0.3 & 0.5.
4. We got training accuracy of approximately 88%
5. We got test accuracy of approximately 90%