

```

1 Gaussian Naive Bayes Algorithm
2 Machine Language Programming assignment 2
3 Author: Larry Bilodeau
4
5
6 Data set: spambase.data
7 rawdata shape (4601, 58)
8 test data shape (2300, 58)
9 training data shape (2301, 57)
10
11 Hypothesis      Frequency      Ratio
12 0                1251          0.54
13
14 Prior Probability of each hypothesis
15 hyposis: spam,  P(h1): 0.45652173913043476
16 hyposis: not spam, P(h0): 0.5439130434782609
17 P(+|h1): 0.5434782608695652 P(-|h1): 0.45652173913043476
18 P(+|h0): 0.45608695652173914 P(-|h0): 0.5439130434782609
19
20 Hypothesis      Frequency      Ratio
21 1                1050          0.46
22
23 Prior Probability of each hypothesis
24 hyposis: spam,  P(h1): 0.45652173913043476
25 hyposis: not spam, P(h0): 0.5439130434782609
26 P(+|h1): 0.5434782608695652 P(-|h1): 0.45652173913043476
27 P(+|h0): 0.45608695652173914 P(-|h0): 0.5439130434782609
28
29 class: 0          feature: word_freq_make          mean: 0.0756      std_dev:
30 0.2998              variance: 0.0899
31 class: 0          feature: word_freq_address        mean: 0.2234      std_dev:
32 1.4983              variance: 2.2448
33 class: 0          feature: word_freq_all           mean: 0.1958      std_dev:
34 0.5168              variance: 0.2671
35 class: 0          feature: word_freq_3d           mean: 0.0008      std_dev:
36 0.0183              variance: 0.0003
37 class: 0          feature: word_freq_our          mean: 0.1648      std_dev:
38 0.5550              variance: 0.3080
39 class: 0          feature: word_freq_over         mean: 0.0441      std_dev:
40 0.1992              variance: 0.0397
41 class: 0          feature: word_freq_remove       mean: 0.0101      std_dev:
42 0.1122              variance: 0.0126
43 class: 0          feature: word_freq_internet     mean: 0.0442      std_dev:
44 0.2753              variance: 0.0758
45 class: 0          feature: word_freq_order        mean: 0.0330      std_dev:
46 0.1567              variance: 0.0246
47 class: 0          feature: word_freq_mail         mean: 0.1650      std_dev:
48 0.7054              variance: 0.4976
49 class: 0          feature: word_freq_receive      mean: 0.0275      std_dev:
50 0.1752              variance: 0.0307
51 class: 0          feature: word_freq_will         mean: 0.5413      std_dev:
52 0.9480              variance: 0.8987
53 class: 0          feature: word_freq_people       mean: 0.0668      std_dev:
54 0.2951              variance: 0.0871
55 class: 0          feature: word_freq_report       mean: 0.0404      std_dev:
56 0.2761              variance: 0.0762
57 class: 0          feature: word_freq_addresses    mean: 0.0095      std_dev:
58 0.1043              variance: 0.0109
59 class: 0          feature: word_freq_free         mean: 0.0780      std_dev:
60 0.6615              variance: 0.4375
61 class: 0          feature: word_freq_business     mean: 0.0537      std_dev:
62 0.2263              variance: 0.0512
63 class: 0          feature: word_freq_email        mean: 0.0963      std_dev:
64 0.3814              variance: 0.1455
65 class: 0          feature: word_freq_you          mean: 1.2464      std_dev:
66 1.7233              variance: 2.9699
67 class: 0          feature: word_freq_credit       mean: 0.0083      std_dev:
68 0.1001              variance: 0.0100
69 class: 0          feature: word_freq_your         mean: 0.4289      std_dev:

```

	1.0268	variance: 1.0542		
50	class: 0	feature: word_freq_font	mean: 0.0586	std_dev:
	0.6757	variance: 0.4566		
51	class: 0	feature: word_freq_000	mean: 0.0065	std_dev:
	0.0489	variance: 0.0024		
52	class: 0	feature: word_freq_money	mean: 0.0154	std_dev:
	0.2685	variance: 0.0721		
53	class: 0	feature: word_freq_hp	mean: 0.9776	std_dev:
	2.2224	variance: 4.9391		
54	class: 0	feature: word_freq_hpl	mean: 0.4506	std_dev:
	1.1798	variance: 1.3920		
55	class: 0	feature: word_freq_george	mean: 1.4046	std_dev:
	4.4583	variance: 19.8760		
56	class: 0	feature: word_freq_650	mean: 0.2260	std_dev:
	0.7081	variance: 0.5014		
57	class: 0	feature: word_freq_lab	mean: 0.1975	std_dev:
	0.8984	variance: 0.8070		
58	class: 0	feature: word_freq_labs	mean: 0.1799	std_dev:
	0.6069	variance: 0.3684		
59	class: 0	feature: word_freq_telnet	mean: 0.1221	std_dev:
	0.6004	variance: 0.3605		
60	class: 0	feature: word_freq_857	mean: 0.0925	std_dev:
	0.4682	variance: 0.2193		
61	class: 0	feature: word_freq_data	mean: 0.1766	std_dev:
	0.8569	variance: 0.7343		
62	class: 0	feature: word_freq_415	mean: 0.0927	std_dev:
	0.4685	variance: 0.2195		
63	class: 0	feature: word_freq_85	mean: 0.1745	std_dev:
	0.5950	variance: 0.3541		
64	class: 0	feature: word_freq_technology	mean: 0.1427	std_dev:
	0.4964	variance: 0.2465		
65	class: 0	feature: word_freq_1999	mean: 0.1863	std_dev:
	0.4888	variance: 0.2389		
66	class: 0	feature: word_freq_parts	mean: 0.0184	std_dev:
	0.2427	variance: 0.0589		
67	class: 0	feature: word_freq_pm	mean: 0.1034	std_dev:
	0.4989	variance: 0.2489		
68	class: 0	feature: word_freq_direct	mean: 0.0962	std_dev:
	0.4647	variance: 0.2160		
69	class: 0	feature: word_freq_cs	mean: 0.0837	std_dev:
	0.5432	variance: 0.2951		
70	class: 0	feature: word_freq_meeting	mean: 0.2323	std_dev:
	1.0313	variance: 1.0636		
71	class: 0	feature: word_freq_original	mean: 0.0676	std_dev:
	0.2688	variance: 0.0723		
72	class: 0	feature: word_freq_project	mean: 0.1458	std_dev:
	0.9437	variance: 0.8905		
73	class: 0	feature: word_freq_re	mean: 0.4472	std_dev:
	1.3308	variance: 1.7711		
74	class: 0	feature: word_freq_edu	mean: 0.2709	std_dev:
	1.1222	variance: 1.2592		
75	class: 0	feature: word_freq_table	mean: 0.0049	std_dev:
	0.0552	variance: 0.0031		
76	class: 0	feature: word_freq_conference	mean: 0.0460	std_dev:
	0.3557	variance: 0.1265		
77	class: 0	feature: char_freq_;	mean: 0.0598	std_dev:
	0.3610	variance: 0.1303		
78	class: 0	feature: char_freq_(mean: 0.1505	std_dev:
	0.2346	variance: 0.0550		
79	class: 0	feature: char_freq_[mean: 0.0297	std_dev:
	0.1841	variance: 0.0339		
80	class: 0	feature: char_freq_!	mean: 0.0908	std_dev:
	0.3924	variance: 0.1540		
81	class: 0	feature: char_freq_\$	mean: 0.0092	std_dev:
	0.0523	variance: 0.0027		
82	class: 0	feature: char_freq_#	mean: 0.0223	std_dev:
	0.2164	variance: 0.0468		
83	class: 0	feature: capital_run_length_average	mean: 2.4060	std_dev:
	2.4974	variance: 6.2371		

84	class: 0 29.5708	feature: capital_run_length_longest variance: 874.4319	mean: 18.1135	std_dev:
85	class: 0 332.4764	feature: capital_run_length_total variance: 110540.5569	mean: 161.3030	std_dev:
86	class: 1 0.2863	feature: word_freq_make variance: 0.0820	mean: 0.1502	std_dev:
87	class: 1 0.3101	feature: word_freq_address variance: 0.0962	mean: 0.1650	std_dev:
88	class: 1 0.4861	feature: word_freq_all variance: 0.2363	mean: 0.4149	std_dev:
89	class: 1 2.1899	feature: word_freq_3d variance: 4.7957	mean: 0.1673	std_dev:
90	class: 1 0.7176	feature: word_freq_our variance: 0.5150	mean: 0.5087	std_dev:
91	class: 1 0.3302	feature: word_freq_over variance: 0.1090	mean: 0.1848	std_dev:
92	class: 1 0.5543	feature: word_freq_remove variance: 0.3072	mean: 0.2643	std_dev:
93	class: 1 0.5659	feature: word_freq_internet variance: 0.3203	mean: 0.2035	std_dev:
94	class: 1 0.3453	feature: word_freq_order variance: 0.1192	mean: 0.1671	std_dev:
95	class: 1 0.6554	feature: word_freq_mail variance: 0.4296	mean: 0.3521	std_dev:
96	class: 1 0.2380	feature: word_freq_receive variance: 0.0566	mean: 0.1145	std_dev:
97	class: 1 0.6361	feature: word_freq_will variance: 0.4046	mean: 0.5568	std_dev:
98	class: 1 0.3516	feature: word_freq_people variance: 0.1236	mean: 0.1467	std_dev:
99	class: 1 0.3214	feature: word_freq_report variance: 0.1033	mean: 0.0837	std_dev:
100	class: 1 0.3671	feature: word_freq_addresses variance: 0.1347	mean: 0.1065	std_dev:
101	class: 1 0.9482	feature: word_freq_free variance: 0.8991	mean: 0.5210	std_dev:
102	class: 1 0.6425	feature: word_freq_business variance: 0.4129	mean: 0.2874	std_dev:
103	class: 1 0.6349	feature: word_freq_email variance: 0.4031	mean: 0.3111	std_dev:
104	class: 1 1.5354	feature: word_freq_you variance: 2.3575	mean: 2.2081	std_dev:
105	class: 1 0.8545	feature: word_freq_credit variance: 0.7302	mean: 0.2067	std_dev:
106	class: 1 1.2533	feature: word_freq_your variance: 1.5709	mean: 1.3896	std_dev:
107	class: 1 1.2716	feature: word_freq_font variance: 1.6169	mean: 0.2062	std_dev:
108	class: 1 0.4984	feature: word_freq_000 variance: 0.2484	mean: 0.2516	std_dev:
109	class: 1 0.5887	feature: word_freq_money variance: 0.3465	mean: 0.2065	std_dev:
110	class: 1 0.1535	feature: word_freq_hp variance: 0.0236	mean: 0.0172	std_dev:
111	class: 1 0.1189	feature: word_freq_hpl variance: 0.0141	mean: 0.0115	std_dev:
112	class: 1 0.0183	feature: word_freq_george variance: 0.0003	mean: 0.0013	std_dev:
113	class: 1 0.3865	feature: word_freq_650 variance: 0.1494	mean: 0.0226	std_dev:
114	class: 1 0.0062	feature: word_freq_lab variance: 0.0000	mean: 0.0004	std_dev:
115	class: 1 0.1337	feature: word_freq_labs variance: 0.0179	mean: 0.0091	std_dev:
116	class: 1 0.0210	feature: word_freq_telnet variance: 0.0004	mean: 0.0009	std_dev:
117	class: 1 0.0205	feature: word_freq_857 variance: 0.0004	mean: 0.0009	std_dev:
118	class: 1	feature: word_freq_data	mean: 0.0185	std_dev:

	0.1236	variance: 0.0153		
119	class: 1	feature: word_freq_415	mean: 0.0022	std_dev:
	0.0445	variance: 0.0020		
120	class: 1	feature: word_freq_85	mean: 0.0047	std_dev:
	0.0655	variance: 0.0043		
121	class: 1	feature: word_freq_technology	mean: 0.0314	std_dev:
	0.1564	variance: 0.0244		
122	class: 1	feature: word_freq_1999	mean: 0.0439	std_dev:
	0.2916	variance: 0.0850		
123	class: 1	feature: word_freq_parts	mean: 0.0045	std_dev:
	0.0384	variance: 0.0015		
124	class: 1	feature: word_freq_pm	mean: 0.0111	std_dev:
	0.0819	variance: 0.0067		
125	class: 1	feature: word_freq_direct	mean: 0.0357	std_dev:
	0.1448	variance: 0.0210		
126	class: 1	feature: word_freq_cs	mean: 0.0001	std_dev:
	0.0031	variance: 0.0000		
127	class: 1	feature: word_freq_meeting	mean: 0.0024	std_dev:
	0.0277	variance: 0.0008		
128	class: 1	feature: word_freq_original	mean: 0.0087	std_dev:
	0.0468	variance: 0.0022		
129	class: 1	feature: word_freq_project	mean: 0.0049	std_dev:
	0.0474	variance: 0.0022		
130	class: 1	feature: word_freq_re	mean: 0.1243	std_dev:
	0.3313	variance: 0.1098		
131	class: 1	feature: word_freq_edu	mean: 0.0118	std_dev:
	0.1191	variance: 0.0142		
132	class: 1	feature: word_freq_table	mean: 0.0012	std_dev:
	0.0190	variance: 0.0004		
133	class: 1	feature: word_freq_conference	mean: 0.0023	std_dev:
	0.0296	variance: 0.0009		
134	class: 1	feature: char_freq_;	mean: 0.0202	std_dev:
	0.0871	variance: 0.0076		
135	class: 1	feature: char_freq_()	mean: 0.1199	std_dev:
	0.3556	variance: 0.1264		
136	class: 1	feature: char_freq_[]	mean: 0.0082	std_dev:
	0.0523	variance: 0.0027		
137	class: 1	feature: char_freq_!	mean: 0.5005	std_dev:
	0.7233	variance: 0.5231		
138	class: 1	feature: char_freq_\$	mean: 0.1871	std_dev:
	0.4111	variance: 0.1690		
139	class: 1	feature: char_freq_#	mean: 0.0643	std_dev:
	0.4557	variance: 0.2076		
140	class: 1	feature: capital_run_length_average	mean: 9.0670	std_dev:
	37.5061	variance: 1406.7101		
141	class: 1	feature: capital_run_length_longest	mean: 109.0362	std_dev:
	352.8741	variance: 124520.1417		
142	class: 1	feature: capital_run_length_total	mean: 485.2552	std_dev:
	905.8244	variance: 820517.7994		

143

144 Variance Table

	class 0		class 1	
Feature	Spam_mean	Spam_variance	Not Spam_mean	Not Spam_variance
147 word_freq_make	-0.9244	0.0899	-0.8498	0.0820
148 word_freq_address	-0.7766	2.2448	-0.8350	0.0962
149 word_freq_all	-0.8042	0.2671	-0.5851	0.2363
150 word_freq_3d	-0.9992	0.0003	-0.8327	4.7957
151 word_freq_our	-0.8352	0.3080	-0.4913	0.5150
152 word_freq_over	-0.9559	0.0397	-0.8152	0.1090
153 word_freq_remove	-0.9899	0.0126	-0.7357	0.3072
154 word_freq_internet	-0.9558	0.0758	-0.7965	0.3203
155 word_freq_order	-0.9670	0.0246	-0.8329	0.1192
156 word_freq_mail	-0.8350	0.4976	-0.6479	0.4296
157 word_freq_receive	-0.9725	0.0307	-0.8855	0.0566
158 word_freq_will	-0.4587	0.8987	-0.4432	0.4046
159 word_freq_people	-0.9332	0.0871	-0.8533	0.1236
160 word_freq_report	-0.9596	0.0762	-0.9163	0.1033
161 word_freq_addresses	-0.9905	0.0109	-0.8935	0.1347
162 word_freq_free	-0.9220	0.4375	-0.4790	0.8991

163	word_freq_business	-0.9463	0.0512	-0.7126	0.4129
164	word_freq_email	-0.9037	0.1455	-0.6889	0.4031
165	word_freq_you	0.2464	2.9699	1.2081	2.3575
166	word_freq_credit	-0.9917	0.0100	-0.7933	0.7302
167	word_freq_your	-0.5711	1.0542	0.3896	1.5709
168	word_freq_font	-0.9414	0.4566	-0.7938	1.6169
169	word_freq_000	-0.9935	0.0024	-0.7484	0.2484
170	word_freq_money	-0.9846	0.0721	-0.7935	0.3465
171	word_freq_hp	-0.0224	4.9391	-0.9828	0.0236
172	word_freq_hpl	-0.5494	1.3920	-0.9885	0.0141
173	word_freq_george	0.4046	19.8760	-0.9987	0.0003
174	word_freq_650	-0.7740	0.5014	-0.9774	0.1494
175	word_freq_lab	-0.8025	0.8070	-0.9996	0.0000
176	word_freq_labs	-0.8201	0.3684	-0.9909	0.0179
177	word_freq_telnet	-0.8779	0.3605	-0.9991	0.0004
178	word_freq_857	-0.9075	0.2193	-0.9991	0.0004
179	word_freq_data	-0.8234	0.7343	-0.9815	0.0153
180	word_freq_415	-0.9073	0.2195	-0.9978	0.0020
181	word_freq_85	-0.8255	0.3541	-0.9953	0.0043
182	word_freq_technology	-0.8573	0.2465	-0.9686	0.0244
183	word_freq_1999	-0.8137	0.2389	-0.9561	0.0850
184	word_freq_parts	-0.9816	0.0589	-0.9955	0.0015
185	word_freq_pm	-0.8966	0.2489	-0.9889	0.0067
186	word_freq_direct	-0.9038	0.2160	-0.9643	0.0210
187	word_freq_cs	-0.9163	0.2951	-0.9999	0.0000
188	word_freq_meeting	-0.7677	1.0636	-0.9976	0.0008
189	word_freq_original	-0.9324	0.0723	-0.9913	0.0022
190	word_freq_project	-0.8542	0.8905	-0.9951	0.0022
191	word_freq_re	-0.5528	1.7711	-0.8757	0.1098
192	word_freq_edu	-0.7291	1.2592	-0.9882	0.0142
193	word_freq_table	-0.9951	0.0031	-0.9988	0.0004
194	word_freq_conference	-0.9540	0.1265	-0.9977	0.0009
195	char_freq_;	-0.9402	0.1303	-0.9798	0.0076
196	char_freq_(-0.8495	0.0550	-0.8801	0.1264
197	char_freq_[-0.9703	0.0339	-0.9918	0.0027
198	char_freq_!	-0.9092	0.1540	-0.4995	0.5231
199	char_freq_\$	-0.9908	0.0027	-0.8129	0.1690
200	char_freq_#	-0.9777	0.0468	-0.9357	0.2076
201	capital_run_length_average	1.4060	6.2371	8.0670	1406.7101
202	capital_run_length_longest	17.1135	874.4319	108.0362	124520.1417

203
204
205 Training_data class probabilities

206	class 0	class 1
207	-74.6436	17.4100
208	-607.7075	-0.4713
209	-25.1317	17.1130
210	-36.9822	8.5877
211	-42.1584	9.5037
212	-785.0648	-15.8611
213	-32.3877	17.8796
214	-74.2296	-13.7194
215	-18.8898	18.1364
216	-23.5606	15.7583

217
218 Test_data class probabilities

219	class 0	class 1
220	-22.3112	-602.3293
221	-48.7308	-61.2579
222	-18.8834	13.5184
223	-61.7625	-inf
224	-inf	-149.4920
225	-137.6091	13.3769
226	-35.3379	-inf
227	-43.5377	-inf
228	-42.0118	-inf
229	-34.6401	16.1318

230
231

```
232 Accuracy      Precision    Recall
233 0.7952        0.9646      0.3317
234
235
236 Confusion Matrix
237 [[1093.  444.]
238  [  27.  736.]]
239
```