Kmeans and Kmeans++: Comparison and deployment

Eduardo A. Castillo Perera

Instituto Tecnológico de Estudios Superiores de Monterrey

Professor: Benjamín Valdés Aguirre

Table of Contents

Abstract

Machine learning is without a doubt revolutionizing our world. Each day new discoveries are being made and there are a lot of techniques and algorithms. Usually, ML techniques are divided into **supervised learning** & **unsupervised learning.** Unsupervised learning techniques do not require labeling and are useful to train and group batches of data for further analysis and decision taking. In this document I describe the implementation of the KMeans algorithm, and the deployment of an example of such algorithm.

Keywords: AI, Clustering, ML, Unsupervised learning, Web, Cloud, unit testing.

Kmeans and Kmeans++: Comparison and deployment

[The body of your paper uses a half-inch first line indent and is double-spaced. APA style provides for up to five heading levels, shown in the paragraphs that follow. Note that the word Introduction should not be used as an initial heading, as it's assumed that your paper begins with an introduction.]

KMeans Clustering and its variant

K Means clustering is an **unsupervised learning technique** to group samples of data without requiring labeling. It takes a cluster of data points and attempts to cluster them by computing centroids, and from those centroids, it creates.

Implementation & deployment[1]

The code was implemented in C#, using 2 classes: Row and cluster. The github folders contain more information about their contents, including useful texts.

Annex: Results from running the KMEans classic and KMeans++

ID	sepal_length	sepal_width	petal_length	petal_width	Label
1	5.1	3.5	1.4	0.2	2
2	4.9	3	1.4	0.2	2
3	4.7	3.2	1.3	0.2	2
4	4.6	3.1	1.5	0.2	2
5	5	3.6	1.4	0.2	2
6	5.4	3.9	1.7	0.4	2
7	4.6	3.4	1.4	0.3	2
8	5	3.4	1.5	0.2	2
9	4.4	2.9	1.4	0.2	2

10	4.9	3.1	1.5	0.1	2
11	5.4	3.7	1.5	0.2	2
12	4.8	3.4	1.6	0.2	2
13	4.8	3	1.4	0.1	2
14	4.3	3	1.1	0.1	2
15	5.8	4	1.2	0.2	2
16	5.7	4.4	1.5	0.4	2
17	5.4	3.9	1.3	0.4	2
18	5.1	3.5	1.4	0.3	2
19	5.7	3.8	1.7	0.3	2
20	5.1	3.8	1.5	0.3	2
21	5.4	3.4	1.7	0.2	2
22	5.1	3.7	1.5	0.4	2
23	4.6	3.6	1	0.2	2
24	5.1	3.3	1.7	0.5	2
25	4.8	3.4	1.9	0.2	2
26	5	3	1.6	0.2	2
27	5	3.4	1.6	0.4	2
28	5.2	3.5	1.5	0.2	2
29	5.2	3.4	1.4	0.2	2
30	4.7	3.2	1.6	0.2	2
31	4.8	3.1	1.6	0.2	2
32	5.4	3.4	1.5	0.4	2
33	5.2	4.1	1.5	0.1	2
34	5.5	4.2	1.4	0.2	2
35	4.9	3.1	1.5	0.1	2
36	5	3.2	1.2	0.2	2
37	5.5	3.5	1.3	0.2	2
38	4.9	3.1	1.5	0.1	2
39	4.4	3	1.3	0.2	2
40	5.1	3.4	1.5	0.2	2

41	5	3.5	1.3	0.3	2
42	4.5	2.3	1.3	0.3	2
43	4.4	3.2	1.3	0.2	2
44	5	3.5	1.6	0.6	2
45	5.1	3.8	1.9	0.4	2
46	4.8	3	1.4	0.3	2
47	5.1	3.8	1.6	0.2	2
48	4.6	3.2	1.4	0.2	2
49	5.3	3.7	1.5	0.2	2
50	5	3.3	1.4	0.2	2
51	7	3.2	4.7	1.4	1
52	6.4	3.2	4.5	1.5	0
53	6.9	3.1	4.9	1.5	1
54	5.5	2.3	4	1.3	0
55	6.5	2.8	4.6	1.5	0
56	5.7	2.8	4.5	1.3	0
57	6.3	3.3	4.7	1.6	0
58	4.9	2.4	3.3	1	0
59	6.6	2.9	4.6	1.3	0
60	5.2	2.7	3.9	1.4	0
61	5	2	3.5	1	0
62	5.9	3	4.2	1.5	0
63	6	2.2	4	1	0
64	6.1	2.9	4.7	1.4	0
65	5.6	2.9	3.6	1.3	0
66	6.7	3.1	4.4	1.4	0
67	5.6	3	4.5	1.5	0
68	5.8	2.7	4.1	1	0
69	6.2	2.2	4.5	1.5	0
70	5.6	2.5	3.9	1.1	0
71	5.9	3.2	4.8	1.8	0
72	6.1	2.8	4	1.3	0
73	6.3	2.5	4.9	1.5	0
74	6.1	2.8	4.7	1.2	0
75	6.4	2.9	4.3	1.3	0
76	6.6	3	4.4	1.4	0
77	6.8	2.8	4.8	1.4	0
78	6.7	3	5	1.7	1
79	6	2.9	4.5	1.5	0
80	5.7	2.6	3.5	1	0
81	5.5	2.4	3.8	1.1	0
82	5.5	2.4	3.7	1	0
83	5.8	2.7	3.9	1.2	0

84	6	2.7	5.1	1.6	0
85	5.4	3	4.5	1.5	0
86	6	3.4	4.5	1.6	0
87	6.7	3.1	4.7	1.5	0
88	6.3	2.3	4.4	1.3	0
89	5.6	3	4.1	1.3	0
90	5.5	2.5	4	1.3	0
91	5.5	2.6	4.4	1.2	0
92	6.1	3	4.6	1.4	0
93	5.8	2.6	4	1.2	0
94	5	2.3	3.3	1	0
95	5.6	2.7	4.2	1.3	0
96	5.7	3	4.2	1.2	0
97	5.7	2.9	4.2	1.3	0
98	6.2	2.9	4.3	1.3	0
99	5.1	2.5	3	1.1	0
100	5.7	2.8	4.1	1.3	0
101	6.3	3.3	6	2.5	1
102	5.8	2.7	5.1	1.9	0
103	7.1	3	5.9	2.1	1
104	6.3	2.9	5.6	1.8	1
105	6.5	3	5.8	2.2	1
106	7.6	3	6.6	2.1	1
107	4.9	2.5	4.5	1.7	0
108	7.3	2.9	6.3	1.8	1
109	6.7	2.5	5.8	1.8	1
110	7.2	3.6	6.1	2.5	1
111	6.5	3.2	5.1	2	1
112	6.4	2.7	5.3	1.9	1
113	6.8	3	5.5	2.1	1
114	5.7	2.5	5	2	0
115	5.8	2.8	5.1	2.4	0
116	6.4	3.2	5.3	2.3	1
117	6.5	3	5.5	1.8	1
118	7.7	3.8	6.7	2.2	1
119	7.7	2.6	6.9	2.3	1
120	6	2.2	5	1.5	0
121	6.9	3.2	5.7	2.3	1
122	5.6	2.8	4.9	2	0
123	7.7	2.8	6.7	2	1
124	6.3	2.7	4.9	1.8	0
125	6.7	3.3	5.7	2.1	1
126	7.2	3.2	6	1.8	1

127	6.2	2.8	4.8	1.8	0
128	6.1	3	4.9	1.8	0
129	6.4	2.8	5.6	2.1	1
130	7.2	3	5.8	1.6	1
131	7.4	2.8	6.1	1.9	1
132	7.9	3.8	6.4	2	1
133	6.4	2.8	5.6	2.2	1
134	6.3	2.8	5.1	1.5	0
135	6.1	2.6	5.6	1.4	1
136	7.7	3	6.1	2.3	1
137	6.3	3.4	5.6	2.4	1
138	6.4	3.1	5.5	1.8	1
139	6	3	4.8	1.8	0
140	6.9	3.1	5.4	2.1	1
141	6.7	3.1	5.6	2.4	1
142	6.9	3.1	5.1	2.3	1
143	5.8	2.7	5.1	1.9	0
144	6.8	3.2	5.9	2.3	1
145	6.7	3.3	5.7	2.5	1
146	6.7	3	5.2	2.3	1
147	6.3	2.5	5	1.9	0
148	6.5	3	5.2	2	1
149	6.2	3.4	5.4	2.3	1
150	5.9	3	5.1	1.8	0

ID	sepal_length	sepal_width	petal_length	petal_width	Label
1	5.1	3.5	1.4	0.2	2
2	4.9	3	1.4	0.2	2
3	4.7	3.2	1.3	0.2	2
4	4.6	3.1	1.5	0.2	2
5	5	3.6	1.4	0.2	2
6	5.4	3.9	1.7	0.4	2

7	4.6	3.4	1.4	0.3	2
8	5	3.4	1.5	0.2	2
9	4.4	2.9	1.4	0.2	2
10	4.9	3.1	1.5	0.1	2
11	5.4	3.7	1.5	0.2	2
12	4.8	3.4	1.6	0.2	2
13	4.8	3	1.4	0.1	2
14	4.3	3	1.1	0.1	2
15	5.8	4	1.2	0.2	2
16	5.7	4.4	1.5	0.4	2
17	5.4	3.9	1.3	0.4	2
18	5.1	3.5	1.4	0.3	2

19	5.7	3.8	1.7	0.3	2
20	5.1	3.8	1.5	0.3	2
21	5.4	3.4	1.7	0.2	2
22	5.1	3.7	1.5	0.4	2
23	4.6	3.6	1	0.2	2
24	5.1	3.3	1.7	0.5	2
25	4.8	3.4	1.9	0.2	2
26	5	3	1.6	0.2	2
27	5	3.4	1.6	0.4	2
28	5.2	3.5	1.5	0.2	2
29	5.2	3.4	1.4	0.2	2

	3	4.7	3.2	1.6	0.2	2
0						
	3	4.8	3.1	1.6	0.2	2
1						
	3	5.4	3.4	1.5	0.4	2
2						
	3	5.2	4.1	1.5	0.1	2
3						
	3	5.5	4.2	1.4	0.2	2
4						
	3	4.9	3.1	1.5	0.1	2
5						
	3	5	3.2	1.2	0.2	2
6						
	3	5.5	3.5	1.3	0.2	2
7						
	3	4.9	3.1	1.5	0.1	2
8						
	3	4.4	3	1.3	0.2	2
9						
	4	5.1	3.4	1.5	0.2	2
0						

	4	5	3.5	1.3	0.3	2
1						
	4	4.5	2.3	1.3	0.3	2
2						
	4	4.4	3.2	1.3	0.2	2
3						
	4	5	3.5	1.6	0.6	2
4						
	4	5.1	3.8	1.9	0.4	2
5						
	4	4.8	3	1.4	0.3	2
6						
	4	5.1	3.8	1.6	0.2	2
7						
	4	4.6	3.2	1.4	0.2	2
8						
	4	5.3	3.7	1.5	0.2	2
9						
	5	5	3.3	1.4	0.2	2
0						
	5	7	3.2	4.7	1.4	1
1						

	5	6.4	3.2	4.5	1.5	0
2						
	5	6.9	3.1	4.9	1.5	1
3						
	5	5.5	2.3	4	1.3	0
4						
	5	6.5	2.8	4.6	1.5	0
5						
	5	5.7	2.8	4.5	1.3	0
6						
	5	6.3	3.3	4.7	1.6	0
7						
	5	4.9	2.4	3.3	1	0
8						
	5	6.6	2.9	4.6	1.3	0
9						
	6	5.2	2.7	3.9	1.4	0
0						
	6	5	2	3.5	1	0
1						
	6	5.9	3	4.2	1.5	0
2						

63	6	2.2	4	1	0
64	6.1	2.9	4.7	1.4	0
65	5.6	2.9	3.6	1.3	0
66	6.7	3.1	4.4	1.4	0
67	5.6	3	4.5	1.5	0
68	5.8	2.7	4.1	1	0
69	6.2	2.2	4.5	1.5	0
70	5.6	2.5	3.9	1.1	0
71	5.9	3.2	4.8	1.8	0
72	6.1	2.8	4	1.3	0
73	6.3	2.5	4.9	1.5	0

74	6.1	2.8	4.7	1.2	0
75	6.4	2.9	4.3	1.3	0
76	6.6	3	4.4	1.4	0
77	6.8	2.8	4.8	1.4	0
78	6.7	3	5	1.7	1
79	6	2.9	4.5	1.5	0
80	5.7	2.6	3.5	1	0
81	5.5	2.4	3.8	1.1	0
82	5.5	2.4	3.7	1	0
83	5.8	2.7	3.9	1.2	0
84	6	2.7	5.1	1.6	0

	8	5.4	3	4.5	1.5	0
5						
	8	6	3.4	4.5	1.6	0
6						
	8	6.7	3.1	4.7	1.5	0
7						
	8	6.3	2.3	4.4	1.3	0
8						
	8	5.6	3	4.1	1.3	0
9						
	9	5.5	2.5	4	1.3	0
0						
	9	5.5	2.6	4.4	1.2	0
1						
	9	6.1	3	4.6	1.4	0
2						
	9	5.8	2.6	4	1.2	0
3						
	9	5	2.3	3.3	1	0
4						
	9	5.6	2.7	4.2	1.3	0
5						

96	5.7	3	4.2	1.2	0
97	5.7	2.9	4.2	1.3	0
98	6.2	2.9	4.3	1.3	0
99	5.1	2.5	3	1.1	0
100	5.7	2.8	4.1	1.3	0
101	6.3	3.3	6	2.5	1
102	5.8	2.7	5.1	1.9	0
103	7.1	3	5.9	2.1	1
104	6.3	2.9	5.6	1.8	1
105	6.5	3	5.8	2.2	1
106	7.6	3	6.6	2.1	1

107	4.9	2.5	4.5	1.7	0
108	7.3	2.9	6.3	1.8	1
109	6.7	2.5	5.8	1.8	1
110	7.2	3.6	6.1	2.5	1
111	6.5	3.2	5.1	2	1
112	6.4	2.7	5.3	1.9	1
113	6.8	3	5.5	2.1	1
114	5.7	2.5	5	2	0
115	5.8	2.8	5.1	2.4	0
116	6.4	3.2	5.3	2.3	1
117	6.5	3	5.5	1.8	1

	1	7.7	3.8	6.7	2.2	1
18						
	1	7.7	2.6	6.9	2.3	1
19						
	1	6	2.2	5	1.5	0
20						
	1	6.9	3.2	5.7	2.3	1
21						
	1	5.6	2.8	4.9	2	0
22						
	1	7.7	2.8	6.7	2	1
23						
	1	6.3	2.7	4.9	1.8	0
24						
	1	6.7	3.3	5.7	2.1	1
25						
	1	7.2	3.2	6	1.8	1
26						
	1	6.2	2.8	4.8	1.8	0
27						
	1	6.1	3	4.9	1.8	0
28						

	1	6.4	2.8	5.6	2.1	1
29						
	1	7.2	3	5.8	1.6	1
30						
	1	7.4	2.8	6.1	1.9	1
31						
	1	7.9	3.8	6.4	2	1
32						
	1	6.4	2.8	5.6	2.2	1
33						
	1	6.3	2.8	5.1	1.5	0
34						
	1	6.1	2.6	5.6	1.4	1
35						
	1	7.7	3	6.1	2.3	1
36						
	1	6.3	3.4	5.6	2.4	1
37						
	1	6.4	3.1	5.5	1.8	1
38						
	1	6	3	4.8	1.8	0
39						

40	1	6.9	3.1	5.4	2.1	1
41	1	6.7	3.1	5.6	2.4	1
42	1	6.9	3.1	5.1	2.3	1
43	1	5.8	2.7	5.1	1.9	0
44	1	6.8	3.2	5.9	2.3	1
45	1	6.7	3.3	5.7	2.5	1
46	1	6.7	3	5.2	2.3	1
47	1	6.3	2.5	5	1.9	0
48	1	6.5	3	5.2	2	1
49	1	6.2	3.4	5.4	2.3	1
50	1	5.9	3	5.1	1.8	0