

# Olympic Medal SVM

Yijia Dai

March 3, 2022

We have created a matrix whose rows are indexed by Olympic teams and whose columns are indexed by sports. The  $(i, j)$  entry of the matrix counts the combined number of medals (gold, silver, and bronze) that team  $i$  won in sport  $j$  from 1980 to 2016.

An important step of preprocessing is to remove the centroid of the data vectors. The SVM is seeking a best-fit linear subspace, as we proved in question 3, SVM could help to find the best-fit subspace when the centroid is . After removing the mean for the vectors, we also normalized each column. It is done under the assumption that the sports are relatively independent from each other for each country. We make this decision with tradeoff between the dependency of different sports and the clearance of the countries ranking.

Then, we compute the singular value decomposition of this matrix. For the top four singular values, we take the associated left and right singular vectors  $u$  and  $v$ , and generate tables ranking Olympic teams  $i$  in ascending order of  $u_i$  and ranking sports  $j$  in ascending order of  $v_j$ .

Figure 1 shows the 38 singular values. For the convenience of visualizing, two zero values are patched to form a  $(8 \times 5)$  table. The actual singular values are stored in a  $(38 \times 51)$  matrix.

	0	1	2	3	4
0	0.886364	0.678023	0.553199	0.513609	0.507386
1	0.465595	0.436586	0.412766	0.348220	0.331071
2	0.290833	0.262573	0.258286	0.239348	0.236724
3	0.221460	0.209522	0.185786	0.171585	0.154596
4	0.147084	0.131733	0.122548	0.118610	0.104547
5	0.098577	0.089517	0.081266	0.069348	0.059773
6	0.052158	0.051287	0.041817	0.034885	0.029742
7	0.018495	0.016170	0.000000	0.000000	0.000000

Figure 1: Table for all singular values.

The tables are shown in Figure 2-5 have the first columns listing team names or sports names, the second column listing the corresponding components of the singular vector in ascending order.

Interesting patterns are shown within the ranking of teams and sports. For example, the ranking of the teams corresponding to the first left singular vector has the precedence of "The United States", "China", "Russia", "Canada", "Australia", "Great Britain", "Germany"... which is an "accurate" approximate indicator of the ranking of countries performance in recent decades. We can see the absolute value of the vector's dimensions are in a wide range with the U.S. about 0.75. It means that the characteristic of these teams are very prominent in this direction. The sports with absolute value large are reasonably the sports like Swimming that the top-ranking teams are good at.

At the second singular vector, China plays a dominant role with a dimension value of 0.78. And the corresponding sports, Badminton, Table Tennis, and Weighting, are the sports that China is good at. Similarly, there exists a strong correlation between Spain and Beach Volleyball on the third singular vector.

Moreover, the third right singular vector shows positive components are predominantly summer sports and whose negative components are predominantly winter sports. And many European and high-latitude teams are listed in negative components correspondingly.

		Sport	
Team	Vector		Vector
0	United States	0	Swimming
1	China	1	Figure Skating
2	Russia	2	Athletics
3	Canada	3	Basketball
4	Australia	4	Diving
5	Great Britain	5	Softball
6	Germany	6	Snowboarding
7	South Korea	7	Shooting
8	Japan	8	Freestyle Skiing
9	France	9	Short Track Speed Skating
10	Soviet Union	10	Skeleton
11	Italy	11	Gymnastics
12	Cuba	12	Synchronized Swimming
13	East Germany	13	Archery
14	Netherlands	14	Volleyball
15	Spain	15	Taekwondo
16	Switzerland	16	Boxing
17	Sweden	17	Trampoline
18	Brazil	18	Wrestling
19	Romania	19	Tennis
20	Unified Team	20	Equestrianism
21	Bulgaria	21	Table Tennis
22	Ukraine	22	Badminton
23	Hungary	23	Water Polo
24	New Zealand	24	Golf
25	Poland	25	Sailing
26	West Germany	26	Baseball
27	Norway	27	Cycling
28	Belarus	28	Rowing
29	Denmark	29	Weightlifting
30	Finland	30	Ice Hockey
31	Yugoslavia	31	Speed Skating
32	Austria	32	Football
33	Argentina	33	Judo
34	Croatia	34	Fencing
35	Czech Republic	35	Triathlon
36	Jamaica	36	Rhythmic Gymnastics
37	Czechoslovakia	37	Curling
		38	Rugby Sevens
		39	Modern Pentathlon
		40	Bobsleigh
		41	Alpine Skiing
		42	Luge
		43	Biathlon
		44	Hockey
		45	Canoeing
		46	Nordic Combined
		47	Handball
		48	Cross Country Skiing
		49	Beach Volleyball
		50	Ski Jumping

Figure 2: Table for the teams and sports corresponding to sorted 1st singular vectors.

		Sport	
Team	Vector	Sport	Vector
0	United States	Swimming	-0.184702
1	Australia	Snowboarding	-0.171256
2	Great Britain	Equestrianism	-0.167908
3	Switzerland	Tennis	-0.160743
4	France	Athletics	-0.155628
5	Germany	Sailing	-0.151126
6	Spain	Triathlon	-0.138781
7	Austria	Basketball	-0.134870
8	Netherlands	Cycling	-0.129323
9	Canada	Skeleton	-0.126566
10	Norway	Alpine Skiing	-0.111916
11	New Zealand	Bobsleigh	-0.103086
12	Sweden	Rowing	-0.098650
13	Finland	Water Polo	-0.094537
14	Brazil	Rugby Sevens	-0.093688
15	Argentina	Ice Hockey	-0.087297
16	East Germany	Softball	-0.085472
17	West Germany	Figure Skating	-0.083657
18	Czech Republic	Nordic Combined	-0.075953
19	Soviet Union	Football	-0.072224
20	Jamaica	Beach Volleyball	-0.062523
21	Czechoslovakia	Hockey	-0.062523
22	Croatia	Freestyle Skiing	-0.061617
23	Hungary	Canoeing	-0.058563
24	Yugoslavia	Speed Skating	-0.055602
25	Italy	Curling	-0.051311
26	Poland	Luge	-0.042242
27	Unified Team	Ski Jumping	-0.041652
28	Cuba	Modern Pentathlon	-0.031873
29	Romania	Biathlon	-0.029564
30	Belarus	Baseball	-0.021879
31	Ukraine	Cross Country Skiing	-0.019991
32	Denmark	Boxing	-0.007017
33	Bulgaria	Wrestling	0.022981
34	Japan	Fencing	0.023039
35	Russia	Volleyball	0.023443
36	South Korea	Handball	0.043312
37	China	Golf	0.062732
		Judo	0.077187
		Synchronized Swimming	0.081530
		Gymnastics	0.089246
		Rhythmic Gymnastics	0.089408
		Taekwondo	0.110958
		Short Track Speed Skating	0.134833
		Archery	0.148430
		Shooting	0.176081
		Diving	0.242945
		Trampoline	0.267266
		Weightlifting	0.292731
		Table Tennis	0.383267
		Badminton	0.414862

Figure 3: Table for the teams and sports corresponding to sorted 2nd singular vectors.

		Sport	
Team	Vector		Vector
0	Germany	0	Biathlon -0.193827
1	Austria	1	Luge -0.188395
2	Norway	2	Nordic Combined -0.176426
3	Canada	3	Ski Jumping -0.163661
4	France	4	Cross Country Skiing -0.133744
5	Finland	5	Alpine Skiing -0.130516
6	Switzerland	6	Curling -0.113260
7	Great Britain	7	Canoeing -0.112995
8	Sweden	8	Bobsleigh -0.091653
9	Italy	9	Speed Skating -0.090281
10	East Germany	10	Equestrianism -0.081732
11	Soviet Union	11	Rowing -0.078802
12	Russia	12	Skeleton -0.074771
13	Hungary	13	Fencing -0.074205
14	Poland	14	Trampolineing -0.072674
15	West Germany	15	Snowboarding -0.071611
16	Netherlands	16	Freestyle Skiing -0.054976
17	China	17	Ice Hockey -0.054365
18	Romania	18	Modern Pentathlon -0.048963
19	New Zealand	19	Shooting -0.045762
20	Czech Republic	20	Cycling -0.042011
21	Ukraine	21	Table Tennis -0.041000
22	Czechoslovakia	22	Rugby Sevens -0.040449
23	Belarus	23	Diving -0.028654
24	Unified Team	24	Triathlon -0.027229
25	Bulgaria	25	Weightlifting -0.019785
26	Jamaica	26	Figure Skating -0.014305
27	Denmark	27	Judo -0.007410
28	Japan	28	Short Track Speed Skating 0.000401
29	Yugoslavia	29	Boxing 0.002014
30	Croatia	30	Wrestling 0.008546
31	Brazil	31	Gymnastics 0.008730
32	Cuba	32	Golf 0.009444
33	Australia	33	Football 0.015260
34	Argentina	34	Athletics 0.017088
35	South Korea	35	Swimming 0.022056
36	United States	36	Volleyball 0.026652
37	Spain	37	Badminton 0.027710
		38	Hockey 0.031001
		39	Handball 0.046028
		40	Archery 0.053470
		41	Softball 0.060187
		42	Baseball 0.066892
		43	Synchronized Swimming 0.079250
		44	Rhythmic Gymnastics 0.080379
		45	Sailing 0.083429
		46	Taekwondo 0.101700
		47	Water Polo 0.104364
		48	Tennis 0.149940
		49	Basketball 0.170168
		50	Beach Volleyball 0.794137

Figure 4: Table for the teams and sports corresponding to sorted 3rd singular vectors.

		Sport	
Team	Vector	Sport	Vector
0	Great Britain	0	Rugby Sevens
1	Canada	1	Bobsleigh
2	Australia	2	Triathlon
3	New Zealand	3	Curling
4	China	4	Golf
5	Switzerland	5	Sailing
6	Denmark	6	Badminton
7	South Korea	7	Cycling
8	Netherlands	8	Trampoline
9	Sweden	9	Short Track Speed Skating
10	France	10	Rowing
11	Argentina	11	Hockey
12	Jamaica	12	Skeleton
13	Czech Republic	13	Table Tennis
14	Czechoslovakia	14	Freestyle Skiing
15	Croatia	15	Softball
16	Ukraine	16	Diving
17	West Germany	17	Taekwondo
18	Yugoslavia	18	Equestrianism
19	Romania	19	Ice Hockey
20	Brazil	20	Archery
21	Poland	21	Swimming
22	Unified Team	22	Tennis
23	Belarus	23	Baseball
24	Finland	24	Judo
25	Cuba	25	Snowboarding
26	East Germany	26	Athletics
27	Japan	27	Speed Skating
28	Bulgaria	28	Modern Pentathlon
29	Spain	29	Figure Skating
30	Hungary	30	Weightlifting
31	United States	31	Handball
32	Austria	32	Basketball
33	Norway	33	Beach Volleyball
34	Soviet Union	34	Shooting
35	Italy	35	Synchronized Swimming
36	Germany	36	Alpine Skiing
37	Russia	37	Gymnastics
		38	Football
		39	Boxing
		40	Canoeing
		41	Volleyball
		42	Ski Jumping
		43	Water Polo
		44	Nordic Combined
		45	Fencing
		46	Wrestling
		47	Cross Country Skiing
		48	Rhythmic Gymnastics
		49	Luge
		50	Biathlon

Figure 5: Table for the teams and sports corresponding to sorted 4th singular vectors.