CS 5350/6350: Machine Learning Fall 2021

Homework 3

Handed out: 19 Oct, 2021
Due date: 11:59pm, 2 Nov, 2021

- You are welcome to talk to other members of the class about the homework. I am more concerned that you understand the underlying concepts. However, you should write down your own solution. Please keep the class collaboration policy in mind.
- Feel free to discuss the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. You do not need to include original problem descriptions in your solutions. You need to show your work, not just the final answer, but you do *not* need to write it in gory detail. Your assignment should be **no more than 15 pages**. Every extra page will cost a point.
- Handwritten solutions will not be accepted.
- *Your code should run on the CADE machines.* **You should include a shell script, `run.sh`, that will execute your code in the CADE environment. Your code should produce similar output to what you include in your report.**

  You are responsible for ensuring that the grader can execute the code using only the included script. If you are using an esoteric programming language, you should make sure that its runtime is available on CADE.
- Please do not hand in binary files! We will *not* grade binary submissions.
- The homework is due by **midnight of the due date**. Please submit the homework on Canvas.

1 Paper Problems [36 points + 15 bonus]

1. [8 points] hyperplane: $2x_1 + 3x_2 - 4 = 0$

   (a) [4 points]

   Yes, the hyperplane does have a margin for the dataset. We can look at the formula, $\text{dist}(x_0) = \frac{|\mathbf{w}^\top \mathbf{x}_0 + b|}{||\mathbf{w}||}$, to compute the distance for each data point and simply take the minimum of those values to find the margin of the hyperplane. In this equation $\mathbf{w} = (2, 3)$, $b = -4$, and $\mathbf{x}_0 = (x_1, x_2)$ of the row which we are calculating the distance of.

   Distance of each data point in order: $0.27, 1.39, 1.11, 0.83$. We can clearly see that the margin of this hyperplane for dataset 2 is 0.27.

(b) [4 points]

No, the hyperplane does not have a margin for the dataset. This is because with the addition of the fifth row, the given hyperplane does not linearly seperate the labels anymore. This is the case in which the margin does not exist for the hyperplane.

2. [8 points]

(a) [4 points]

Yes, we can calculate the margin of the dataset. This data is a perfect square centered at the origin and rotated 90 degrees around the origin. This means that the margin is when the hyperplane goes through the center of the data and due to the shape of the data this is the line $y = x$. Thus with $\mathbf{w} = (1, 1)$ and $b = 0$ we know that the distance of every point is $\gamma = 0.707$ (using formula from above question). Using any of $\mathbf{w}$ and b will result in a smaller gamma value thus we know that we have found the margin for the dataset.

(b) [4 points]

No, we cannot calculate the margin for this dataset. This data is not linearly separable by any hyperplane without performing feature mapping, thus we know that the there is no margin.

3. [**Bonus**] [5 points]

We can look at the information presented in the Preceptron lecture in order to solve this. If $\mathbf{u}$ is not a unit vector, then we scale γ in the mistake bound. This will change the final mistake bound to $(\frac{||\mathbf{u}||R}{\gamma})^2$

4. [10 points]

$$f(x_1, x_2, \ldots, x_n) = \neg x_1 \lor \neg \ldots \neg x_k \lor x_{k+1} \lor \ldots \lor x_{2k} \quad \text{(note that } 2k < n).$$

We will solve this by following the steps in the Perceptron lecture, but modifying it for this question. We will first start by finding R. We can find this by doing $R = \sqrt{1 + 2k}$ as 2k is the number of entries. From here we will find a separating hyperplane with a non-zero margin.

$$-x_1 + -x_2 + \ldots + -x_k + x_{k+1} + \ldots + x_{2k} \geq 1 - k$$

$$-x_1 + -x_2 + \ldots + -x_k + x_{k+1} + \ldots + x_{2k} - \frac{1}{2} + k = 0$$

We have now found a hyperplane with a non-zero margin. We guarntee that the hyperplane has a non-zero margin by changing the 1 to a 1/2. This will result in no points lying on the boundary since a combination of 0's and 1's can never be 1/2. From here we can find $\mathbf{u}$ and $\mathbf{x}$ in augmented space which will give us the equation for γ.

$$\mathbf{u} = \frac{1}{\sqrt{(k + 1/2)^2}}[-1, ..., -1_k, 1_{k+1}, ..., -1/2 + k]$$

$$\mathbf{x} = [x_1, ..., x_2, 1]$$

Now we can finally construct our γ. We know that the minimum margin is 1/2. From here plug that into the equation $(\frac{R}{\gamma})^2$ giving us our upper mistake bound.

$$\gamma = \frac{1/2}{k+1/2} = \frac{1}{2k} + 1$$

$$\text{mistake upper bound } = \left(\frac{\sqrt{1+2k}}{\frac{1}{2k}+1}\right)^2$$

5. [10 points] Prove that linear classifiers in a plane cannot shatter any 4 distinct points.

   In order to prove this we will use Radon's theorem. Using this theorem we know that any set of $d+2$ points in $\mathbf{R}^d$ can we partitioned into two sets whose convex hulls intersect. From here we know $d = 2$ this is because there when using linear classifiers there are two possible labels. This is very useful as we now know that given any 4 distinct points on a plane we can partition those points into two convex hulls that must intersect. Now let us discuss convex hulls. A convex hull is defined to be the shape that can contain smallest convex set (a shape containing a line between two points). This means that when we have intersecting convex hulls we know that a line cannot separate the convex hulls. This means that there is no linear classifier that will be able to shatter the 4 distinct points. Thus we have proven that linear classifiers in a plane cannot shatter any 4 distinct points.

6. [**Bonus**] [10 points]

   Through information presented in lecture we know that a rectangle can be used to shatter any four points and using "the VC dimension of hypothesis space H over instance space X is the size of the largest finite subset of X that is shattered by H" we know that VC($\mathcal{H}$) = 4.

2 Practice [64 points]

1. [2 Points]

   `https://github.com/ericodonoghue/machine_learning_uofu`

2. (a) [16 points]

   $\mathbf{w}$: $[12.0, -15.818, -9.956, -12.491, -2.707]$

   Average Prediction Error: 0.02

   (b) [16 points]

   Average Prediction Error: 0.012

c	**w**				
0	0.000	0.000	0.000	0.000	0.000
0	-0.250	-0.962	-2.538	0.964	1.056

2	-0.500	-0.950	-2.138	-1.155	0.867
1	-0.250	-1.267	-1.433	-1.761	0.395
8	0.000	-1.780	-0.473	-1.960	0.092
3	0.250	-2.597	-3.659	1.929	0.056
7	0.500	-3.577	-2.640	1.870	-0.472
5	0.250	-4.444	-1.622	0.798	-0.858
9	0.000	-4.415	-1.228	-1.210	-0.851
10	0.250	-4.052	-0.326	-2.224	-1.250
2	0.500	-5.221	-1.742	0.519	-1.334
1	0.250	-4.559	-4.276	0.851	0.034
3	0.500	-4.530	-3.471	-0.006	-0.677
1	0.750	-5.177	-2.504	-0.090	-0.997
8	1.000	-5.147	-2.448	-0.208	-0.755
2	1.250	-5.270	-1.737	-1.119	-1.530
10	1.500	-4.721	-0.600	-2.363	-2.211
2	1.750	-5.726	-2.676	0.776	-2.589
3	2.000	-5.699	-2.182	-0.066	-2.752
14	2.250	-5.624	-2.139	-0.504	-2.630
6	2.500	-5.474	-1.655	-1.326	-2.711
11	2.750	-5.210	-1.359	-1.987	-2.683
17	2.500	-5.223	-3.122	-1.473	-1.895
27	2.250	-5.509	-2.187	-2.868	-1.736
30	2.000	-4.847	-4.721	-2.535	-0.369
3	2.250	-4.605	-3.760	-3.768	-1.402
12	2.000	-4.687	-2.647	-4.911	-1.155
6	2.250	-5.115	-3.841	-3.358	-1.055
7	2.500	-4.975	-3.919	-3.312	-0.944
31	2.250	-5.105	-3.103	-4.085	-0.697
13	2.500	-5.890	-6.362	-0.165	-0.863
2	2.250	-6.322	-5.245	-2.221	-0.411
1	2.500	-8.082	-2.945	-2.156	-1.582
30	2.750	-7.662	-1.893	-3.291	-2.180
53	3.000	-7.324	-1.628	-3.877	-2.080
15	3.250	-8.201	-4.770	-0.087	-2.268
24	3.500	-7.768	-3.781	-1.272	-2.893
18	3.250	-8.223	-2.162	-3.285	-2.789
17	3.000	-8.026	-4.554	-2.338	-0.913
5	3.250	-7.649	-4.064	-3.103	-0.944
18	3.500	-7.396	-3.839	-3.690	-0.837
104	3.750	-6.798	-2.699	-4.938	-1.561
21	4.000	-7.700	-4.193	-2.415	-1.769
27	3.750	-7.660	-2.577	-4.504	-1.388
27	4.000	-8.584	-5.997	-0.109	-2.043
12	4.250	-8.360	-4.803	-1.320	-3.440
1	4.000	-8.490	-3.988	-2.092	-3.194

113	3.750	-8.572	-2.874	-3.235	-2.947
1	4.000	-8.064	-2.411	-3.988	-2.946
10	3.750	-7.973	-4.483	-3.508	-2.113
28	4.000	-7.469	-4.034	-4.247	-2.060
48	4.250	-7.164	-3.509	-5.046	-2.028
7	4.000	-6.967	-5.901	-4.100	-0.153
29	4.250	-8.551	-3.579	-4.096	-1.849
24	4.000	-7.889	-6.114	-3.763	-0.481
28	4.250	-7.527	-5.212	-4.777	-0.880
27	4.500	-6.978	-4.074	-6.021	-1.561
2	4.750	-7.984	-6.150	-2.882	-1.939
23	5.000	-7.834	-5.667	-3.705	-2.020
11	5.250	-7.570	-5.371	-4.365	-1.992
11	5.500	-7.198	-4.514	-5.373	-2.349
78	5.250	-7.280	-3.400	-6.516	-2.102
6	5.500	-7.707	-4.594	-4.963	-2.002
38	5.250	-7.837	-3.779	-5.735	-1.756
13	5.500	-8.623	-7.038	-1.816	-1.921
2	5.250	-9.054	-5.920	-3.871	-1.470
6	5.000	-9.287	-4.971	-5.032	-1.396
25	5.250	-8.867	-3.919	-6.167	-1.994
20	5.500	-9.674	-5.723	-3.256	-2.231
23	5.250	-9.513	-4.571	-5.343	-1.553
10	5.500	-9.175	-4.306	-5.929	-1.453
74	5.250	-8.979	-6.698	-4.982	0.423
5	5.500	-8.602	-6.208	-5.747	0.392
58	5.750	-8.431	-4.995	-7.050	-1.134
26	6.000	-9.065	-6.735	-4.849	-0.752
38	6.250	-8.468	-5.596	-6.096	-1.476
53	6.000	-8.597	-4.780	-6.868	-1.230
22	6.250	-9.521	-8.200	-2.474	-1.885
12	6.500	-9.298	-7.006	-3.684	-3.282
1	6.250	-9.428	-6.190	-4.457	-3.036
113	6.000	-9.510	-5.076	-5.600	-2.789
1	6.250	-9.002	-4.613	-6.353	-2.788
38	6.500	-8.498	-4.164	-7.092	-2.736
49	6.750	-8.678	-5.854	-5.632	-2.580
114	7.000	-8.129	-4.716	-6.876	-3.261
2	7.250	-9.135	-6.792	-3.737	-3.639
34	7.500	-8.871	-6.496	-4.397	-3.611
11	7.750	-8.499	-5.638	-5.405	-3.968
63	7.500	-7.837	-8.173	-5.072	-2.600
15	7.250	-7.919	-7.059	-6.215	-2.353
44	7.000	-8.049	-6.243	-6.988	-2.106
49	6.750	-8.179	-5.427	-7.760	-1.860

17	7.000	-8.986	-7.231	-4.849	-2.097
23	6.750	-8.825	-6.079	-6.936	-1.419
211	7.000	-8.227	-4.940	-8.183	-2.144
11	7.250	-9.166	-7.013	-5.607	-2.049
37	7.000	-9.126	-5.397	-7.697	-1.668
27	7.250	-10.050	-8.817	-3.302	-2.323
12	7.500	-9.826	-7.623	-4.513	-3.721
1	7.250	-9.956	-6.807	-5.285	-3.474
113	7.000	-10.039	-5.694	-6.428	-3.227
1	7.250	-9.531	-5.231	-7.181	-3.226
38	7.500	-9.026	-4.781	-7.921	-3.174
49	7.750	-9.207	-6.471	-6.460	-3.018
6	7.500	-9.010	-8.862	-5.514	-1.142
29	7.750	-10.594	-6.541	-5.510	-2.838
52	8.000	-10.231	-5.639	-6.524	-3.237
3	7.750	-9.569	-8.174	-6.191	-1.870
24	8.000	-9.021	-7.036	-7.435	-2.551
125	7.750	-9.103	-5.922	-8.578	-2.304
6	8.000	-9.531	-7.117	-7.026	-2.205
38	7.750	-9.661	-6.301	-7.798	-1.958
13	8.000	-10.446	-9.560	-3.879	-2.124
2	7.750	-10.878	-8.443	-5.934	-1.672
6	7.500	-11.110	-7.493	-7.095	-1.598
25	7.750	-10.690	-6.442	-8.230	-2.196
20	8.000	-11.498	-8.245	-5.319	-2.433
23	7.750	-11.337	-7.094	-7.406	-1.755
49	8.000	-10.903	-6.105	-8.591	-2.381
35	7.750	-10.707	-8.497	-7.644	-0.505
5	8.000	-10.330	-8.007	-8.409	-0.536
122	8.250	-9.732	-6.867	-9.656	-1.260
21	8.500	-10.633	-8.361	-7.133	-1.467
27	8.250	-10.593	-6.745	-9.223	-1.087
27	8.500	-11.517	-10.165	-4.828	-1.741
12	8.750	-11.293	-8.971	-6.038	-3.139
1	8.500	-11.423	-8.156	-6.811	-2.893
113	8.250	-11.505	-7.042	-7.954	-2.646
1	8.500	-10.997	-6.579	-8.707	-2.645
38	8.750	-10.493	-6.129	-9.446	-2.593
52	9.000	-10.953	-8.401	-7.136	-2.619
84	9.250	-10.590	-7.500	-8.150	-3.018
152	9.000	-10.673	-6.386	-9.293	-2.771
6	9.250	-11.100	-7.580	-7.740	-2.671
38	9.000	-11.230	-6.765	-8.512	-2.425
49	8.750	-11.360	-5.949	-9.285	-2.179
17	9.000	-12.168	-7.752	-6.374	-2.415

23	8.750	-12.006	-6.601	-8.461	-1.738
84	8.500	-11.810	-8.992	-7.514	0.138
5	8.750	-11.433	-8.502	-8.279	0.107
82	9.000	-11.248	-7.289	-9.478	-1.309
38	9.250	-12.343	-8.668	-6.744	-1.411
2	9.500	-11.745	-7.529	-7.991	-2.136
53	9.250	-11.875	-6.713	-8.763	-1.890
22	9.500	-12.799	-10.133	-4.368	-2.544
12	9.750	-12.575	-8.939	-5.579	-3.942
1	9.500	-12.705	-8.123	-6.351	-3.696
113	9.250	-12.787	-7.010	-7.494	-3.448
1	9.500	-12.279	-6.547	-8.247	-3.448
38	9.750	-11.775	-6.097	-8.987	-3.395
55	9.500	-11.578	-8.489	-8.040	-1.519
81	9.750	-11.216	-7.587	-9.054	-1.918
61	10.000	-12.055	-9.397	-6.194	-2.061
13	10.250	-11.683	-8.540	-7.201	-2.418
78	10.000	-11.765	-7.426	-8.344	-2.170
44	9.750	-11.895	-6.610	-9.117	-1.924
13	10.000	-12.680	-9.869	-5.197	-2.090
8	9.750	-12.913	-8.920	-6.358	-2.016
25	10.000	-12.493	-7.868	-7.493	-2.614
3	9.750	-12.623	-7.053	-8.265	-2.368
124	9.500	-12.426	-9.444	-7.319	-0.492
5	9.750	-12.049	-8.954	-8.083	-0.523
82	10.000	-11.865	-7.741	-9.283	-1.939
40	10.250	-11.267	-6.602	-10.530	-2.664
11	10.500	-12.206	-8.675	-7.954	-2.569
37	10.250	-12.166	-7.059	-10.044	-2.188
27	10.500	-13.090	-10.479	-5.649	-2.843
12	10.750	-12.866	-9.285	-6.860	-4.240
1	10.500	-12.996	-8.470	-7.632	-3.994
113	10.250	-13.078	-7.356	-8.775	-3.747
1	10.500	-12.571	-6.893	-9.528	-3.746
93	10.250	-12.374	-9.284	-8.581	-1.870
81	10.500	-12.011	-8.383	-9.595	-2.270
27	10.750	-11.463	-7.245	-10.839	-2.951
2	11.000	-12.468	-9.321	-7.700	-3.328
45	11.250	-12.096	-8.464	-8.708	-3.685
78	11.000	-12.178	-7.350	-9.851	-3.438
44	10.750	-12.308	-6.534	-10.624	-3.191
13	11.000	-13.094	-9.793	-6.704	-3.357
8	10.750	-13.326	-8.844	-7.865	-3.283
25	11.000	-12.906	-7.792	-9.000	-3.881
3	10.750	-13.036	-6.977	-9.772	-3.635

17	11.000	-13.843	-8.780	-6.861	-3.871
23	10.750	-13.682	-7.628	-8.948	-3.194
84	10.500	-13.486	-10.020	-8.002	-1.318
5	10.750	-13.109	-9.530	-8.766	-1.349
122	11.000	-12.511	-8.391	-10.013	-2.073
53	10.750	-12.641	-7.575	-10.786	-1.827
22	11.000	-13.565	-10.995	-6.391	-2.482
12	11.250	-13.341	-9.801	-7.602	-3.879
1	11.000	-13.471	-8.985	-8.374	-3.633
113	10.750	-13.553	-7.872	-9.517	-3.386
1	11.000	-13.045	-7.409	-10.270	-3.385
93	10.750	-12.849	-9.800	-9.323	-1.509
81	11.000	-12.486	-8.898	-10.337	-1.909
27	11.250	-11.937	-7.761	-11.581	-2.590
2	11.500	-12.943	-9.837	-8.442	-2.967
45	11.750	-12.570	-8.980	-9.450	-3.324
78	11.500	-12.653	-7.866	-10.593	-3.077
44	11.250	-12.783	-7.050	-11.365	-2.830
13	11.500	-13.568	-10.309	-7.446	-2.996
8	11.250	-13.801	-9.360	-8.607	-2.922
25	11.500	-13.381	-8.308	-9.742	-3.520
3	11.250	-13.511	-7.492	-10.514	-3.274
17	11.500	-14.318	-9.296	-7.603	-3.511
23	11.250	-14.157	-8.144	-9.690	-2.833
84	11.000	-13.960	-10.536	-8.743	-0.957
5	11.250	-13.583	-10.046	-9.508	-0.988
122	11.500	-12.985	-8.907	-10.755	-1.712
32	11.750	-13.757	-10.566	-8.120	-1.935
16	11.500	-13.717	-8.950	-10.209	-1.555
5	11.250	-13.847	-8.134	-10.982	-1.309
22	11.500	-14.771	-11.554	-6.587	-1.963
12	11.750	-14.547	-10.360	-7.798	-3.361
1	11.500	-14.677	-9.545	-8.570	-3.115
113	11.250	-14.759	-8.431	-9.713	-2.868
1	11.500	-14.251	-7.968	-10.466	-2.867
38	11.750	-13.747	-7.518	-11.206	-2.814
55	11.500	-13.550	-9.910	-10.259	-0.939
81	11.750	-13.188	-9.008	-11.273	-1.338
34	12.000	-14.143	-12.272	-7.033	-1.914
1	12.250	-15.513	-10.226	-6.964	-3.172
39	12.500	-15.141	-9.369	-7.971	-3.529
78	12.250	-15.223	-8.255	-9.114	-3.281
44	12.000	-15.353	-7.440	-9.887	-3.035
49	11.750	-15.483	-6.624	-10.659	-2.789
17	12.000	-16.290	-8.427	-7.748	-3.025

23	11.750	-16.129	-7.276	-9.835	-2.348
84	11.500	-15.932	-9.667	-8.888	-0.472
5	11.750	-15.555	-9.177	-9.653	-0.503
122	12.000	-14.957	-8.038	-10.900	-1.227
32	12.250	-15.729	-9.697	-8.265	-1.450
16	12.000	-15.689	-8.082	-10.354	-1.070
154	12.250	-15.181	-7.619	-11.107	-1.069
15	12.500	-16.185	-9.697	-7.994	-1.429
23	12.750	-15.681	-9.247	-8.733	-1.376
136	13.000	-15.318	-8.345	-9.747	-1.775
27	13.250	-14.770	-7.208	-10.991	-2.457
2	13.500	-15.775	-9.284	-7.852	-2.834
45	13.750	-15.403	-8.427	-8.860	-3.191
78	13.500	-15.485	-7.313	-10.003	-2.943
44	13.250	-15.615	-6.497	-10.776	-2.697
13	13.500	-16.401	-9.756	-6.856	-2.863
8	13.250	-16.633	-8.807	-8.017	-2.789
25	13.500	-16.213	-7.755	-9.152	-3.387
3	13.250	-16.343	-6.939	-9.924	-3.141
124	13.000	-16.146	-9.331	-8.978	-1.265
5	13.250	-15.769	-8.841	-9.742	-1.295
122	13.500	-15.171	-7.702	-10.989	-2.020
32	13.750	-15.943	-9.361	-8.354	-2.243
16	13.500	-15.903	-7.745	-10.444	-1.863
5	13.250	-16.033	-6.930	-11.216	-1.616
22	13.500	-16.957	-10.349	-6.821	-2.271
12	13.750	-16.733	-9.156	-8.032	-3.669
1	13.500	-16.863	-8.340	-8.804	-3.422
31	13.250	-16.398	-10.311	-8.388	-2.963
82	13.000	-16.481	-9.197	-9.531	-2.716
1	13.250	-15.973	-8.734	-10.284	-2.715
38	13.500	-15.468	-8.285	-11.024	-2.662
55	13.250	-15.272	-10.676	-10.077	-0.787

Table 1: 2.2.b Distinct Weight Vectors

(c) [16 points]

$\mathbf{w}$: $[83558.0, -103819.466, -68753.926, -72841.582, -22108.965]$

Average Prediction Error: 0.014

Looking at the list of weight vectors from (b) we can seem some interesting relationships. If we look at the table above and for each column of every $\mathbf{w}$ we sum $c * w[i]$ where i is the current column we get a weight vector similar to the one produced through our average Perceptron.

$\mathbf{w}_{sum} = [78904.500, -100303.191, -61792.710, -68339.875, -18623.140]$. We can conclude that this is intended behaviour as average Perceptron is essentially an aggregate of all the data in the above table.

(d) [14 points]

Looking at the average prediction error for each variation of Perceptron we can see that we got very close and very low prediction errors with standard having an error of 0.02, voted having an error of 0.012, and average having an error of 0.014. Through this we can conclude that if the goal is strictly the lowest possible prediction error the best Perceptron variation would be voted although average has a very close prediction error being only 0.002 worse. As mentioned in the lecture average Perceptron is the most popular. I believe this is because it can get a very similar prediction error to voted, but it does not take up nearly as much space. Looking at voted we can see that for only 837 training examples over 250 distinct weight vectors were generated, on training data of much larger size such as 10,000-50,000 examples the space complexity of vote would be very poor. Thus as the lecture states average Perceptron is the most popular due to the above reasons.