

Machine Learning in Critical Care

Lecture 2

A Quick Note about the Course

- **Session 1**
 - The Nature of Medical Data
 - Data Extraction via SQL
- **Session 2**
 - Algorithm Construction and Application
- **Session 3**
 - Feature Engineering
- **Session 4**
 - Apply your knowledge, win a prize, start a paper!
 - Doing the homework will help with this session!

A Quick Note about the Course

- **Session 1**
 - The Nature of Medical Data
 - Data Extraction via SQL
- **Session 2**
 - Algorithm Construction and Application
- **Session 3**
 - Feature Engineering
- **Session 4**
 - Apply your knowledge, win a prize, start a paper!
 - Doing the homework will help with this session!

Session 2 - Learning Objectives

- Part 1: Build an intuition for how machine learning works
 - Data Pre-processing
 - Feature Selection
 - Model Construction

- Part 2: Apply a machine learning to a simple discriminative task
 - Model Validation

Your Homework For This Week

- **Problem 1:** Hand-craft an algorithm for predicting mortality in the patient population
- **Problem 2:** Apply 3 machine learning techniques to predict patient mortality

Session 2, Part 1:

An Introduction to Human Learning

Download Resources:

<http://bit.ly/1Rk0UnD>

Overview of Matlab IDE

Navigation and toolbars for the MATLAB environment, including options for File, Navigate, Edit, Breakpoints, Run, and Advance.

home > mohammad > Dropbox (MIT) > Machine_Learning_In_Critical_Care

Current Folder

Name	Size	Date M...
makeQuery.m	1 KB	01/04/201...
ConnectMIMIC.m	1 KB	01/04/201...
.dropbox	1 KB	01/04/201...
query.sql	1 KB	01/04/201...
Machine_Learning_I...	1 KB	01/05/201...
mlcc1-problem-set...	6 KB	01/11/201...
MLCCData.mat	215 ...	01/11/201...
MLCCDataMini.mat	6 KB	01/11/201...
MLCC_Lesson1.m	1 KB	01/11/201...
mysql-connector-ja...	960 ...	01/10/201...
postgresql-9.4.120...	593 ...	01/04/201...

```
Editor - /home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/MLCC_Lesson1.m
+11 -----
1 %% MIMIC CONNECTION in Matlab
2 %% Machine Learning in Critical Care.
3 %% Lesson #1: Connecting to the database.
4 %% Mohamad M. Ghassemi
5
6 %% STEP 1: Tell Matlab where the driver is
7 javaclasspath('/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.
8
9 %% STEP 2: Connect to the Database
10 conn = ConnectMIMIC;
11
12
13 %% Step 3: Open your web-browser and Let's look at the schema
14 %% https://mi.mic.physionet.org/about/mi.mic/
15
16 %% Step 4: Let's write out a query.
17 %%create a file query.sql in your folder.
18 query = makeQuery('query.sql');
19
20 %% Step 5: Run your query and get the results.
21 results = fetch(conn,query)
22
23
24 %% Step 6: Do some additional Analysis on the results
25
26
27
28
29
30
31
```

Workspace

Name	Value	Size	Class
conn	Database connection object	...	DatabaseConnection
query	SQL query string	...	String
results	Query results	...	Table

Command History

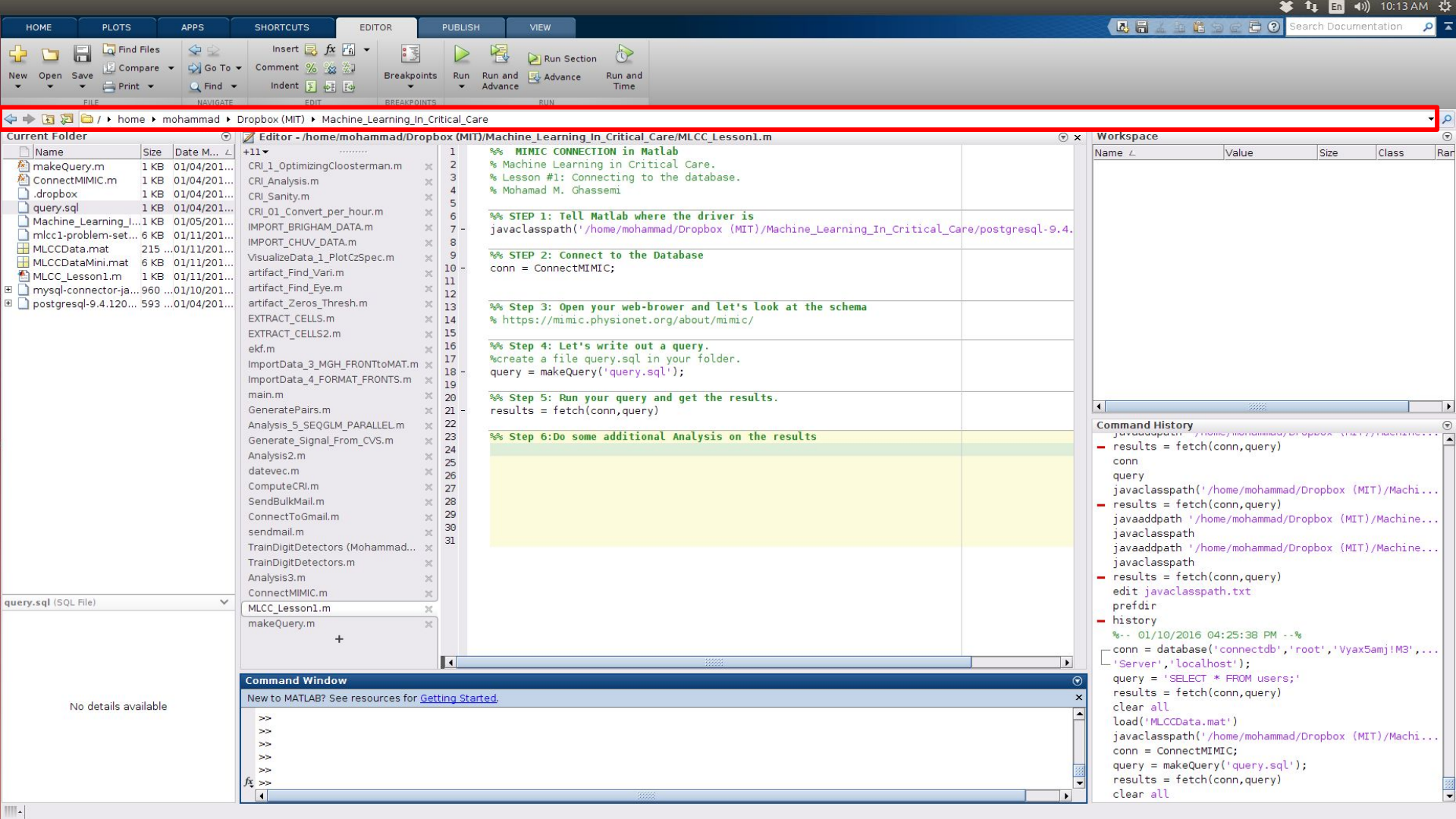
```
javaaddpath('/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.120...
- results = fetch(conn,query)
conn
query
javaclasspath('/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.120...
- results = fetch(conn,query)
javaaddpath '/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.120...
javaclasspath
javaaddpath '/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.120...
javaclasspath
- results = fetch(conn,query)
edit javaclasspath.txt
prefdir
- history
%- 01/10/2016 04:25:38 PM --%
conn = database('connectdb','root','VyaxSamj!M3',...
['Server','localhost'];
query = 'SELECT * FROM users;
results = fetch(conn,query)
clear all
load('MLCCData.mat')
javaclasspath('/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.120...
conn = ConnectMIMIC;
query = makeQuery('query.sql');
results = fetch(conn,query)
clear all
```

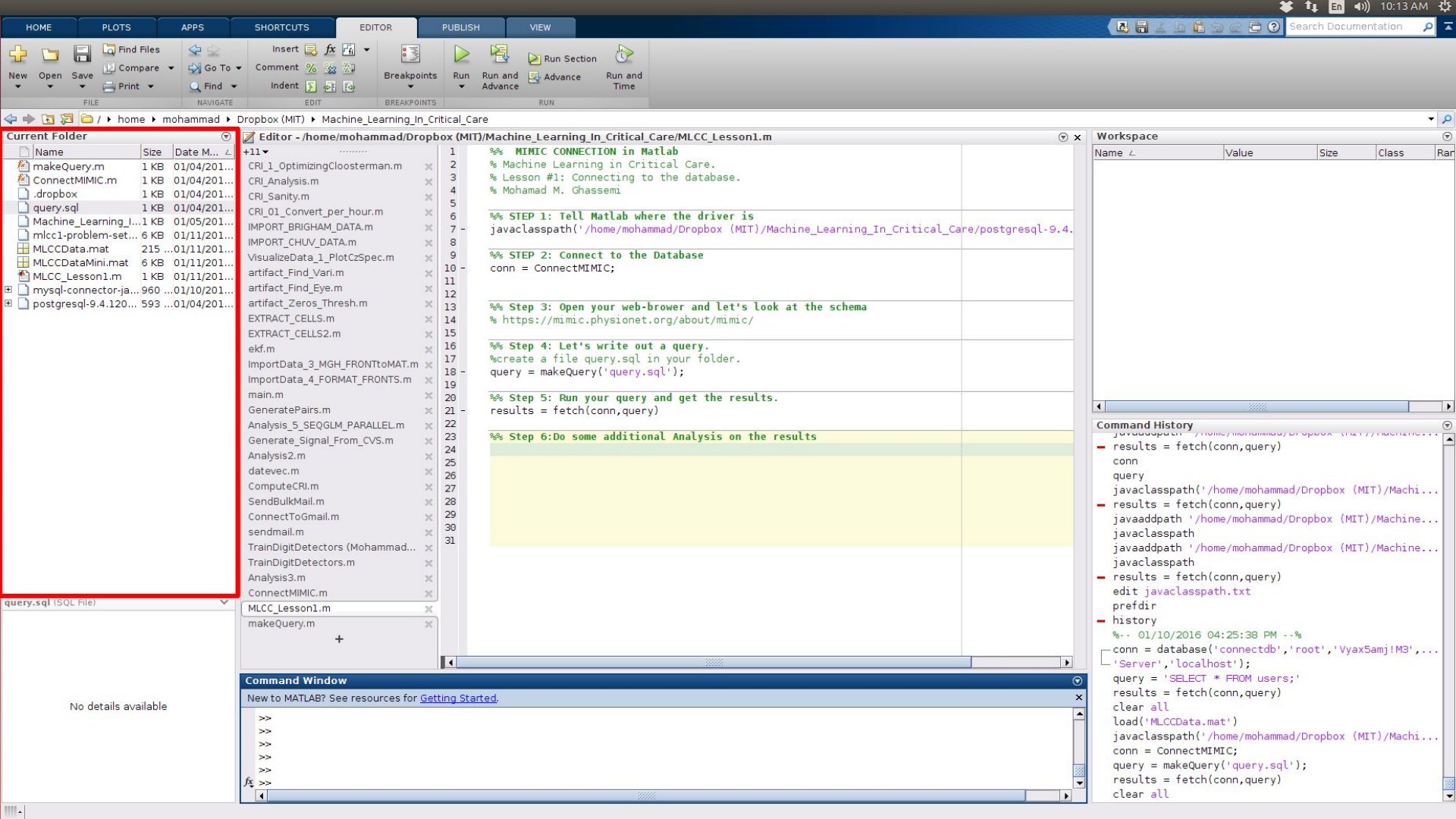
Command Window

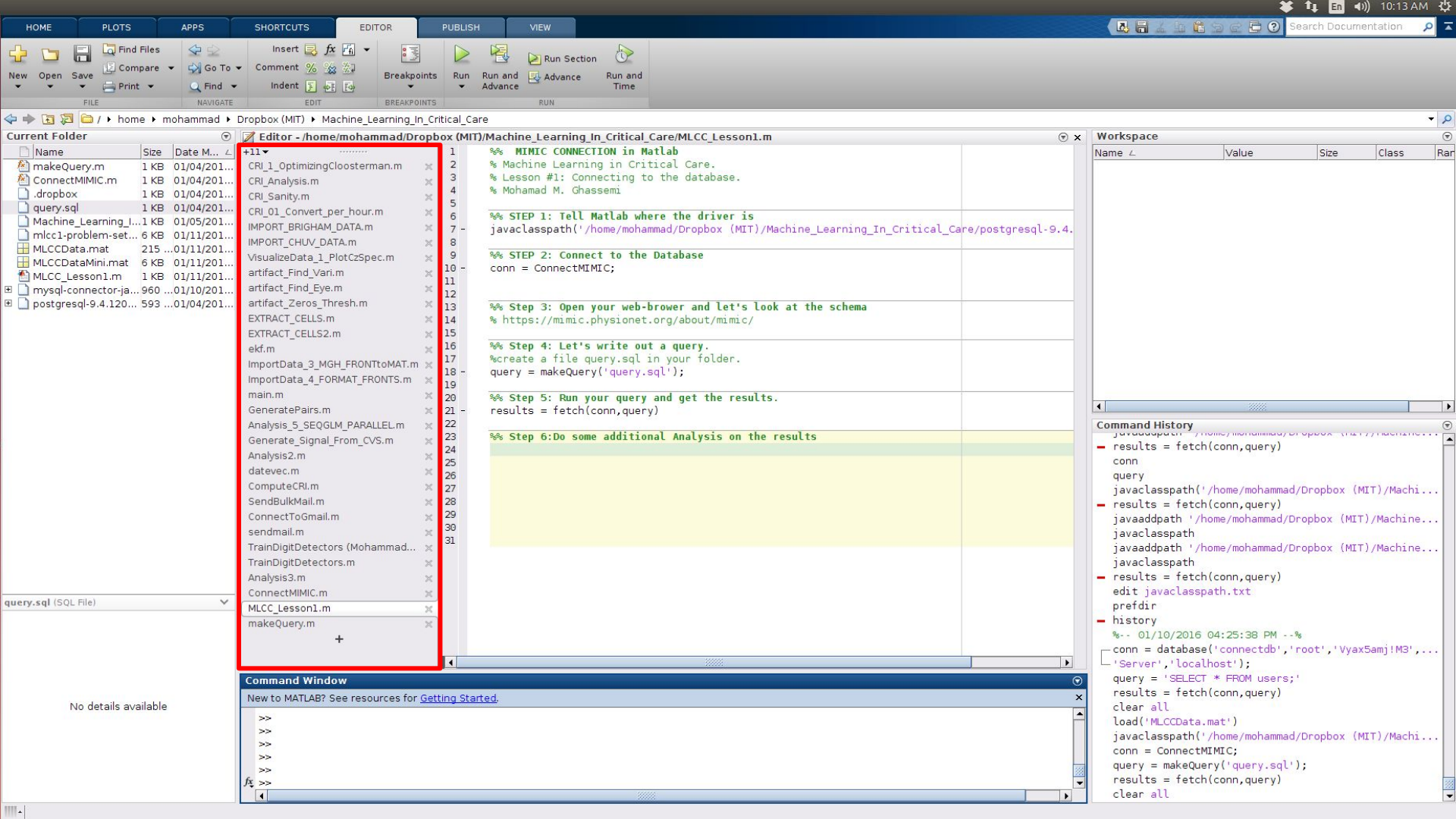
```
New to MATLAB? See resources for Getting Started.
>>
>>
>>
>>
>>
fx >>
```

query.sql (SQL File)

No details available







New Open Save Find Files Compare Go To Print Find Insert Comment Indent Breakpoints Run Run and Advance Advance Run Section Run and Time

home | mohammad | Dropbox (MIT) | Machine_Learning_In_Critical_Care

Current Folder

Name	Size	Date M...
makeQuery.m	1 KB	01/04/201...
ConnectMIMIC.m	1 KB	01/04/201...
.dropbox	1 KB	01/04/201...
query.sql	1 KB	01/04/201...
Machine_Learning_I...	1 KB	01/05/201...
mlcc1-problem-set...	6 KB	01/11/201...
MLCCData.mat	215 ...	01/11/201...
MLCCDataMini.mat	6 KB	01/11/201...
MLCC_Lesson1.m	1 KB	01/11/201...
mysql-connector-ja...	960 ...	01/10/201...
postgresql-9.4.120...	593 ...	01/04/201...

query.sql (SQL File)

```

+11 -----
1 %% MIMIC CONNECTION in Matlab
2 %% Machine Learning in Critical Care.
3 %% Lesson #1: Connecting to the database.
4 %% Mohamad M. Ghassemi
5
6 %% STEP 1: Tell Matlab where the driver is
7 javaclasspath('/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.
8
9 %% STEP 2: Connect to the Database
10 conn = ConnectMIMIC;
11
12
13 %% Step 3: Open your web-browser and let's look at the schema
14 %% https://mi.mic.physionet.org/about/mi.mic/
15
16 %% Step 4: Let's write out a query.
17 %% create a file query.sql in your folder.
18 query = makeQuery('query.sql');
19
20 %% Step 5: Run your query and get the results.
21 results = fetch(conn,query)
22
23 %% Step 6: Do some additional Analysis on the results
24
25
26
27
28
29
30
31
  
```

Workspace

Name	Value	Size	Class	Var

Command History

```

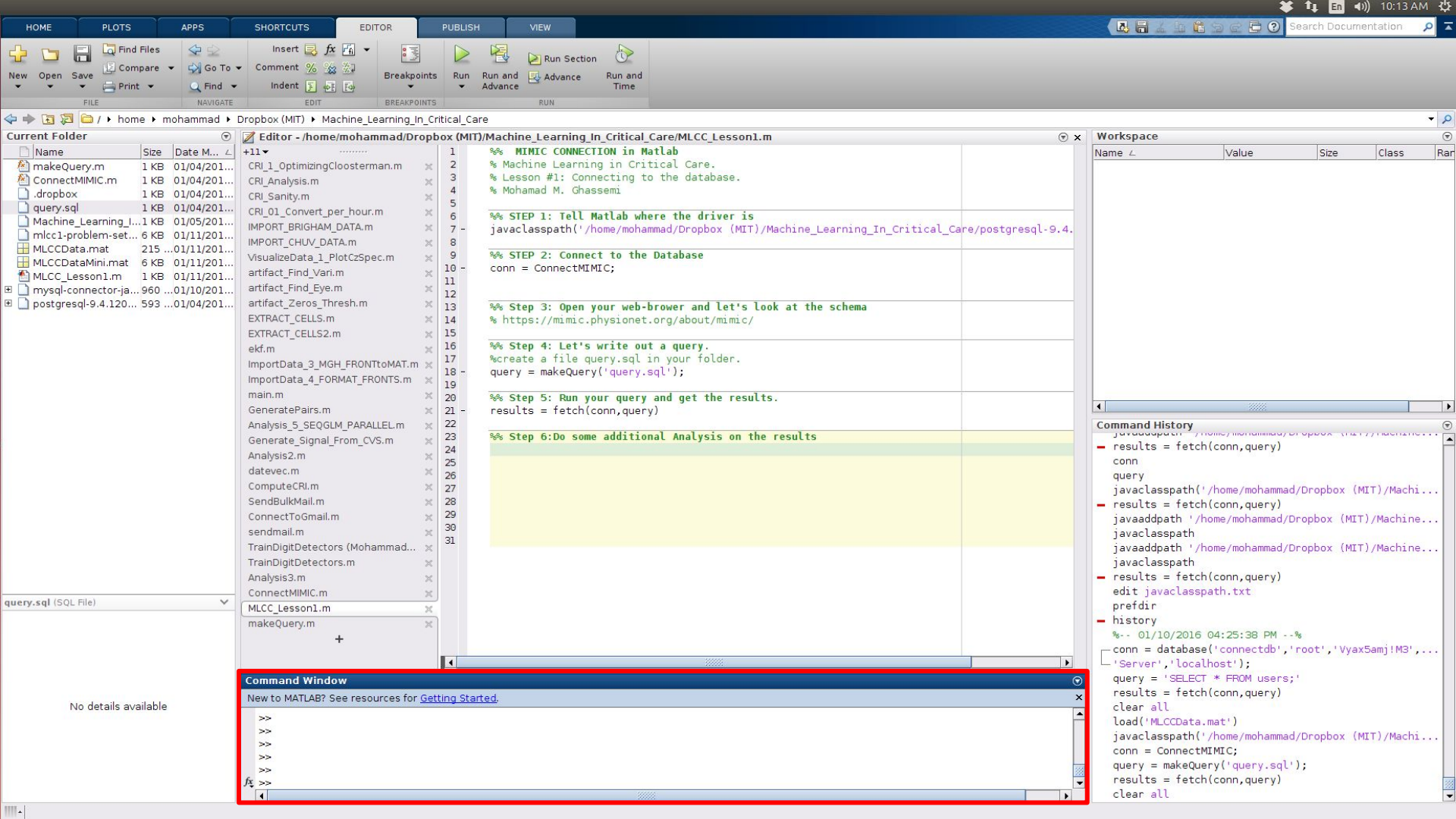
- results = fetch(conn,query)
  conn
  query
  javaclasspath('/home/mohammad/Dropbox (MIT)/Machi...
- results = fetch(conn,query)
  javaaddpath '/home/mohammad/Dropbox (MIT)/Machine...
  javaclasspath
  javaaddpath '/home/mohammad/Dropbox (MIT)/Machine...
  javaclasspath
- results = fetch(conn,query)
  edit javaclasspath.txt
  prefdir
- history
  % - 01/10/2016 04:25:38 PM --%
  conn = database('connectdb','root','VyaxSamj!M3',...
  ['Server','localhost'];
  query = 'SELECT * FROM users;
  results = fetch(conn,query)
  clear all
  load('MLCCData.mat')
  javaclasspath('/home/mohammad/Dropbox (MIT)/Machi...
  conn = ConnectMIMIC;
  query = makeQuery('query.sql');
  results = fetch(conn,query)
  clear all
  
```

Command Window

```

New to MATLAB? See resources for Getting Started.
>>
>>
>>
>>
>>
fx >>
  
```

No details available



Current Folder

Name	Size	Date M...
makeQuery.m	1 KB	01/04/201...
ConnectMIMIC.m	1 KB	01/04/201...
.dropbox	1 KB	01/04/201...
query.sql	1 KB	01/04/201...
Machine_Learning_I...	1 KB	01/05/201...
mlcc1-problem-set...	6 KB	01/11/201...
MLCCData.mat	215 ...	01/11/201...
MLCCDataMini.mat	6 KB	01/11/201...
MLCC_Lesson1.m	1 KB	01/11/201...
mysql-connector-ja...	960 ...	01/10/201...
postgresql-9.4.120...	593 ...	01/04/201...

```

Editor - /home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/MLCC_Lesson1.m
+11-----
CRI_1_.....
1 %% MIMIC CONNECTION in Matlab
2 %% Machine Learning in Critical Care.
3 %% Lesson #1: Connecting to the database.
4 %% Mohamad M. Ghassemi
5
6 %% STEP 1: Tell Matlab where the driver is
7 javaclasspath('/home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/postgresql-9.4.
8
9 %% STEP 2: Connect to the Database
10 conn = ConnectMIMIC;
11
12
13 %% Step 3: Open your web-browser and let's look at the schema
14 %% https://mimic.physionet.org/about/mimic/
15
16 %% Step 4: Let's write out a query.
17 %% create a file query.sql in your folder.
18 query = makeQuery('query.sql');
19
20 %% Step 5: Run your query and get the results.
21 results = fetch(conn,query)
22
23 %% Step 6: Do some additional Analysis on the results
24
25
26
27
28
29
30
31
    
```

Workspace

Name	Value	Size	Class	Par
test	100x100 double	100x100	double	0.9

Command History

```

- results = fetch(conn,query)
  conn
  query
  javaclasspath('/home/mohammad/Dropbox (MIT)/Machi...
- results = fetch(conn,query)
  javaaddpath '/home/mohammad/Dropbox (MIT)/Machine...
  javaclasspath
  javaaddpath '/home/mohammad/Dropbox (MIT)/Machine...
  javaclasspath
- results = fetch(conn,query)
  edit javaclasspath.txt
  prefdir
- history
  %-- 01/10/2016 04:25:38 PM --%
  conn = database('connectdb','root','VyaxSamjIM3',...
  'Server','localhost');
  query = 'SELECT * FROM users;';
  results = fetch(conn,query)
  clear all
  load('MLCCData.mat')
  javaclasspath('/home/mohammad/Dropbox (MIT)/Machi...
  conn = ConnectMIMIC;
  query = makeQuery('query.sql');
  results = fetch(conn,query)
  clear all
  test = rand(100,100);
    
```

No details available

Command Window

New to MATLAB? See resources for [Getting Started](#).

```

>>
>>
>>
>> test = rand(100,100);
fx >>
    
```


Software interface showing tabs: PLOTS, VARIABLE, VIEW. Includes menu options: Open, Rows, Columns, Insert, Delete, Transpose, Print, Sort.

test x 100x100 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
1	0.8147	0.1622	0.6443	0.0596	0.4229	0.5822	0.8507	0.5590	0.6837	0.9879	0.6312	0.7829	0.5038	0.4067	0.5038	0.1403	0.6723	0.8383	0.1982	0.1934	0.7449	0.9852	0.0560	
2	0.9058	0.7943	0.3786	0.6820	0.0942	0.5407	0.5606	0.8541	0.1321	0.1704	0.3551	0.6938	0.4896	0.6669	0.6128	0.2601	0.4315	0.5847	0.1951	0.7544	0.8923	0.5595	0.8169	
3	0.1270	0.3112	0.8116	0.0424	0.5985	0.8699	0.9296	0.3479	0.7227	0.2578	0.0098	0.8770	0.9337	0.8194	0.0868	0.6944	0.9970	0.9481	0.3268	0.3463	0.2426	0.9336	0.5289	
4	0.9134	0.5285	0.5328	0.0714	0.4709	0.2648	0.6967	0.4460	0.1104	0.3968	0.2242	0.8432	0.3531	0.8110	0.5319	0.4294	0.2568	0.0610	0.8803	0.4186	0.1296	0.7203	0.6944	
5	0.6324	0.1656	0.3507	0.5216	0.6959	0.3181	0.5828	0.0542	0.1175	0.0740	0.6525	0.9223	0.4494	0.4845	0.2021	0.2573	0.0098	0.5846	0.4711	0.1557	0.2251	0.4840	0.2124	
6	0.0975	0.6020	0.9390	0.0967	0.6999	0.1192	0.8154	0.1771	0.6407	0.6841	0.6050	0.2851	0.9635	0.7567	0.4539	0.2976	0.5323	0.6050	0.4040	0.8190	0.3500	0.6390	0.5433	
7	0.2785	0.2630	0.8759	0.8181	0.6385	0.9398	0.8790	0.6628	0.3288	0.4024	0.3872	0.0427	0.0423	0.4170	0.4279	0.4249	0.2794	0.8277	0.1792	0.6249	0.2871	0.8876	0.7025	
8	0.5469	0.6541	0.5502	0.8175	0.0336	0.6456	0.9889	0.3308	0.6538	0.9828	0.1422	0.3782	0.9730	0.9718	0.9661	0.1192	0.9462	0.1910	0.9689	0.7386	0.9275	0.1987	0.9564	
9	0.9575	0.6892	0.6225	0.7224	0.0688	0.4795	5.2238e-...	0.8995	0.7491	0.4022	0.0251	0.7043	0.1892	0.9880	0.6201	0.4951	0.9064	0.4425	0.4075	0.8051	0.0513	0.3954	0.4445	
10	0.9649	0.7482	0.5870	0.1499	0.3196	0.6393	0.8654	0.1182	0.5832	0.6207	0.4211	0.7295	0.6671	0.8641	0.6954	0.7064	0.3927	0.3934	0.8445	0.0672	0.5927	0.9922	0.0854	
11	0.1576	0.4505	0.2077	0.6596	0.5309	0.5447	0.6126	0.9884	0.7400	0.1544	0.1841	0.2243	0.5864	0.3889	0.7202	0.2436	0.0249	0.8266	0.6153	0.9508	0.1629	0.4024	0.0573	
12	0.9706	0.0838	0.3012	0.5186	0.6544	0.6473	0.9900	0.5400	0.2348	0.3813	0.7258	0.6291	0.6751	0.4547	0.3469	0.7851	0.6714	0.6769	0.3766	0.4976	0.8384	0.6589	0.6295	
13	0.9572	0.2290	0.4709	0.9730	0.4076	0.5439	0.5277	0.7069	0.7350	0.1611	0.3704	0.6730	0.3610	0.2467	0.5170	0.0741	0.8372	0.2076	0.8772	0.7551	0.1676	0.9013	0.7962	
14	0.4854	0.9133	0.2305	0.6490	0.8200	0.7210	0.4795	0.9995	0.9706	0.7581	0.8416	0.4775	0.6203	0.7844	0.5567	0.3939	0.9715	0.3181	0.7849	0.7424	0.5022	0.9954	0.6912	
15	0.8003	0.1524	0.8443	0.8003	0.7184	0.5225	0.8013	0.2878	0.8669	0.8781	0.7342	0.6237	0.8112	0.8828	0.1565	0.0034	0.0569	0.3138	0.4650	0.8311	0.9993	0.6532	0.3453	
16	0.1419	0.8258	0.1948	0.4538	0.9686	0.9937	0.2278	0.4145	0.0862	0.3508	0.5710	0.2364	0.0193	0.9137	0.5621	0.2207	0.4503	0.6715	0.8140	0.1565	0.3554	0.1084	0.9468	
17	0.4218	0.5383	0.2259	0.4324	0.5313	0.2187	0.4981	0.4648	0.3664	0.6855	0.1769	0.1771	0.0839	0.5583	0.6948	0.0013	0.5825	0.5710	0.8984	0.4573	0.0471	0.0361	0.5202	
18	0.9157	0.9961	0.1707	0.8253	0.3251	0.1058	0.9009	0.7640	0.3692	0.2941	0.9574	0.8296	0.9099	0.5989	0.4265	0.1892	0.6866	0.1698	0.8296	0.6181	0.2137	0.6181	0.9538	
19	0.7922	0.0782	0.2277	0.0835	0.1056	0.1097	0.5747	0.8182	0.6850	0.5306	0.2653	0.7669	0.6513	0.1489	0.8363	0.1425	0.7194	0.1477	0.3343	0.9322	0.3978	0.5671	0.0736	
20	0.9595	0.4427	0.4357	0.1332	0.6110	0.0636	0.8452	0.1002	0.5979	0.8324	0.9246	0.9345	0.2312	0.8997	0.7314	0.2681	0.6500	0.4761	0.5966	0.8351	0.3337	0.9620	0.2070	
21	0.6557	0.1067	0.3111	0.1734	0.7788	0.4046	0.7386	0.1781	0.7894	0.5795	0.1734	0.1079	0.4035	0.4504	0.3600	0.1749	0.9081	0.9020	0.8954	0.2296	0.7461	0.7750		
22	0.0357	0.9619	0.9234	0.3909	0.4235	0.4484	0.5860	0.3596	0.3677	0.3353	0.3736	0.1822	0.1220	0.2057	0.4542	0.1386	0.3738	0.5522	0.7021	0.5825	0.9361	0.6625	0.9142	
23	0.8491	0.0046	0.4302	0.8314	0.0908	0.3658	0.2467	0.0567	0.2060	0.2992	0.0875	0.0991	0.1284	0.8897	0.3864	0.5989	0.5816	0.0329	0.3775	0.5827	0.6832	0.5233	0.7826	
24	0.9340	0.7749	0.1848	0.8034	0.2665	0.7635	0.6664	0.5219	0.0867	0.4526	0.6401	0.4898	0.2578	0.7626	0.7756	0.9011	0.1161	0.0539	0.7350	0.8549	0.9621	0.2599	0.2955	
25	0.6787	0.8173	0.9049	0.0605	0.1537	0.6279	0.0835	0.3358	0.7719	0.4226	0.1806	0.1932	0.3317	0.8825	0.7343	0.9394	0.0577	0.8051	0.9541	0.0349	0.4380	0.9620	0.1518	
26	0.7577	0.8687	0.9797	0.0810	0.2810	0.7720	0.1526	0.1757	0.2057	0.3596	0.9951	0.8959	0.1522	0.2850	0.4303	0.2212	0.9798	0.4514	0.5428	0.8854	0.9403	0.5402	0.8479	
27	0.7431	0.0844	0.4389	0.5269	0.4401	0.9329	0.6609	0.2089	0.3883	0.5583	0.7232	0.0991	0.3480	0.6732	0.6938	0.4827	0.2848	0.3826	0.5401	0.4077	0.0058	0.0303	0.7849	
28	0.3922	0.3998	0.1111	0.4168	0.5271	0.9727	0.7298	0.9052	0.5518	0.7425	0.3474	0.0442	0.1217	0.6643	0.9452	0.3760	0.5950	0.7896	0.3111	0.0364	0.6103	0.6963	0.2708	
29	0.6555	0.2599	0.2581	0.6569	0.4574	0.1920	0.8908	0.6754	0.2290	0.4243	0.6606	0.5573	0.8842	0.1228	0.7042	0.5238	0.9622	0.3643	0.0712	0.7461	0.8011	0.5197	0.2278	
30	0.1712	0.8001	0.4087	0.6280	0.8754	0.1389	0.9823	0.4685	0.6419	0.4294	0.3839	0.7725	0.0943	0.4073	0.7856	0.2649	0.1858	0.5323	0.1820	0.1548	0.2330	0.0590	0.3210	
31	0.7060	0.4314	0.5949	0.2920	0.5181	0.6963	0.7690	0.9121	0.4845	0.1249	0.6273	0.3119	0.9300	0.2753	0.1093	0.0684	0.1930	0.7117	0.0930	0.1439	0.9325	0.8900	0.8296	
32	0.0318	0.9106	0.2622	0.4317	0.9436	0.0938	0.5814	0.1040	0.1518	0.0244	0.0216	0.3990	0.3990	0.7167	0.3899	0.4363	0.3416	0.8715	0.1790	0.4635	0.6060	0.7633	0.3302	0.8222
33	0.2769	0.1818	0.6028	0.0155	0.6377	0.5254	0.9283	0.7455	0.7819	0.2902	0.9106	0.3390	0.0474	0.2834	0.5909	0.1739	0.9329	0.3287	0.0093	0.2545	0.8264	0.2297	0.5707	
34	0.0462	0.2638	0.7112	0.9841	0.9577	0.5303	0.5801	0.7363	0.1006	0.3175	0.8006	0.2101	0.3424	0.8962	0.4594	0.0261	0.3907	0.6501	0.9150	0.3242	0.5735	0.1139	0.5718	
35	0.0971	0.1455	0.2217	0.1672	0.2407	0.8611	0.0170	0.5619	0.2941	0.6537	0.1672	0.5102	0.7360	0.8266	0.0503	0.9547	0.2732	0.9748	0.6427	0.4018	0.7926	0.3109	0.2860	
36	0.8235	0.1361	0.1174	0.1062	0.6761	0.4849	0.1209	0.1842	0.2374	0.9569	0.8131	0.9064	0.7947	0.3900	0.2287	0.4306	0.1519	0.0760	0.0014	0.4064	0.3290	0.2284	0.6991	
37	0.6948	0.8693	0.2967	0.3724	0.2891	0.3935	0.8627	0.5972	0.5309	0.9357	0.3833	0.6289	0.5449	0.4979	0.8342	0.9616	0.3971	0.5870	0.0304	0.3862	0.2235	0.6520	0.7963	
38	0.3171	0.5797	0.3188	0.1981	0.6718	0.6714	0.4843	0.2999	0.0915	0.1615	0.6718	0.1074	0.0862	0.6948	0.0156	0.7624	0.3747	0.4139	0.2085	0.6098	0.3124	0.0662	0.4416	
39	0.9502	0.5499	0.4242	0.4897	0.6951	0.7413	0.8449	0.1341	0.4053	0.2405	0.5755	0.3909	0.8936	0.8344	0.8637	0.0073	0.1311	0.3091	0.4550	0.1669	0.5845	0.2754	0.4462	
40	0.0344	0.1450	0.5079	0.3395	0.0680	0.5201	0.2094	0.2126	0.1048	0.7639	0.5301	0.0546	0.0548	0.6096	0.0781	0.6800	0.4350	0.2638	0.1273	0.1881	0.8299	0.2818	0.4657	

Insert fx f | Comment % | Breakpoints | Run | Run and Advance | Run Section | Advance | Run and Time

FILE NAVIGATE EDIT BREAKPOINTS RUN

home > mohammad > Dropbox (MIT) > Machine_Learning_In_Critical_Care

Current Folder

Name	Size	Date M...
makeQuery.m	1 KB	01/04/201...
ConnectMIMIC.m	1 KB	01/04/201...
.dropbox	1 KB	01/04/201...
query.sql	1 KB	01/04/201...
Machine_Learning_I...	1 KB	01/05/201...
mlcc1-problem-set...	6 KB	01/11/201...
MLCCData.mat	215 ...	01/11/201...
MLCCDataMini.mat	6 KB	01/11/201...
MLCC_Lesson1.m	1 KB	01/11/201...
mysql-connector-ja...	960 ...	01/10/201...
postgresql-9.4.120...	593 ...	01/04/201...

```

Editor - /home/mohammad/Dropbox (MIT)/Machine_Learning_In_Critical_Care/MLCC_Lesson1.m
+11-----
CRI_1_OptimizingCloosterman.m 1  %% MIMIC CONNECTION in Matlab
CRI_Analysis.m 2  %% Machine Learning in Critical Care.
CRI_Sanity.m 3  %% Lesson #1: Connecting to the database.
CRI_01_Convert_per_hour.m 4  %% Mohamad M. Ghassemi
IMPORT_BRIGHAM_DATA.m 5
IMPORT_CHUVU_DATA.m 6
VisualizeData_1_PlotCzSpec.m 7
artifact_Find_Vari.m 8
artifact_Find_Eye.m 9
artifact_Zeros_Thresh.m 10
EXTRACT_CELLS.m 11
EXTRACT_CELLS2.m 12
ekf.m 13
ImportData_3_MGH_FRONTtoMAT.m 14
ImportData_4_FORMAT_FRONTs.m 15
main.m 16
GeneratePairs.m 17
Analysis_5_SEQGLM_PARALLEL.m 18
Generate_Signal_From_CVS.m 19
Analysis2.m 20
datevec.m 21
ComputeCRI.m 22
SendBulkMail.m 23
ConnectToGmail.m 24
sendmail.m 25
TrainDigitDetectors (Mohammad... 26
TrainDigitDetectors.m 27
Analysis3.m 28
ConnectMIMIC.m 29
MLCC_Lesson1.m 30
makeQuery.m 31
  
```

Workspace

Name	Value	Size	Class	Var
test	100x100 double	100x100	double	0.9

Command History

```

- results = fetch(conn,query);
  conn
  query
  javaclasspath('/home/mohammad/Dropbox (MIT)/Machi...
- results = fetch(conn,query)
  javaaddpath '/home/mohammad/Dropbox (MIT)/Machine...
  javaclasspath
  javaaddpath '/home/mohammad/Dropbox (MIT)/Machine...
  javaclasspath
- results = fetch(conn,query)
  edit javaclasspath.txt
  prefdir
- history
  %-- 01/10/2016 04:25:38 PM --%
  conn = database('connectdb','root','VyaxSamjIM3',...
  'Server','localhost');
  query = 'SELECT * FROM users;';
  results = fetch(conn,query)
  clear all
  load('MLCCData.mat')
  javaclasspath('/home/mohammad/Dropbox (MIT)/Machi...
  conn = ConnectMIMIC;
  query = makeQuery('query.sql');
  results = fetch(conn,query)
  clear all
  test = rand(100,100);
  
```

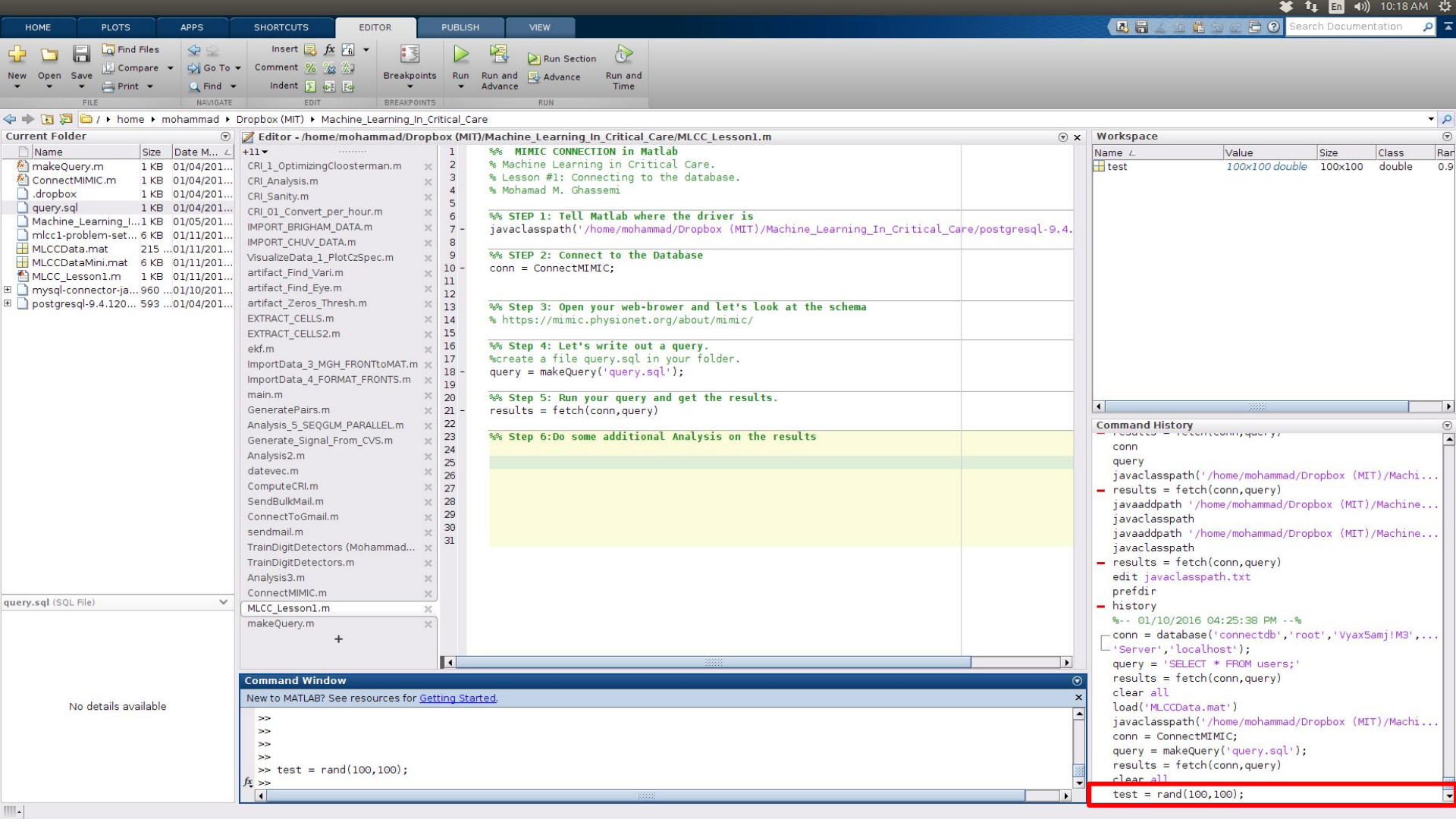
No details available

Command Window

New to MATLAB? See resources for [Getting Started](#).

```

>>
>>
>>
>> test = rand(100,100);
fx >>
  
```



The Human Learning Task

To the **Mat**Mobile





The Human Learning Task

- I want all of you to forget everything you ever learned about machine learning, and treat mortality prediction as a *human learning* task.
- Jim had a cardiac arrest, and is currently on life support. Nobody knows if Jim will wake up from his coma, and taking care of him is very expensive ~20K/day.
- Your boss wants you to generate a procedure to predict Jim's outcome.



The Human Learning Task

- We are going to design an algorithm together to predict the outcomes of the patients by visual inspection alone.
- We want to figure out ***which features*** we extracted are predictive of outcome, and ***how to choose thresholds for their values.***

To the **Mat**Mobile



%% STEP 1: Let's load the data.

Code will be referenced

Load the Data

%% STEP 1: Let's load the data.

% you can also do this in the GUI by double clicking the '.mat' file
% on the left hand side panel. Or you can run...
load('MLCCData.mat');

%% STEP 2: Let's Unpack the Data

```
for i = 1:size(data,2)
    name = header{i};
    eval([name '= data(:,i);'])
end
```

Load the Data

```
%% STEP 1: Let's load the data.
```

```
% you can also do this in the GUI by double clicking the '.mat' file  
% on the left hand side panel. Or you can run...
```

```
load('MLCCData.mat');
```

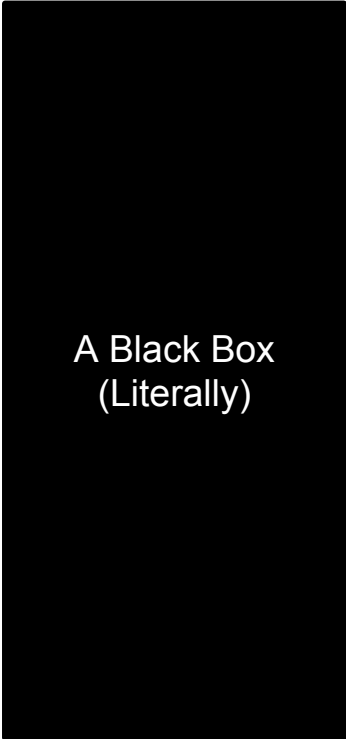
```
%% STEP 2: Let's Unpack the Data
```

```
for i = 1:size(data,2)  
    name = header{i};  
    eval([name '= data(:,i);'])  
end
```

**NOW LOOK AT YOUR
WORKSPACE!**

We want to find features that predict outcome

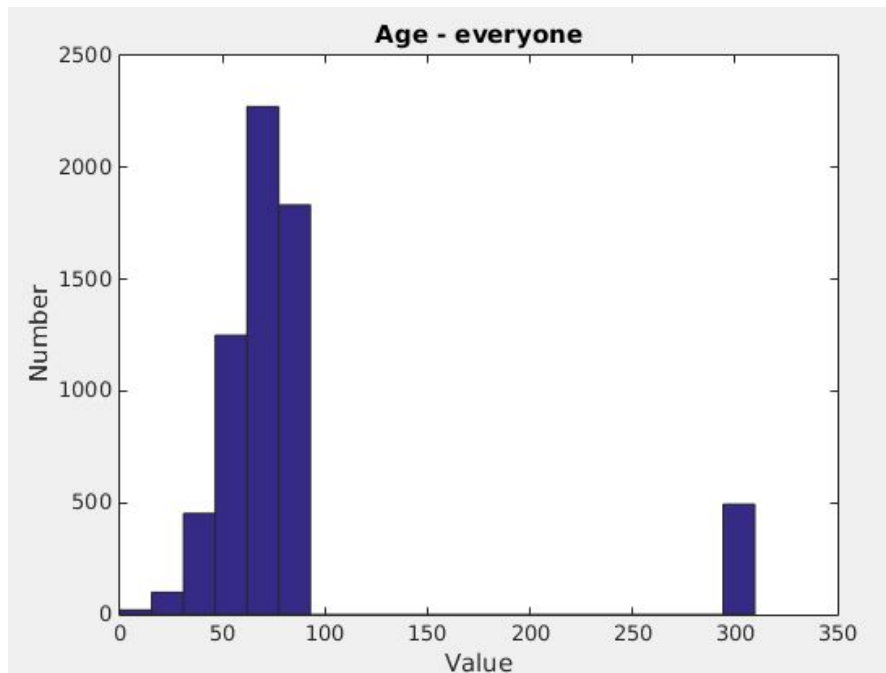
```
'ICUSTAYID'  
'OUTCOME'  
'Age'  
'HeartRateMin'  
'HeartRateMax'  
'MeanBPMIn'  
'MeanBPMax'  
'RespRateMin'  
'RespRateMax'  
'GCSMin'  
'GCSMax'  
'BILIRUBINmin'  
'BILIRUBINmax'  
'CREATININemin'  
'CREATININemax'  
'HEMOGLOBINmin'  
'HEMOGLOBINmax'  
'SODIUMmin'  
'SODIUMmax'  
'WBCmin'  
'WBCmax'
```



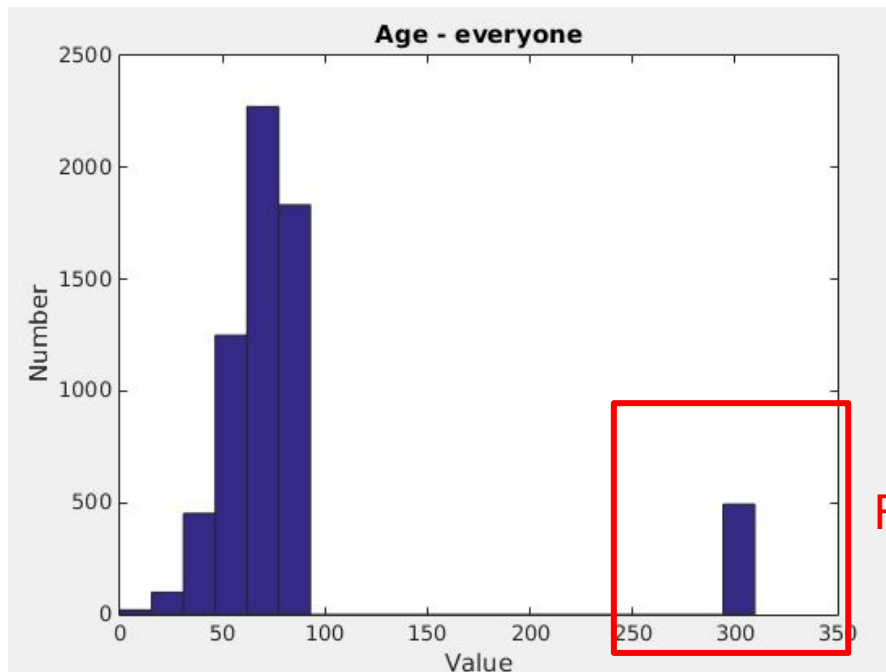
A Black Box
(Literally)

%% STEP 3: Let's Visualize some features, starting with Age

Let's Look at Age

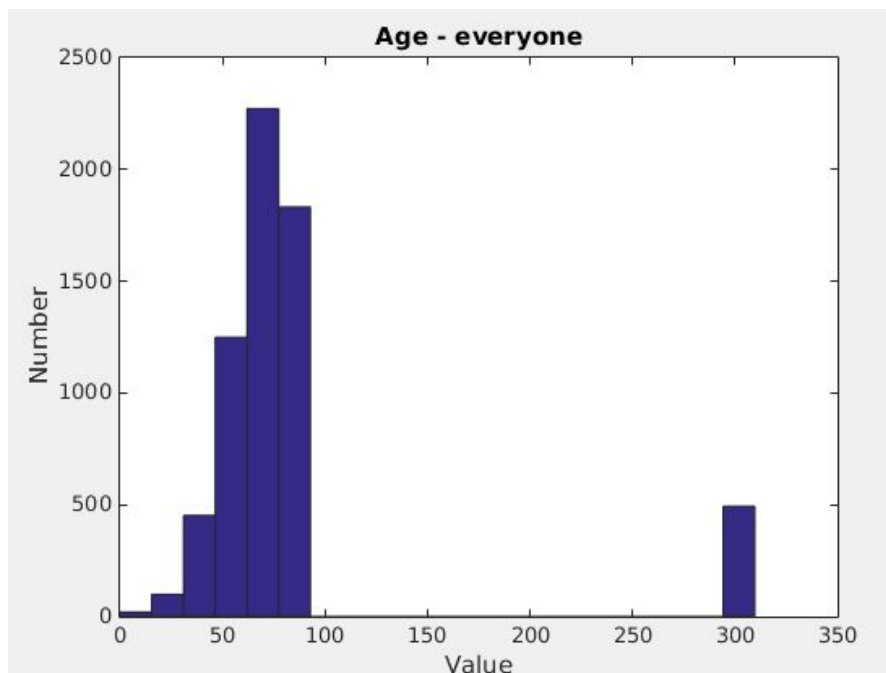


Strange Outliers



Fountain of Youth?!

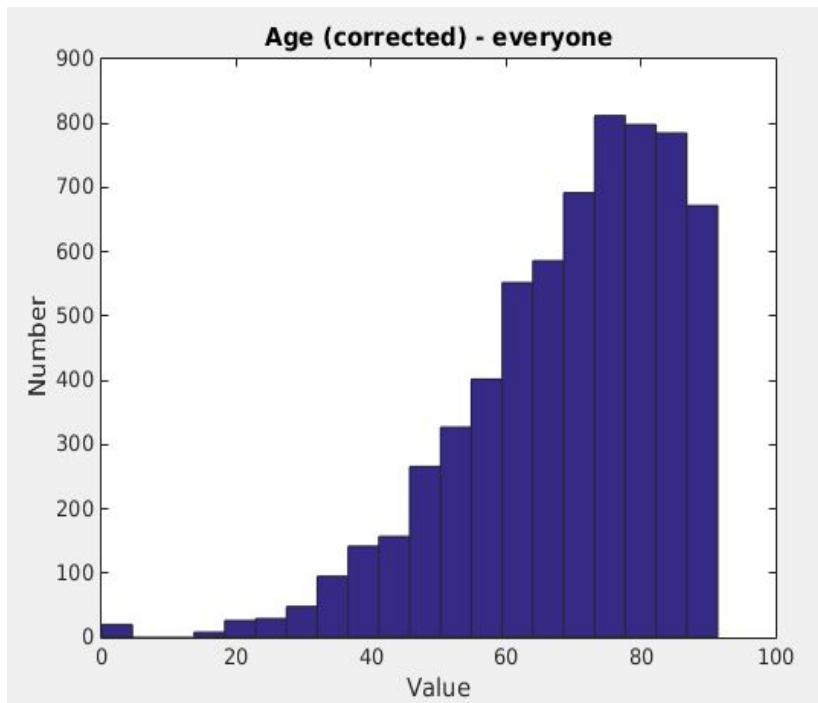
Outliers exist to protect identity of older patients



- DOB has been shifted for patients older than 89. The median age for the patients whose date of birth was shifted is 91.4.

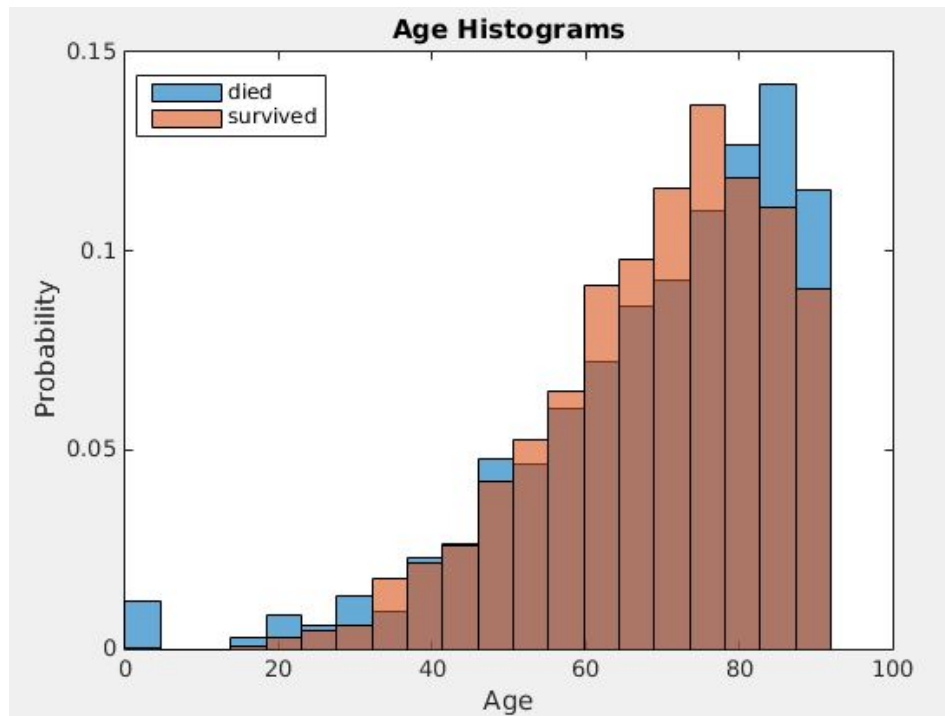
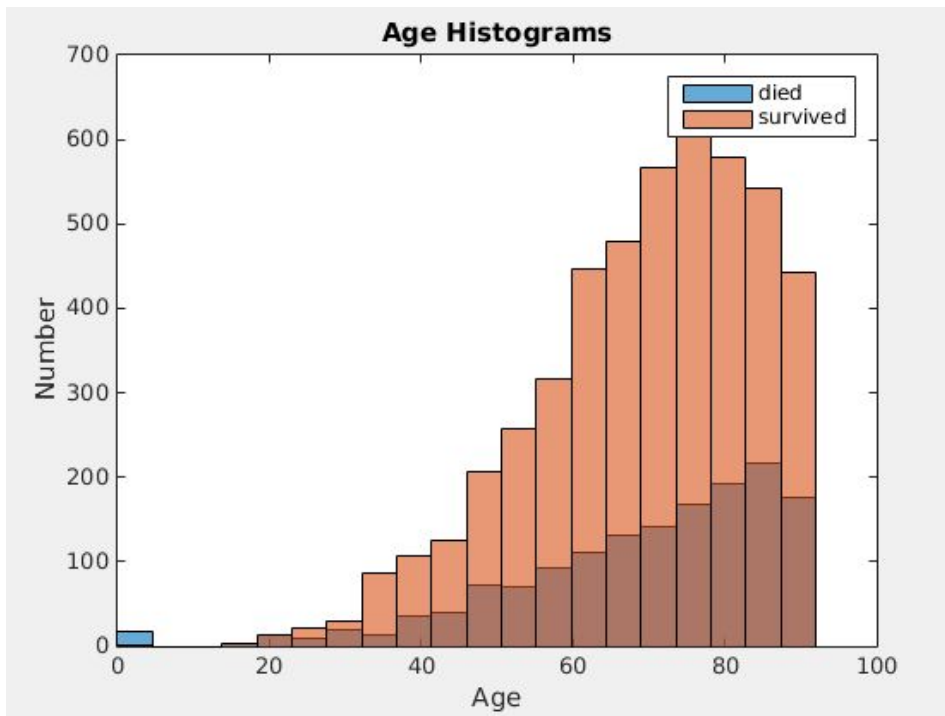
<https://mimic.physionet.org/mimictables/patients/>

Corrected, the distributions of ages make sense

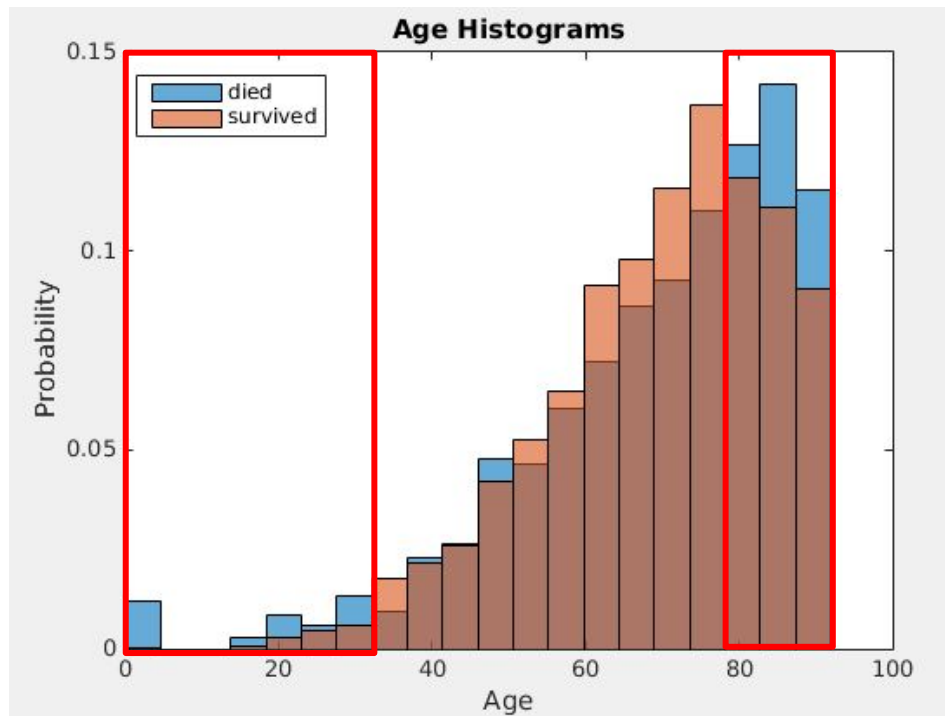
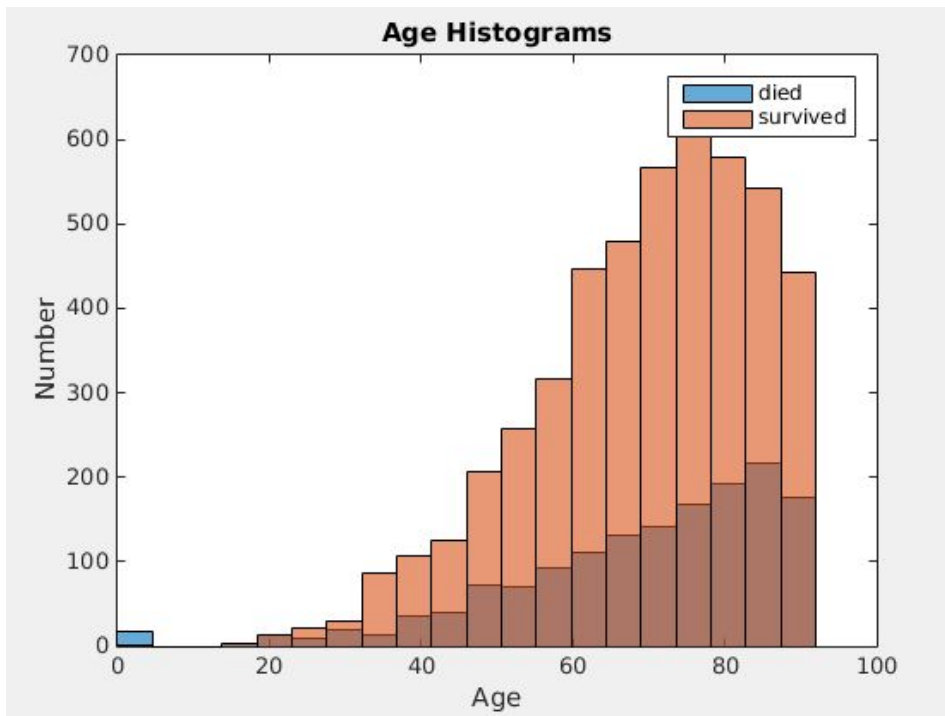


- Much Better!

Compare histograms of each outcome class



Survival is more likely when $30 < \text{Age} < 75$



What features should be considered...?

'ICUSTAYID'
'OUTCOME'
'Age'
'HeartRateMin'
'HeartRateMax'
'MeanBPMin'
'MeanBPMax'
'RespRateMin'
'RespRateMax'
'GCSMin'
'GCSMax'
'BILIRUBINmin'
'BILIRUBINmax'
'CREATININEmin'
'CREATININEmax'
'HEMOGLOBINmin'
'HEMOGLOBINmax'
'SODIUMmin'
'SODIUMmax'
'WBCmin'
'WBCmax'



A Black Box
(Literally)

Inside The Box

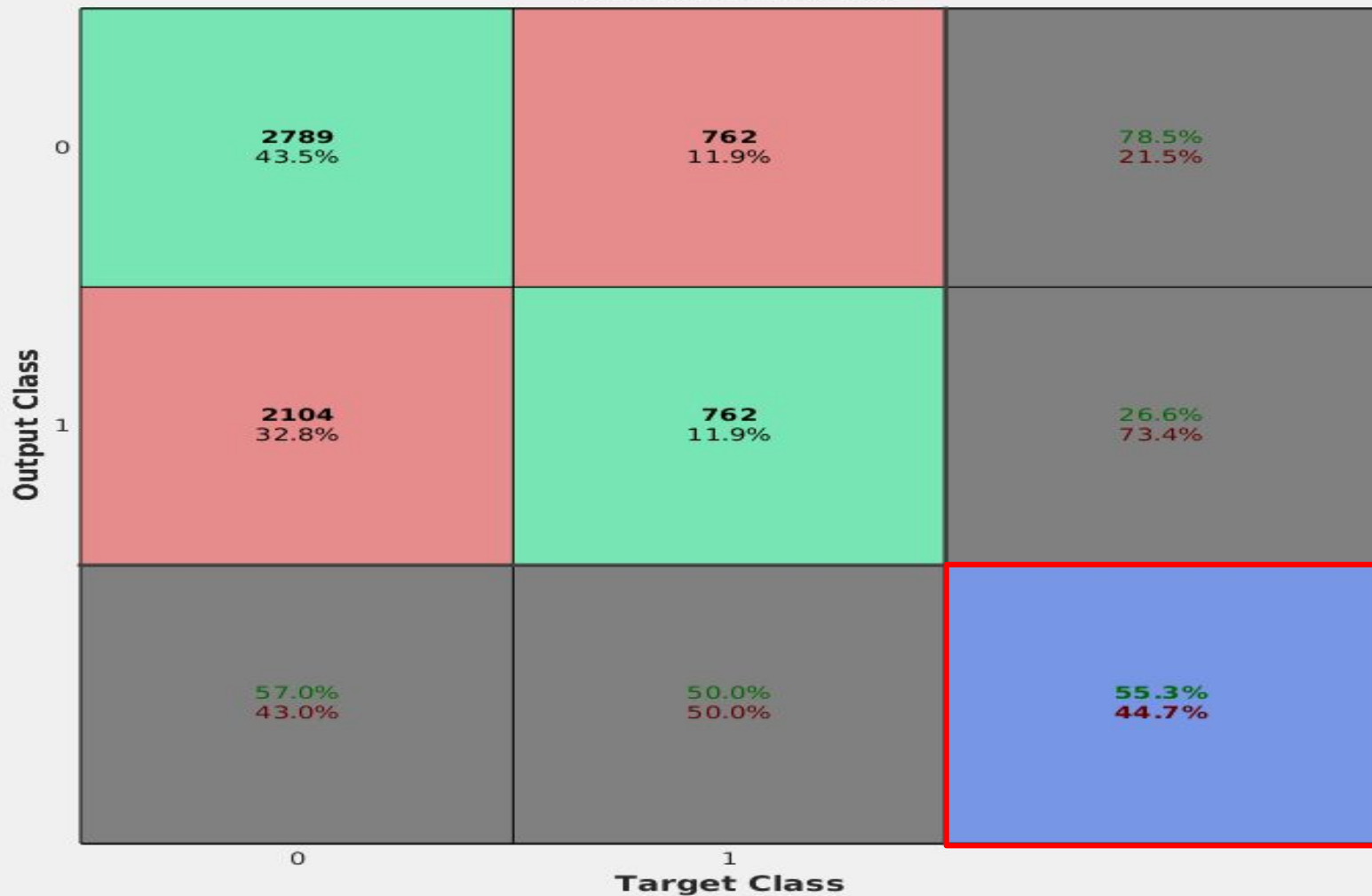
prediction = death if
(Age < 30 | Age > 75)

Confusion Matrix

A confusion matrix heatmap with 'Output Class' on the y-axis (0, 1) and 'Target Class' on the x-axis (0, 1). The cells contain counts and percentages. The diagonal cells (0,0) and (1,1) are green, while the off-diagonal cells (0,1) and (1,0) are red. The bottom-right cell (1,2) is blue, and the bottom-left cell (2,0) is grey.

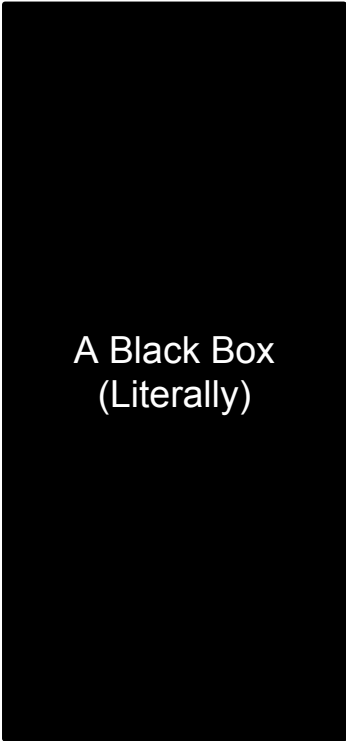
Output Class	Target Class 0	Target Class 1	Target Class 2
0	2789 43.5%	762 11.9%	78.5% 21.5%
1	2104 32.8%	762 11.9%	26.6% 73.4%
2	57.0% 43.0%	50.0% 50.0%	55.3% 44.7%

Confusion Matrix



What features should be considered...?

'ICUSTAYID'
'OUTCOME'
'Age'
'HeartRateMin'
'HeartRateMax'
'MeanBPMin'
'MeanBPMax'
'RespRateMin'
'RespRateMax'
'GCSMin'
'GCSMax'
'BILIRUBINmin'
'BILIRUBINmax'
'CREATININEmin'
'CREATININEmax'
'HEMOGLOBINmin'
'HEMOGLOBINmax'
'SODIUMmin'
'SODIUMmax'
'WBCmin'
'WBCmax'



A Black Box
(Literally)

What features should be considered...?

'ICUSTAYID'
'OUTCOME'
'Age'
'HeartRateMin'
'HeartRateMax'
'MeanBPMin'
'MeanBPMax'
'RespRateMin'
'RespRateMax'
'GCSMin'
'GCSMax'
'BILIRUBINmin'
'BILIRUBINmax'
'CREATININEmin'
'CREATININEmax'
'HEMOGLOBINmin'
'HEMOGLOBINmax'
'SODIUMmin'
'SODIUMmax'
'WBCmin'
'WBCmax'



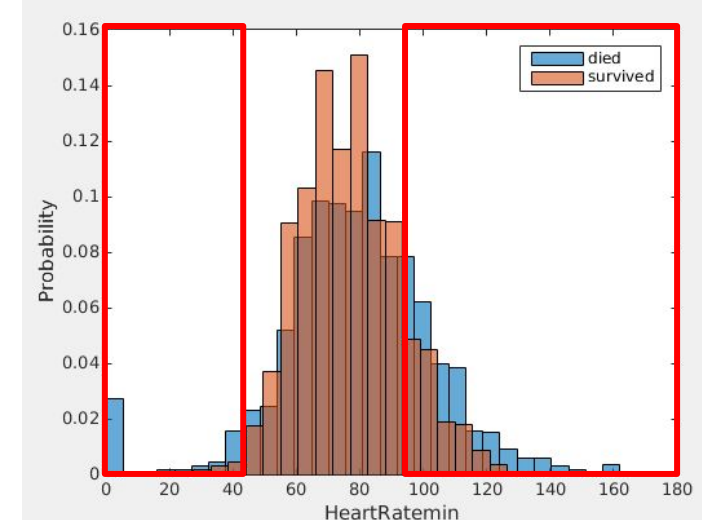
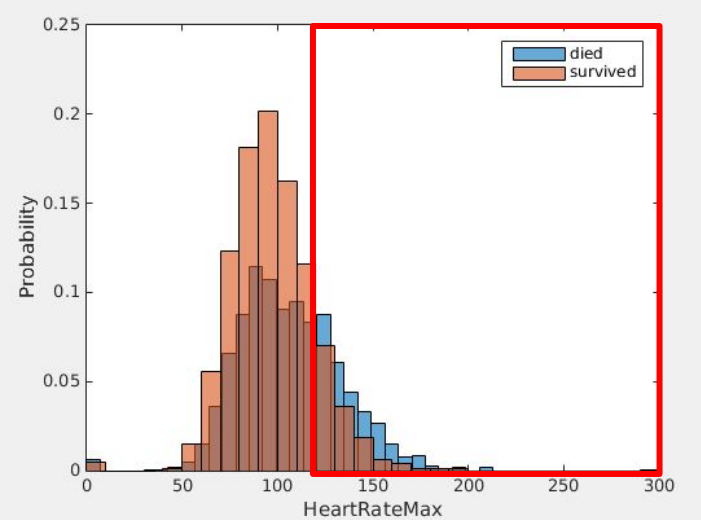
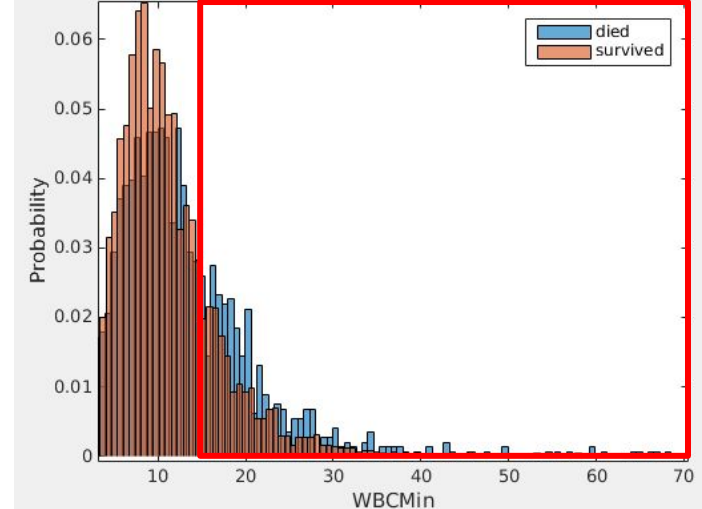
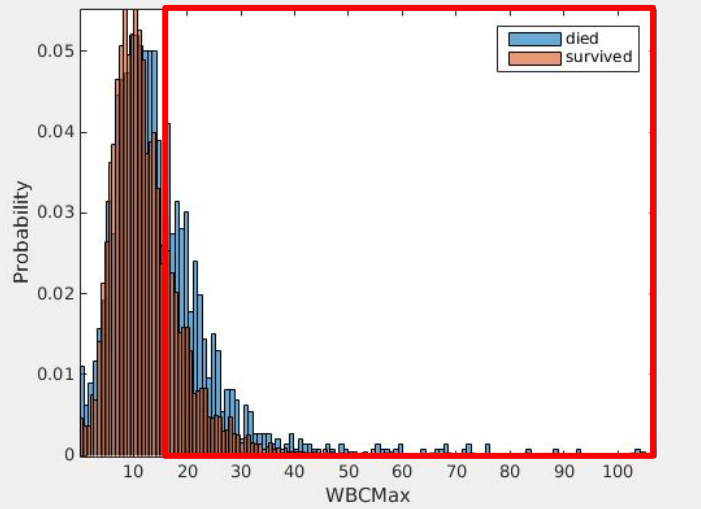
Lived!

Died!

We can do better!

STEP 7: More Features

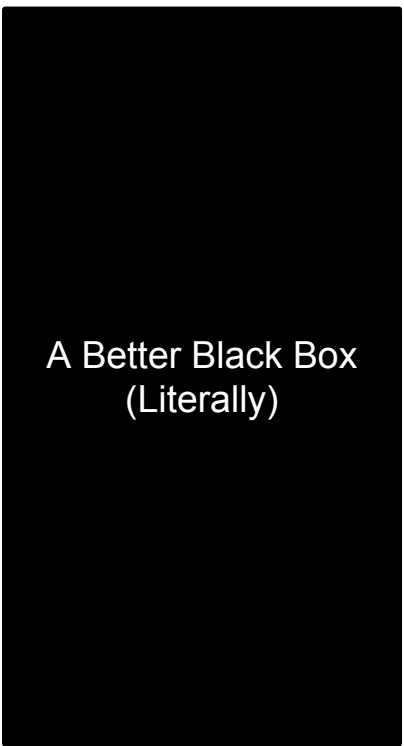
Interesting patterns exist for other features



What would the histograms
of a perfect feature look
like?

Let's Make a Better Black Box

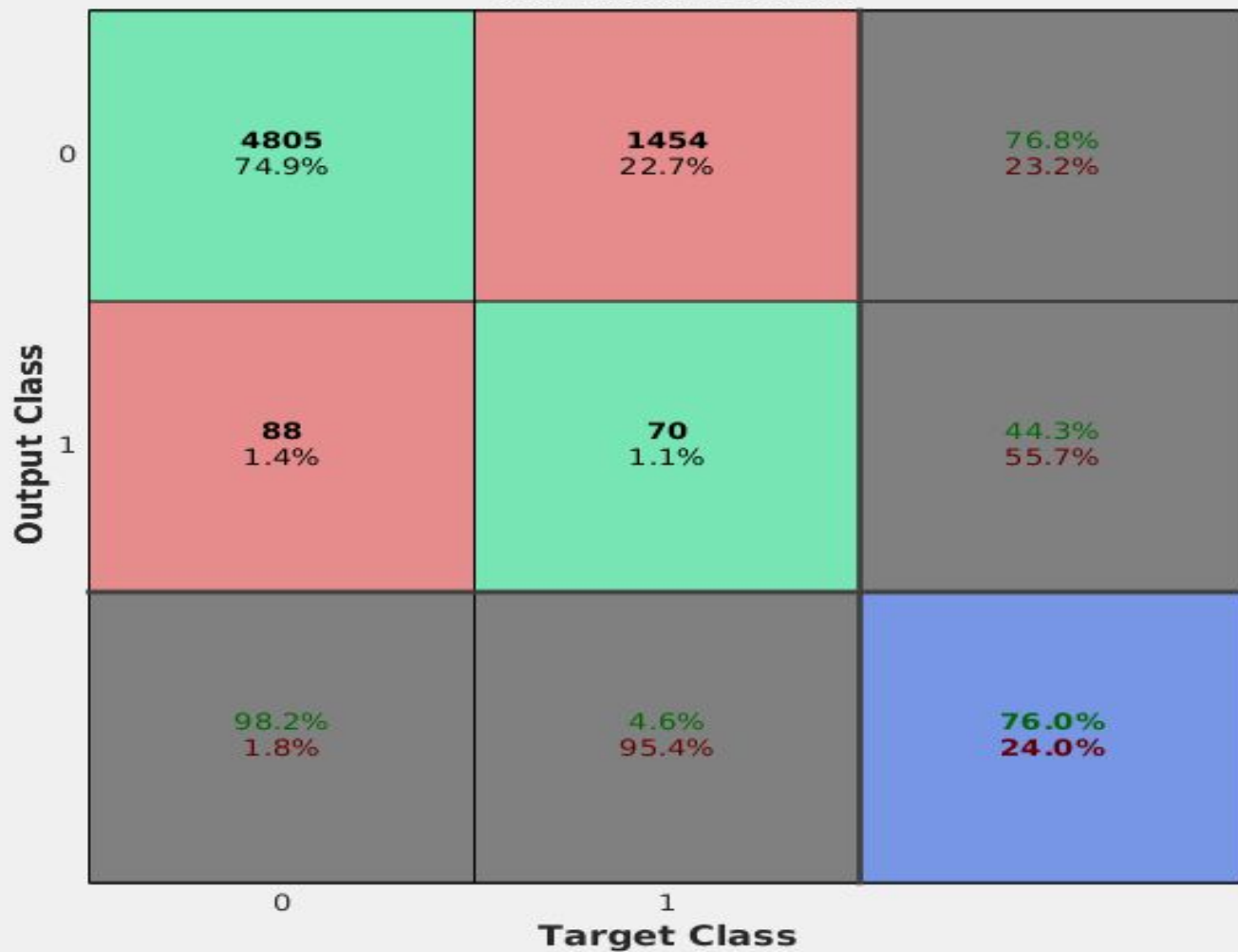
' ICUSTAYID'
' OUTCOME'
' Age'
' HeartRateMin' →
' HeartRateMax' →
' MeanBPMin'
' MeanBPMax'
' RespRateMin'
' RespRateMax'
' GCSMin'
' GCSMax'
' BILIRUBINmin'
' BILIRUBINmax'
' CREATININEmin
' CREATININEmax
' HEMOGLOBINmin
' HEMOGLOBINmax
' SODIUMmin'
' SODIUMmax'
' WBCmin' →
' WBCmax' →



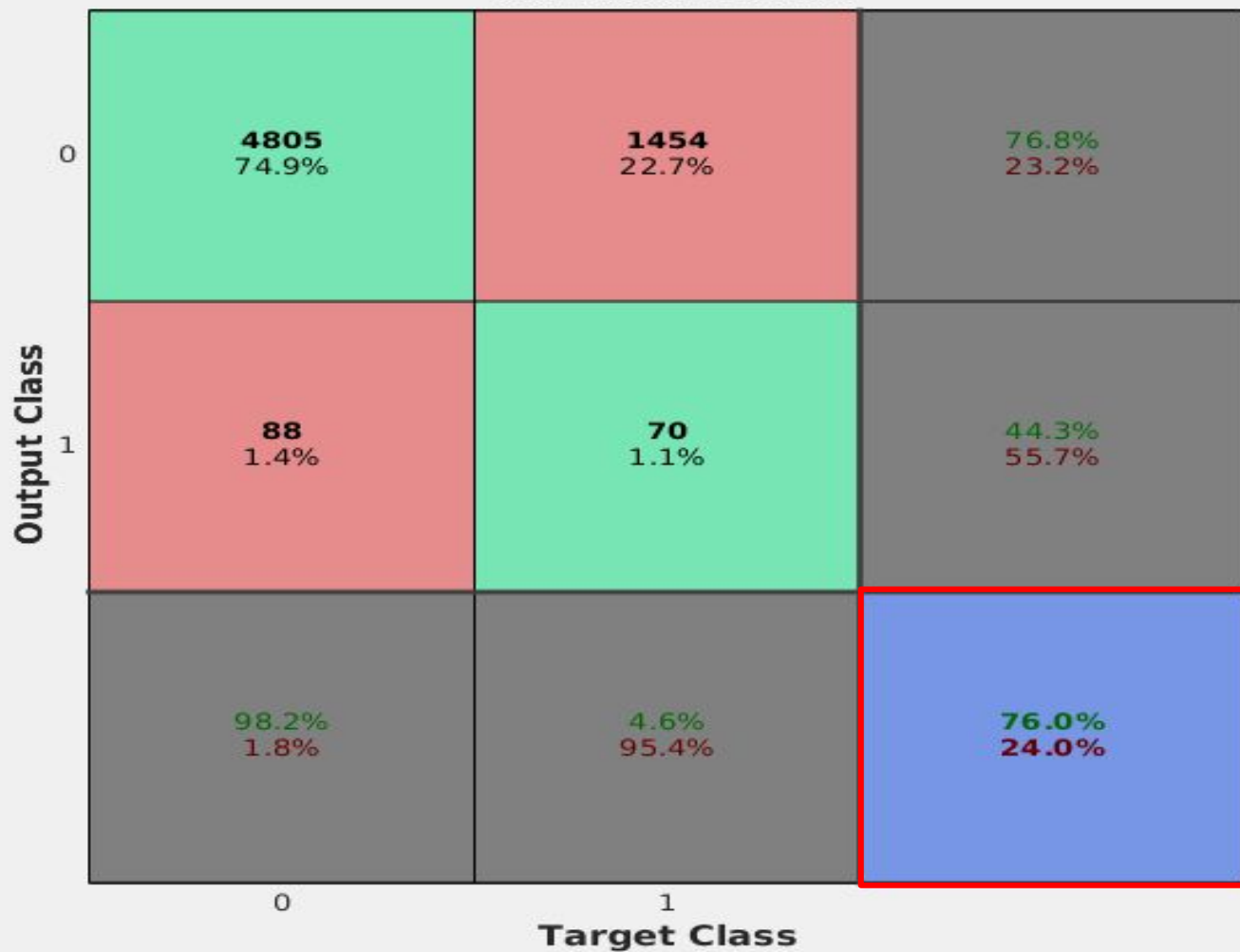
Inside The Better Black Box

prediction = death if
WBCmax > 15 &
WBCmin > 15 &
HeartRateMin > 100

Confusion Matrix



Confusion Matrix

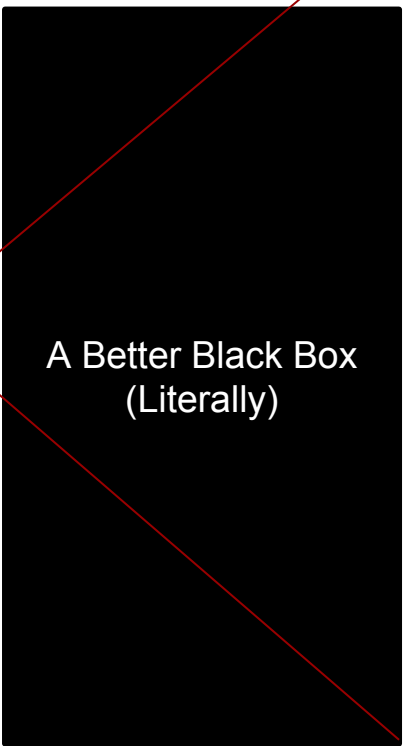


Victory!

Or is it...?

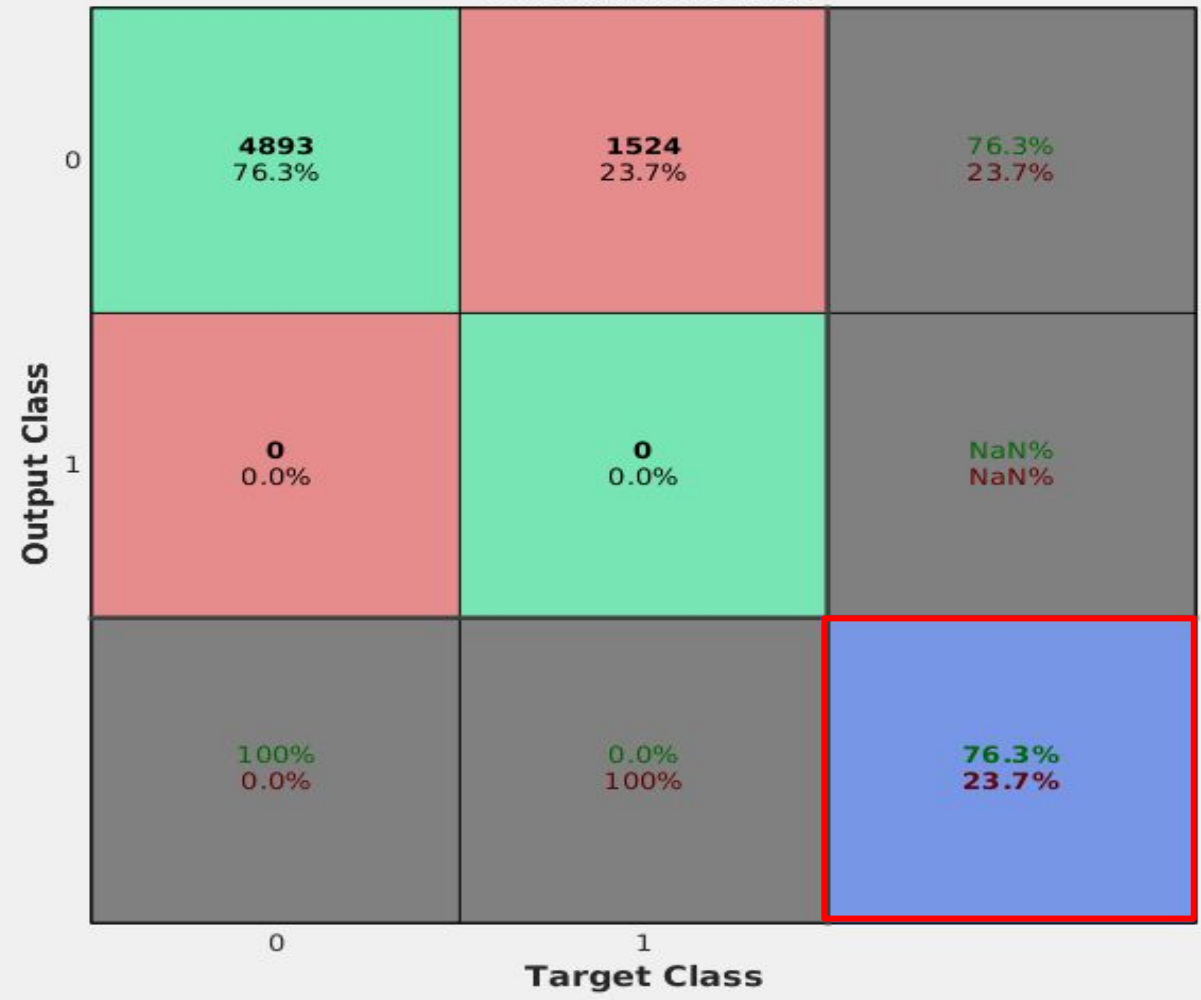
Let's Try The Optimist's Algorithm

'ICUSTAYID'
'OUTCOME'
'Age'
'HeartRateMin' →
'HeartRateMax' →
'MeanBPMin'
'MeanBPMax'
'RespRateMin'
'RespRateMax'
'GCSMin'
'GCSMax'
'BILIRUBINmin'
'BILIRUBINmax'
'CREATININEmin'
'CREATININEmax'
'HEMOGLOBINmin'
'HEMOGLOBINmax'
'SODIUMmin'
'SODIUMmax'
'WBCmin' →
'WBCmax' →



Everyone Always Survives!

Confusion Matrix



Oh dear...
0.3% better...

A silhouette of a hand holding a string of bubbles against a sunset sky. The bubbles are translucent and reflect the colors of the sky. The text "Choose Optimism." is overlaid in a large, white, sans-serif font with a slight drop shadow.

**Choose
Optimism.**

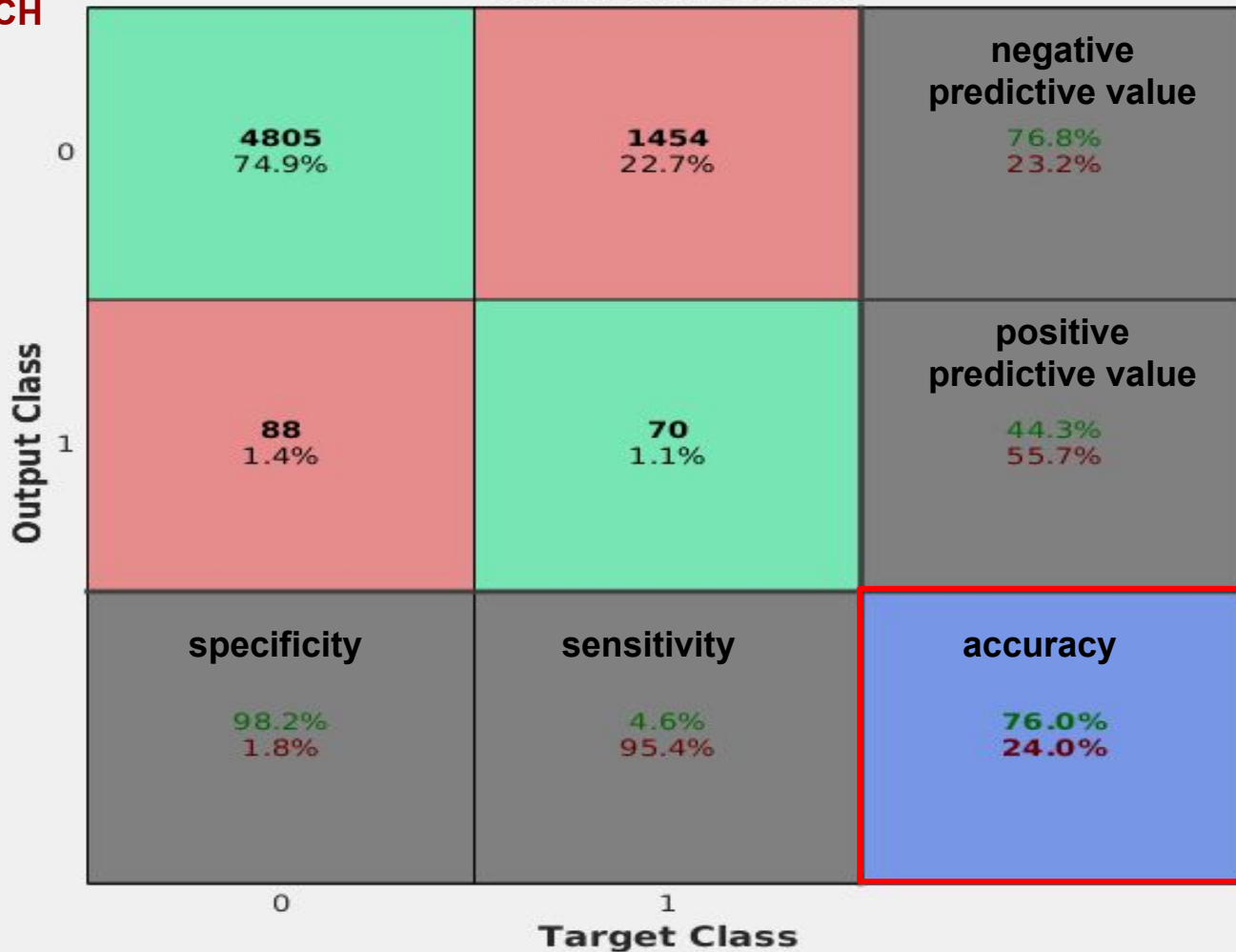
A hand is shown in silhouette, holding a large, iridescent bubble. The background is a bright, cloudy sky at sunset or sunrise, with a warm yellow and orange glow. The text "Choose Optimism." is written in a large, white, sans-serif font across the center of the image. Two thin red diagonal lines cross the entire image from corner to corner, forming an 'X' shape.

**Choose
Optimism.**

Why is this happening?

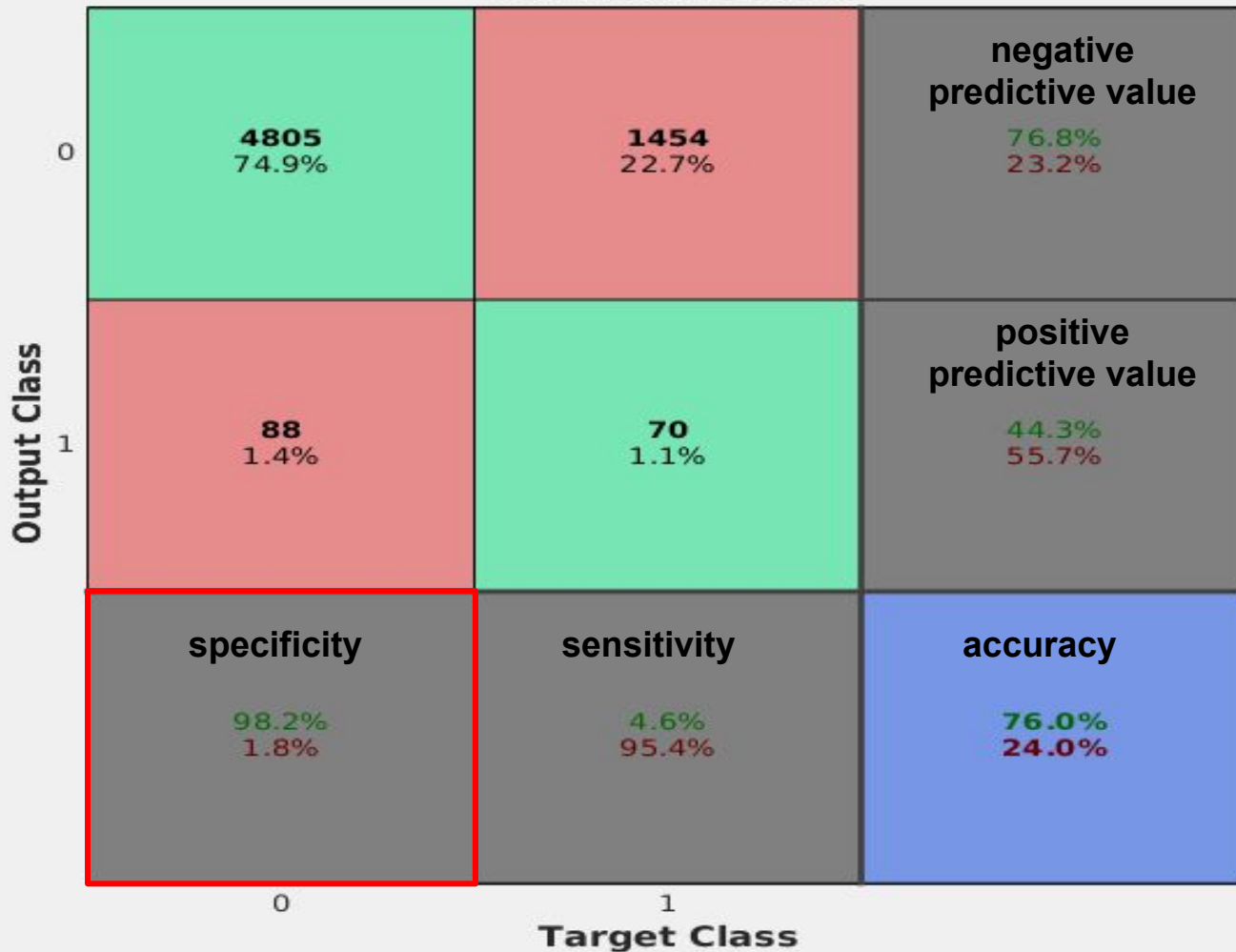
OUR APPROACH

Confusion Matrix



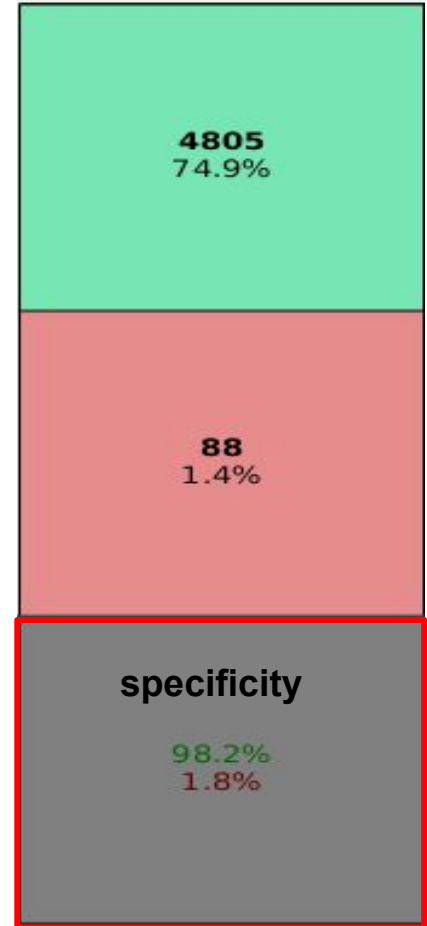
Accuracy may not be a good measure of performance when you have uneven class sizes.

Confusion Matrix

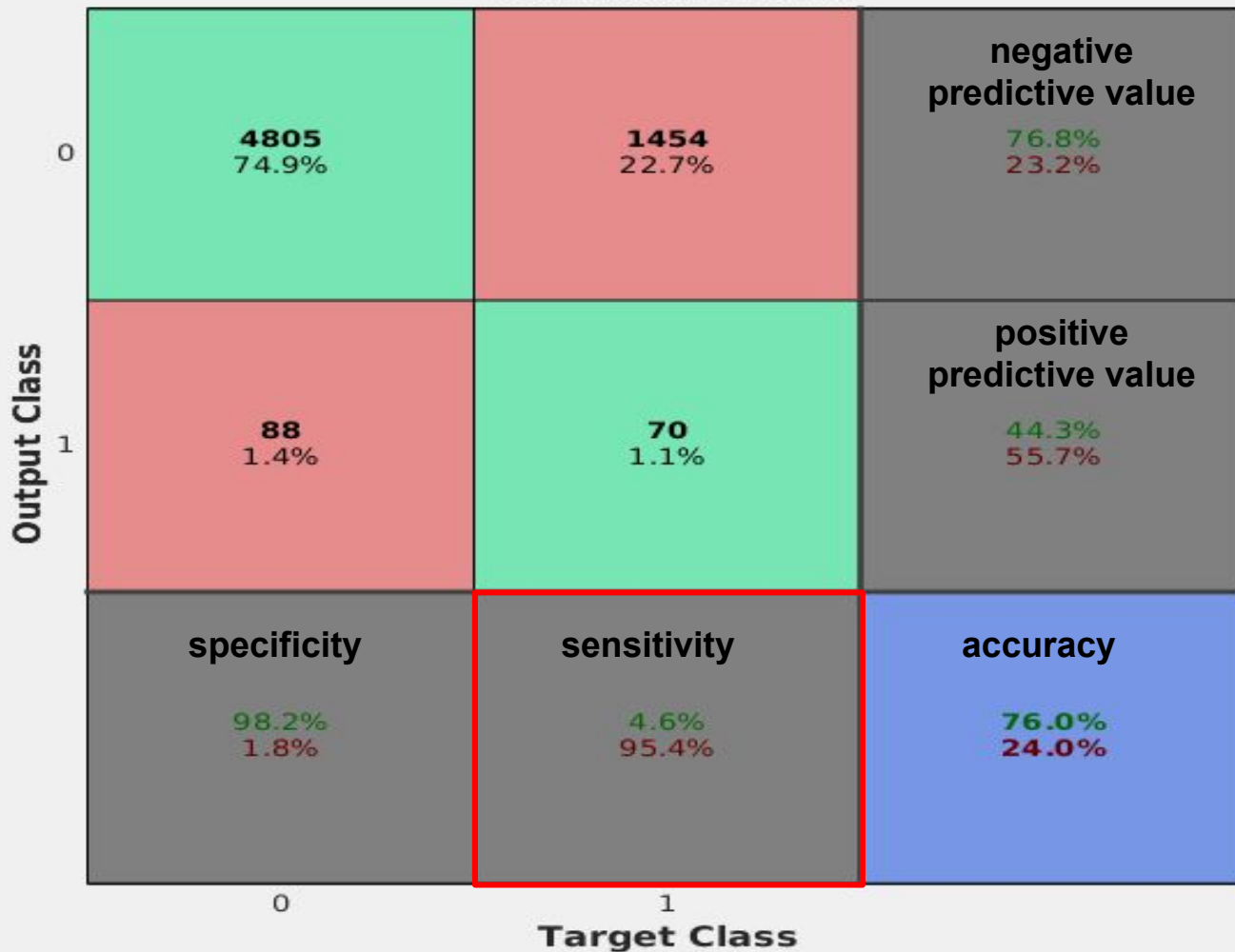


Specificity

- $\text{Specificity} = \frac{\text{true survivors}}{(\text{true survivors} + \text{false deaths})}$
- Given that a person is going to survive, how often will the test indicate survival?

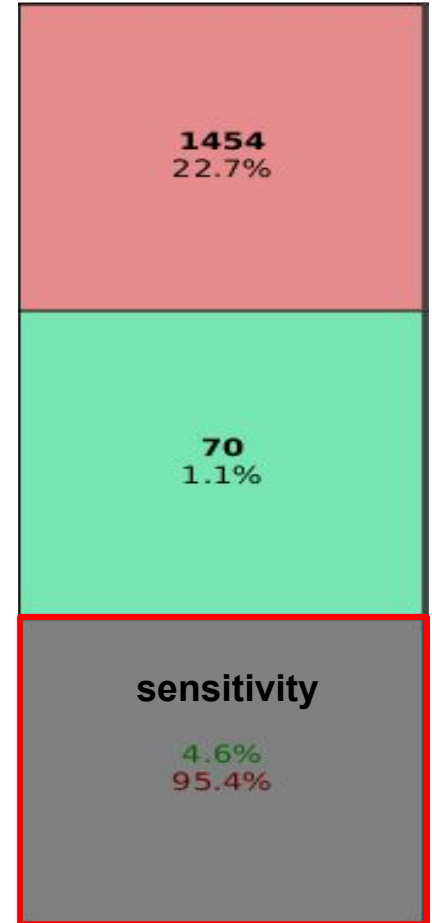


Confusion Matrix

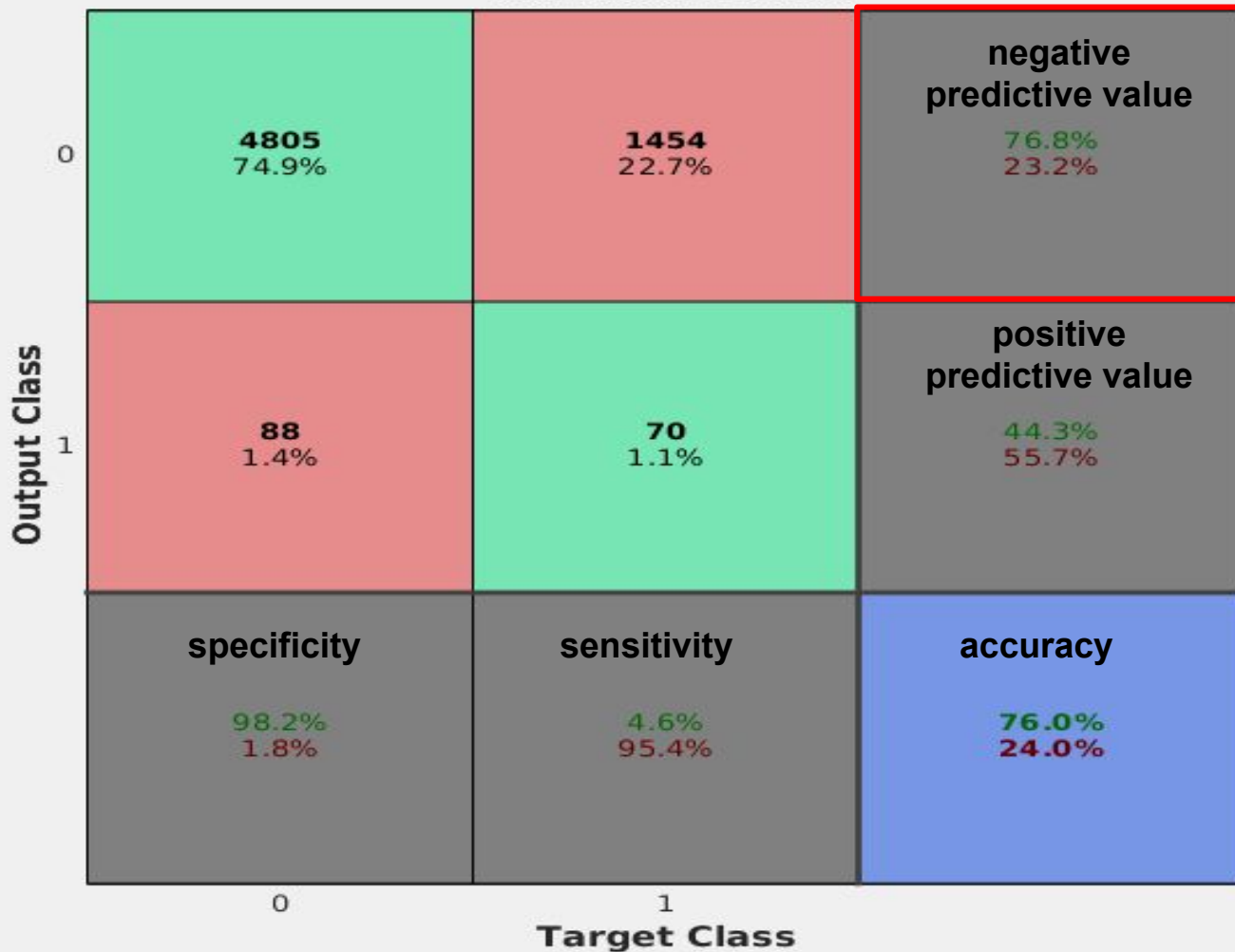


Sensitivity

- $Sensitivity = \frac{\text{true deaths}}{(\text{true deaths} + \text{false survivors})}$
- Given that a person is going to die, how often will the test indicate death?

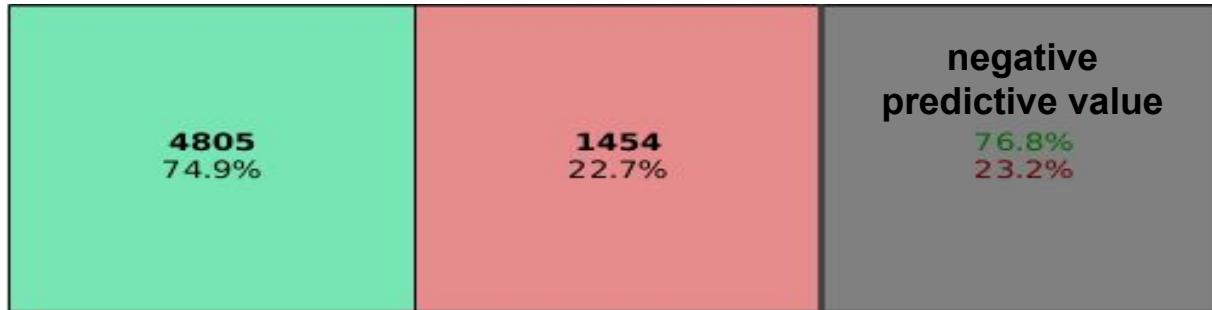


Confusion Matrix

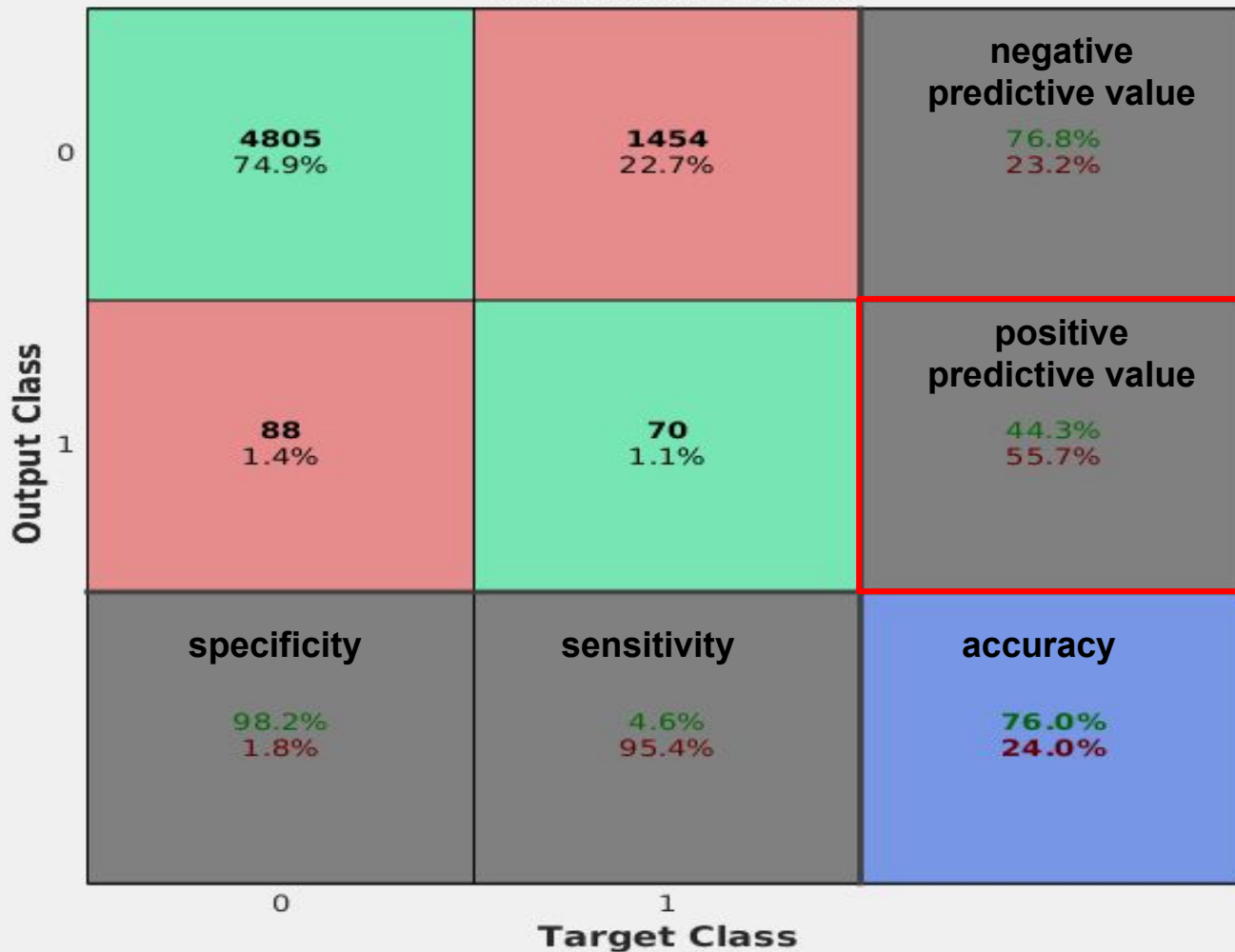


Negative Predictive value:

- $PV^- = \text{true survivals} / (\text{true survivals} + \text{false deaths})$
- Given that the test predicts survival, what is the probability that the patient actually survives?



Confusion Matrix



Positive Predictive value (PV+):

- $PV+= \text{true deaths} / (\text{true deaths} + \text{false deaths})$
- Given that the test predicts death, what is the probability that the patient will die?



If we compare our approach against the Optimistic Approach, using these other metrics, how do we do?

Optimistic

Confusion Matrix

	0	1	
0	4893 76.3%	1524 23.7%	negative predictive value 76.3% 23.7%
1	0 0.0%	0 0.0%	positive predictive value NaN% NaN%
	specificity	sensitivity	accuracy
	100% 0.0%	0.0% 100%	76.3% 23.7%
	0	1	
	Target Class		

Ours

Confusion Matrix

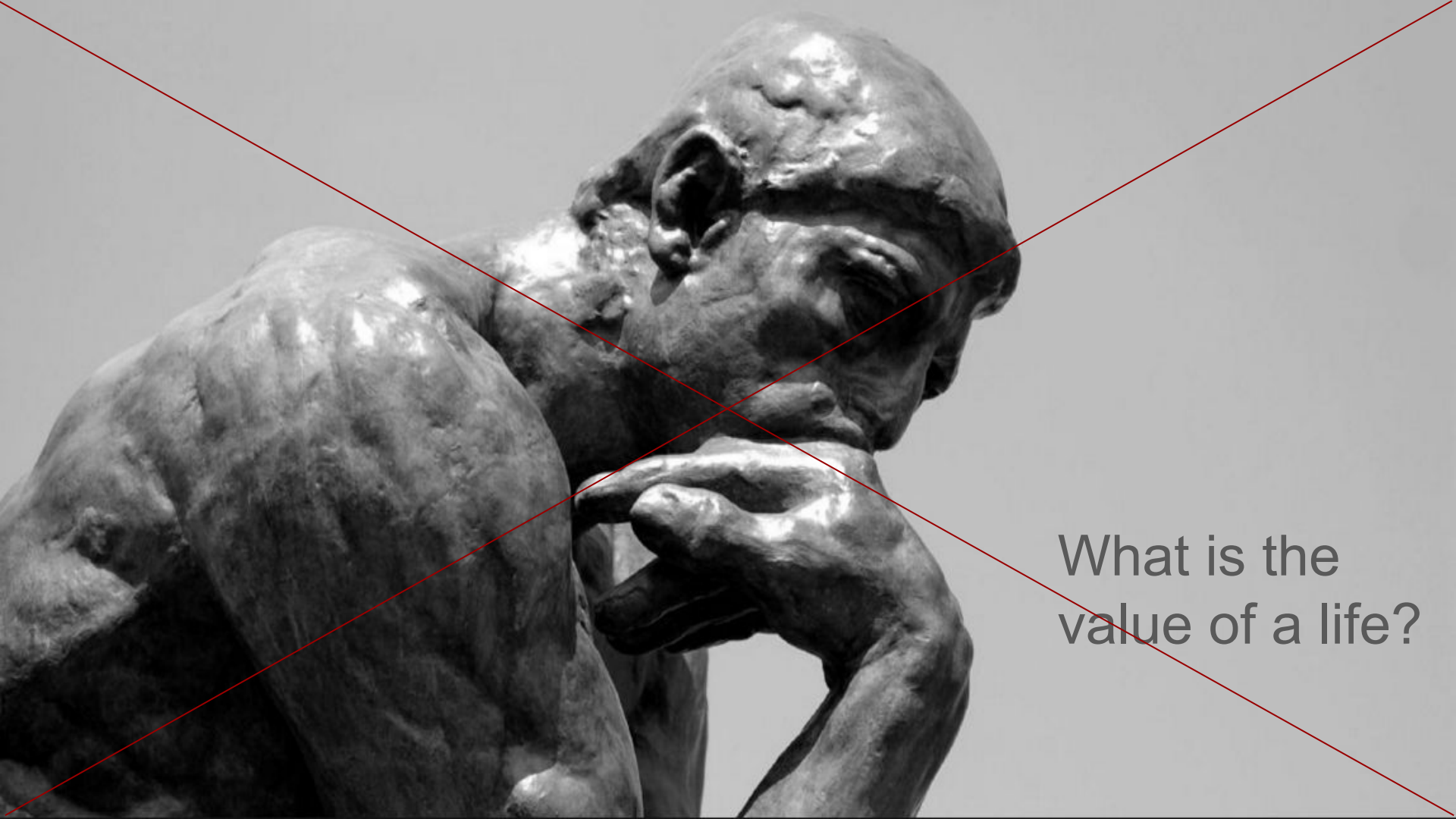
	0	1	
0	4805 74.9%	1454 22.7%	negative predictive value 76.8% 23.2%
1	88 1.4%	70 1.1%	positive predictive value 44.3% 55.7%
	specificity	sensitivity	accuracy
	98.2% 1.8%	4.6% 95.4%	76.0% 24.0%
	0	1	
	Target Class		

Given all the options, what is the standard performance metric when working with medical data?

- A good algorithm is both **sensitive** and **specific**.
- But deciding relative importance can be tricky
- Is it as bad to
 - predict survival given death, as it is to
 - predict death given survival?



What is the
value of a life?



What is the
value of a life?

- If everyone has an opinion about the relative importance of predicting survival vs. death, then let's add a range of biases to our predictions and see how it does.
- We can do this, by adding or subtracting a constant value from our predictions

Let's Bias The Optimist's Algorithm

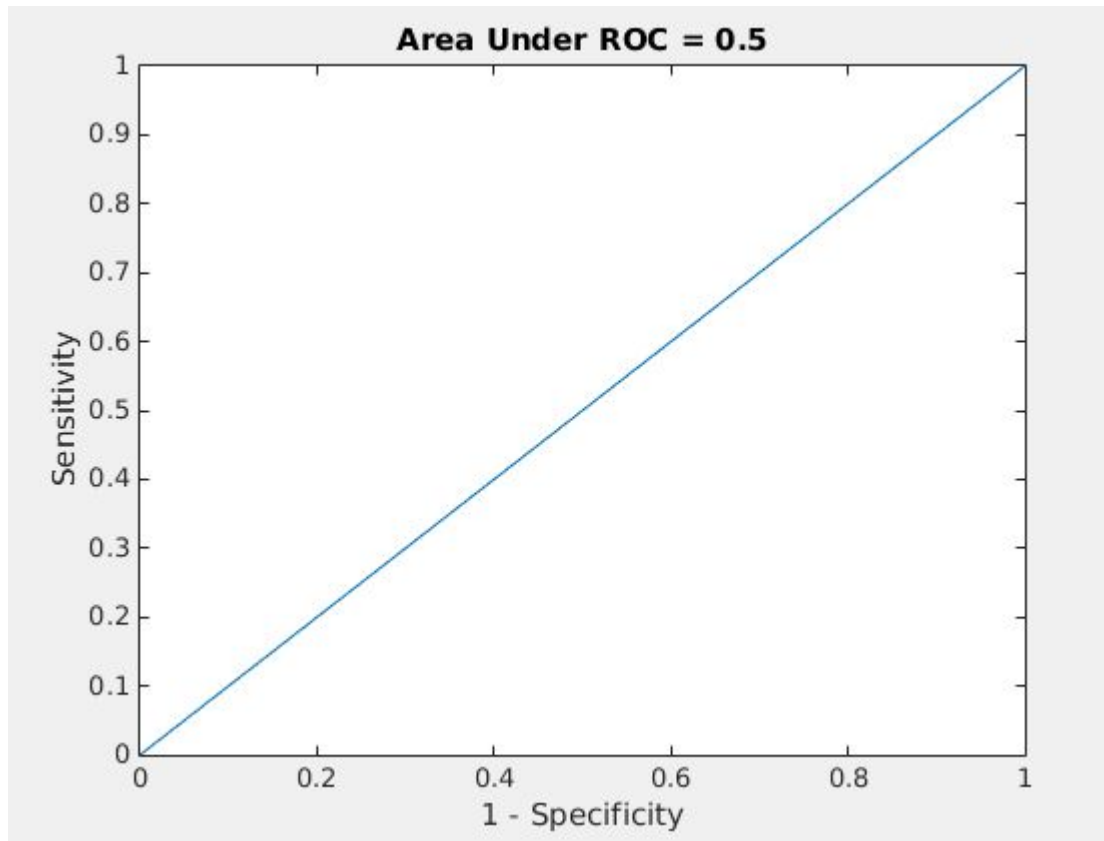
Optimist's Original Prediction	Slight Bias	Modified Prediction	Modified Prediction > 0.5	Outcome
0	.1	0.1	0	1
0	.1	0.1	0	0
0	.1	0.1	0	1
0	.1	0.1	0	0

Optimist's Original Prediction	Pessimist Bias	Modified Prediction	Modified Prediction > 0.5	Outcome
0	1	1	1	1
0	1	1	1	0
0	1	1	1	1
0	1	1	1	0

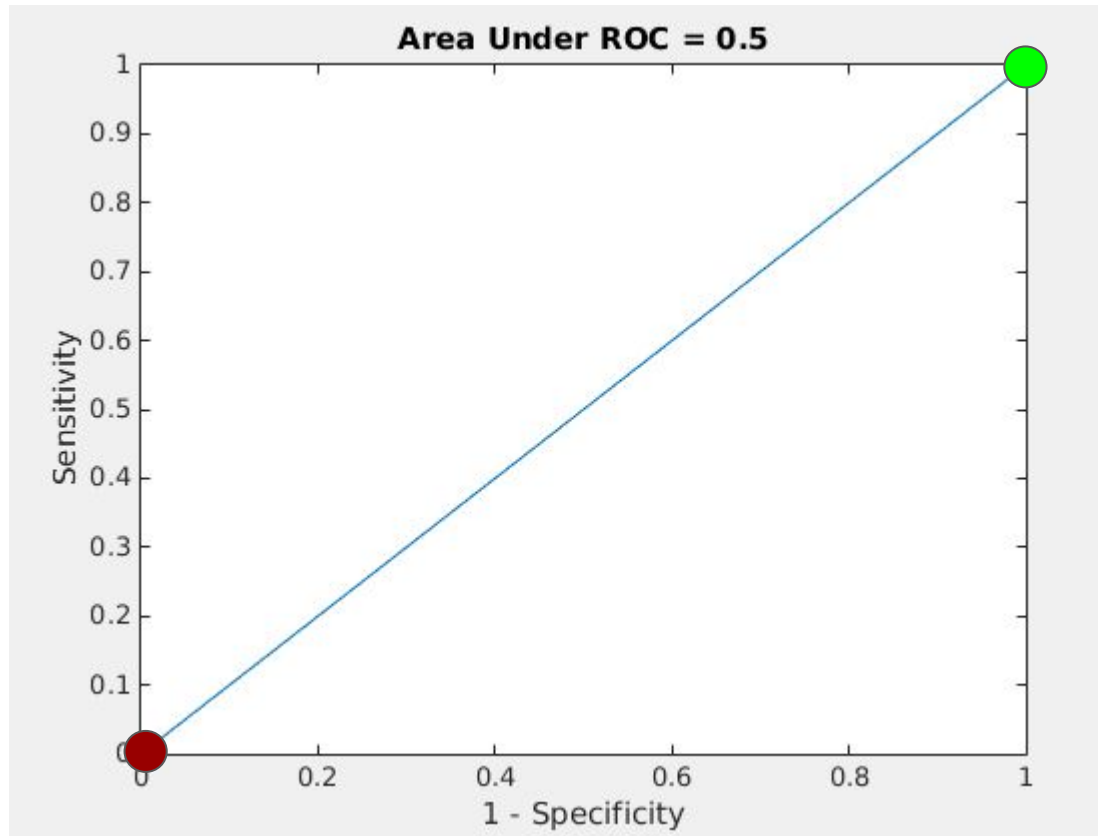
Optimist's Original Prediction	Optimist Bias	Modified Prediction	Modified Prediction > 0.5	Outcome
0	-1	-1	0	1
0	-1	-1	0	0
0	-1	-1	0	1
0	-1	-1	0	0

The Optimist's Algorithm

- Bias create two distinct predictions
- All 1s, or all 0s



The Optimist's Algorithm



If predicting survival given survival is infinitely valuable.

If predicting death given death is infinitely valuable.

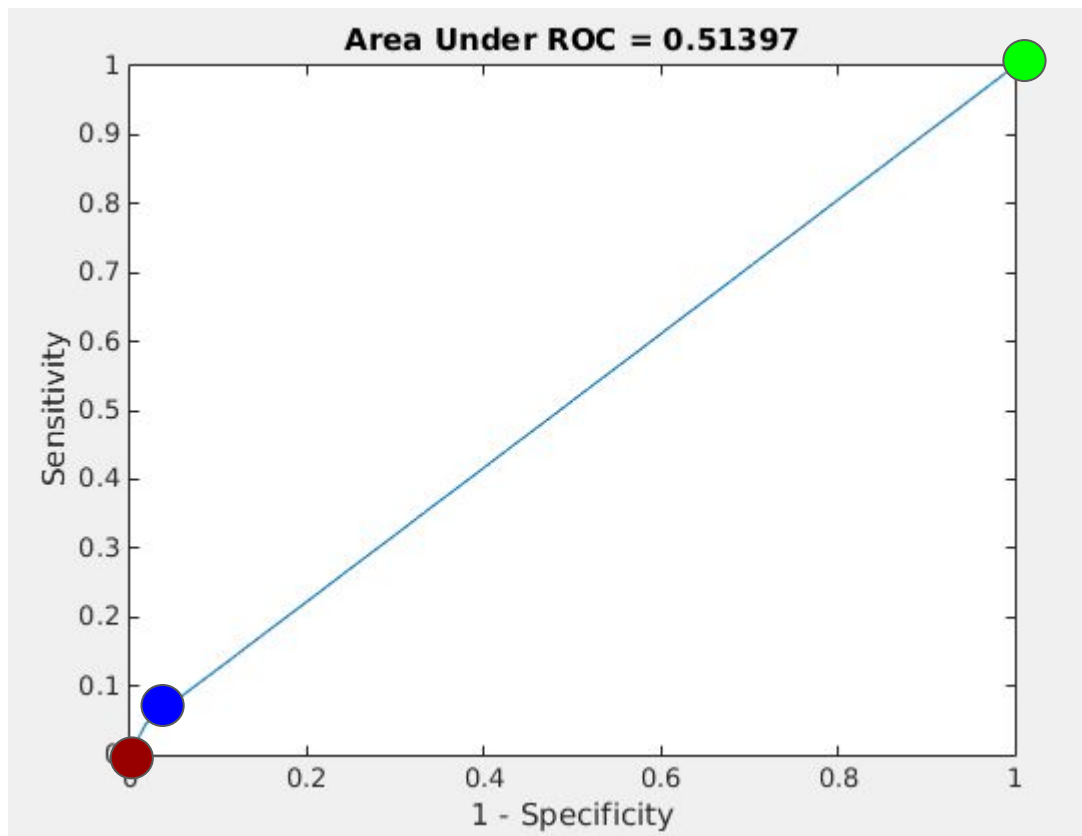
What about our Algorithm?

Our Original Prediction	Bias	Modified Prediction	Modified Prediction > 0.5	Outcome
1	.1	1.1	1	1
1	.1	1.1	1	0
0	.1	0.1	0	1
0	.1	0.1	0	0

Our Original Prediction	Pessimist Bias	Modified Prediction	Modified Prediction > 0.5	Outcome
1	1	2	1	1
1	1	2	1	0
0	1	1	1	1
0	1	1	1	0

Our Original Prediction	Optimist Bias	Modified Prediction	Modified Prediction > 0.5	Outcome
1	-1	0	0	1
1	-1	0	0	0
0	-1	-1	0	1
0	-1	-1	0	0

Our Algorithm: Three Options



If predicting survival given survival is infinitely valuable.

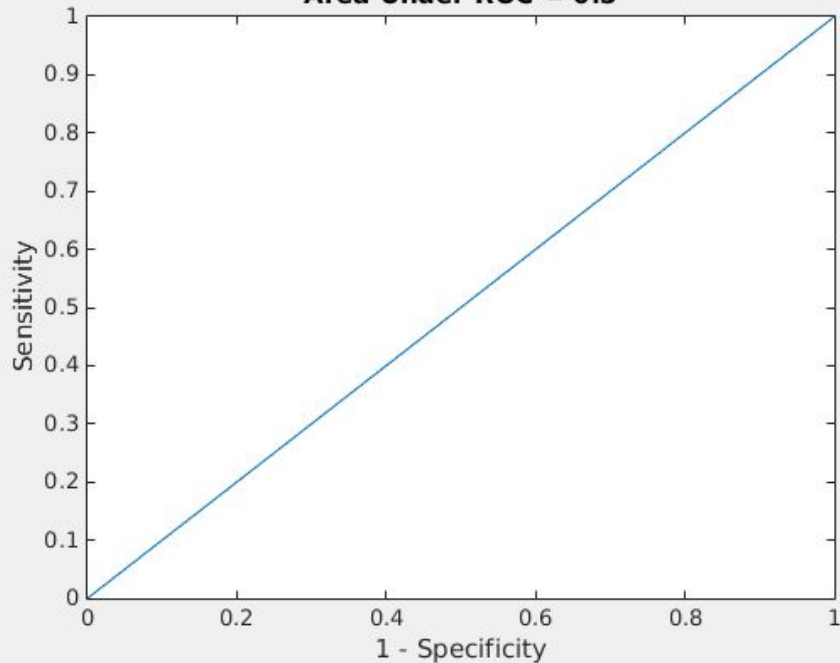
If predicting death given death is infinitely valuable.

How do we know which is better?

Compare Every Option via *Area Under the Curve*

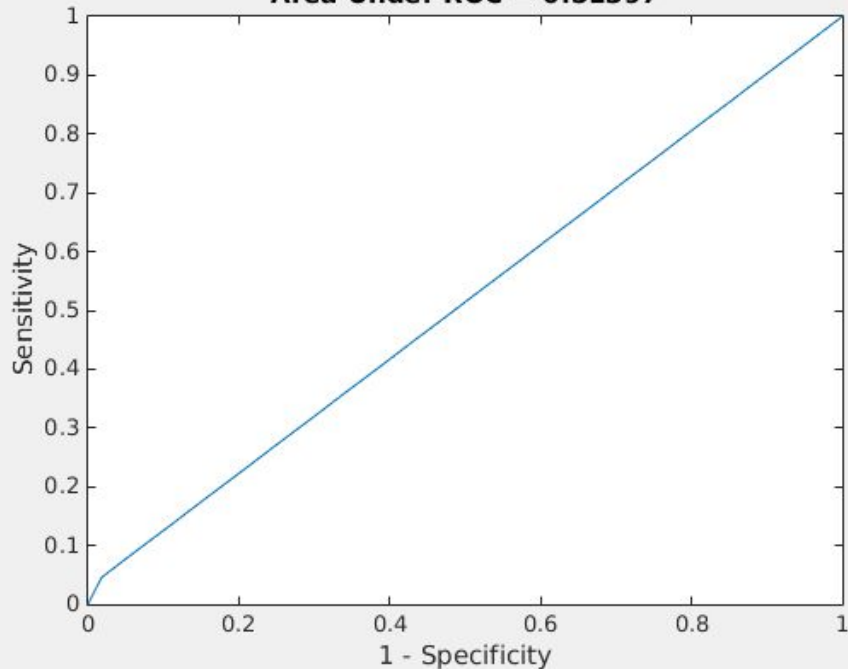
Optimist's

Area Under ROC = 0.5



Our Approach

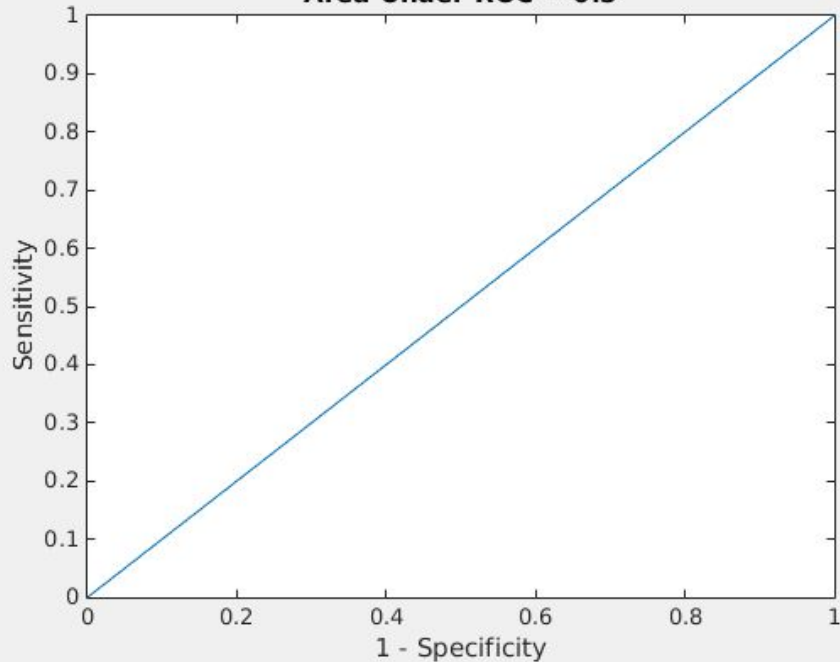
Area Under ROC = 0.51397



Compare Using the Area Under the Curve

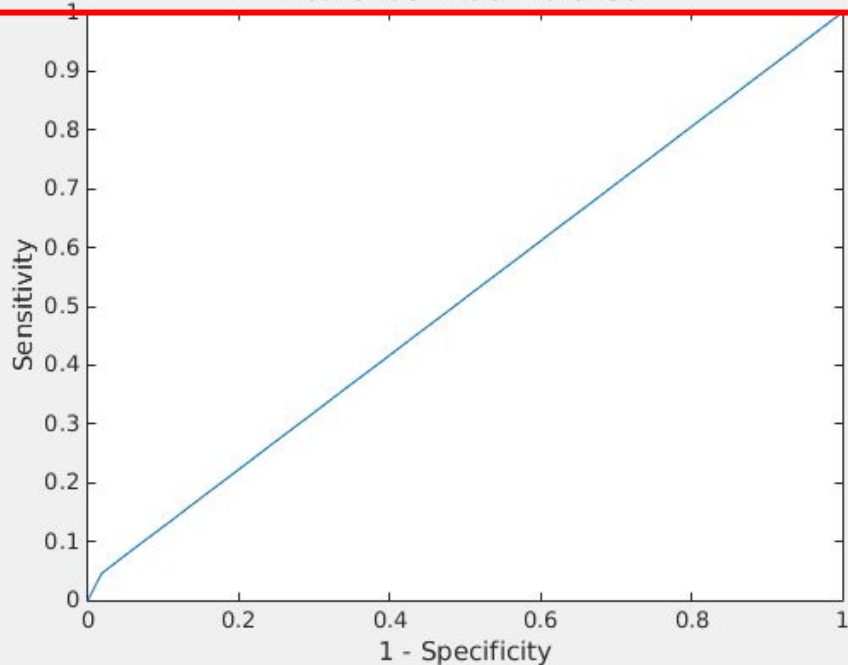
Optimist's

Area Under ROC = 0.5



Our Approach

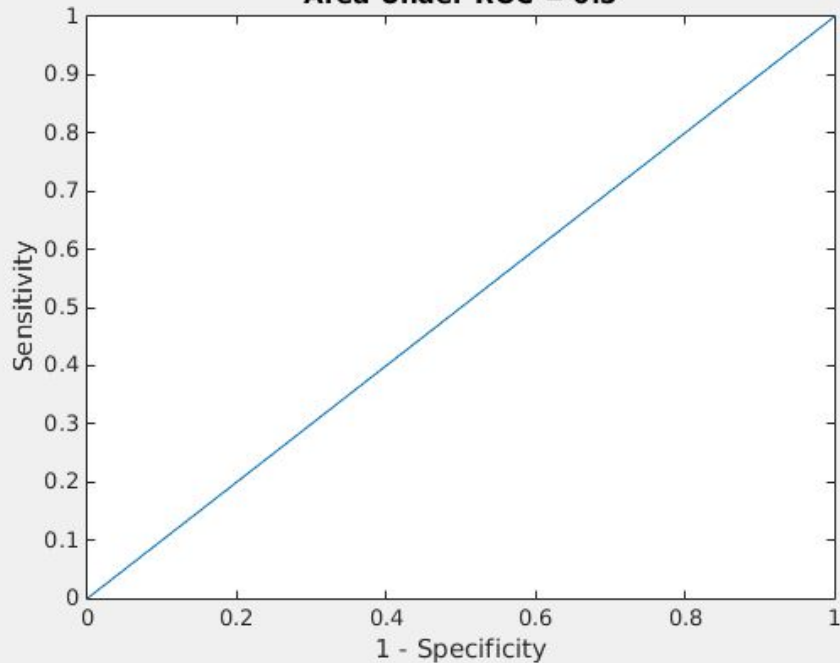
Area Under ROC = 0.51397



Compare Using the Area Under the Curve

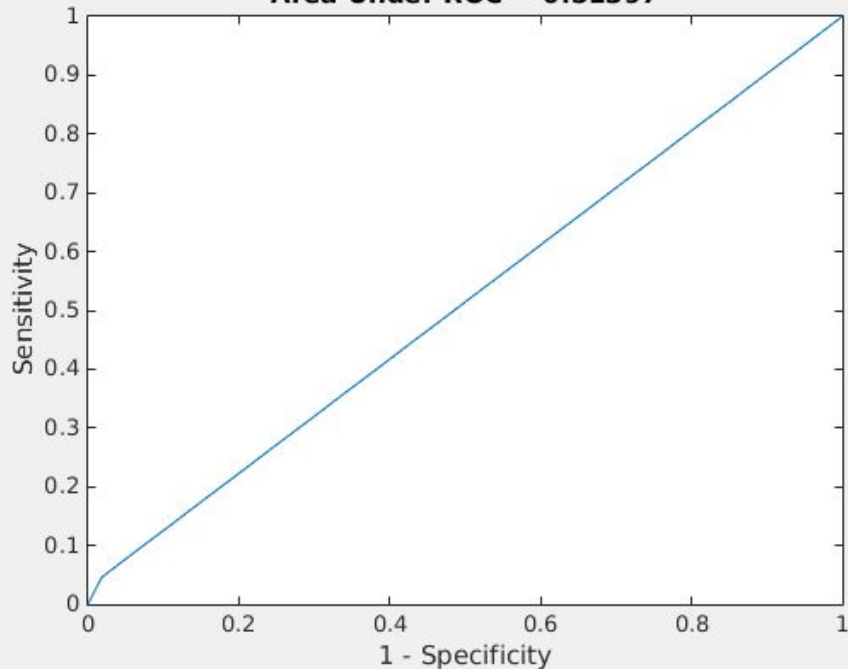
Optimist's

Area Under ROC = 0.5



Our Approach

Area Under ROC = 0.51397





**DON'T
KEEP CALM
IT'S
PARTY
TIME**

Wait... Why are we so similar to the optimist's algorithm?

Confusion Matrix

	0	1	negative predictive value 76.3% 23.7%
0	4893 76.3%	1524 23.7%	
1	0 0.0%	0 0.0%	positive predictive value NaN% NaN%
specificity	100% 0.0%	sensitivity	0.0% 100%
	0	1	accuracy 76.3% 23.7%
	Target Class		

Confusion Matrix

	0	1	negative predictive value 76.8% 23.2%
0	4805 74.9%	1454 22.7%	
1	88 1.4%	70 1.1%	positive predictive value 44.3% 55.7%
specificity	98.2% 1.8%	sensitivity	4.6% 95.4%
	0	1	accuracy 76.0% 24.0%
	Target Class		

Requiring all conditions to
be satisfied is too rare.

Democracy to the rescue!

Let's Make a BETTER Better Black Box

' ICUSTAYID'
' OUTCOME'
' Age'
' HeartRateMin' →
' HeartRateMax' →
' MeanBPMin'
' MeanBPMax'
' RespRateMin'
' RespRateMax'
' GCSMin'
' GCSMax'
' BILIRUBINmin'
' BILIRUBINmax'
' CREATININEmin
' CREATININEmax
' HEMOGLOBINmin
' HEMOGLOBINmax
' SODIUMmin'
' SODIUMmax'
' WBCmin' →
' WBCmax' →



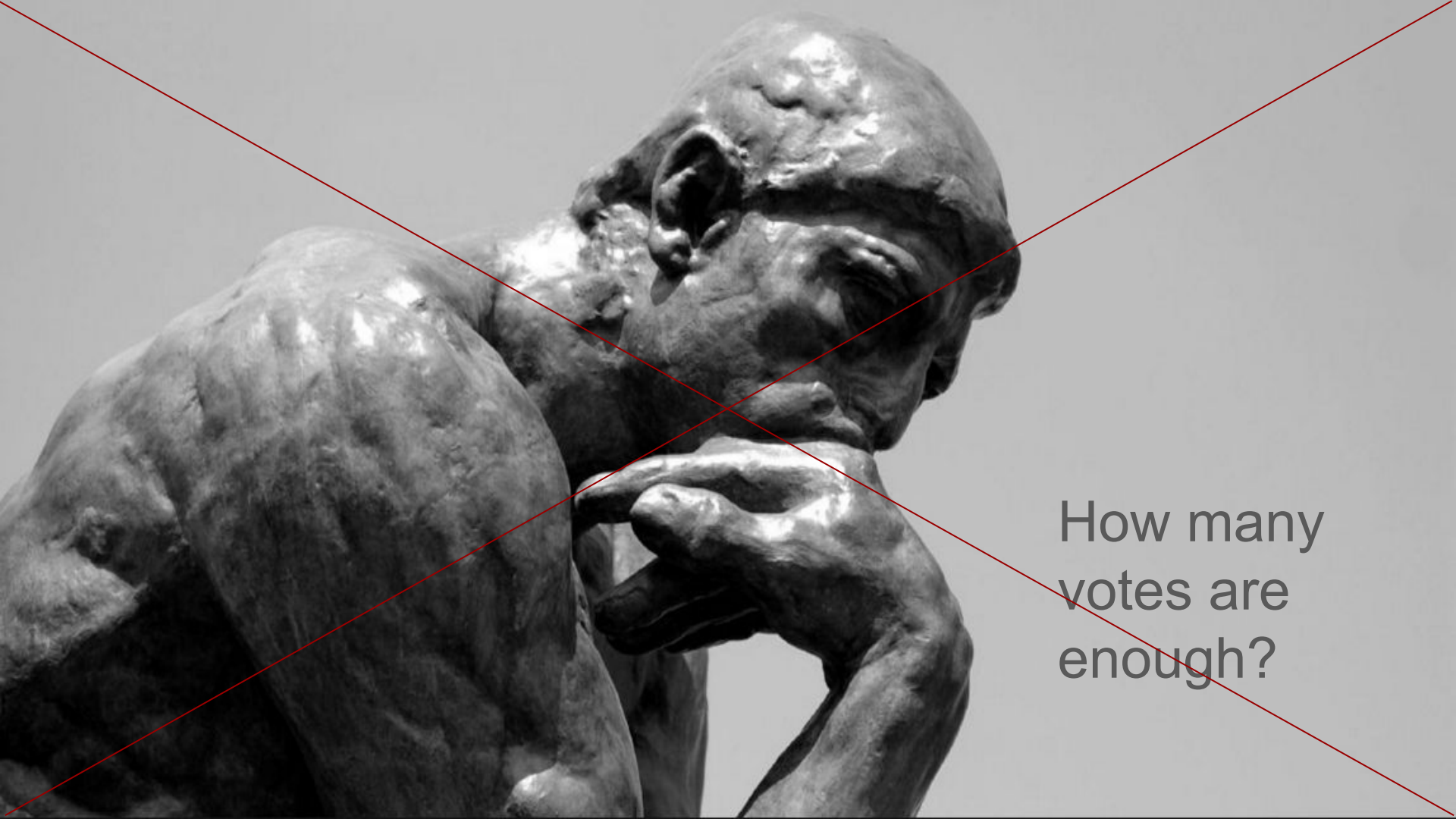
Inside The BETTER Better Black Box

```
votes = ( (Age < 30) +  
          (Age > 75) +  
          (WBCmax > 15) +  
          (WBCmin > 15) +  
          (HeartRateMax > 125) +  
          (HeartRateMin > 100) +  
          (HeartRateMin < 40)  
        ) / 7
```

```
prediction = death if votes > ...
```

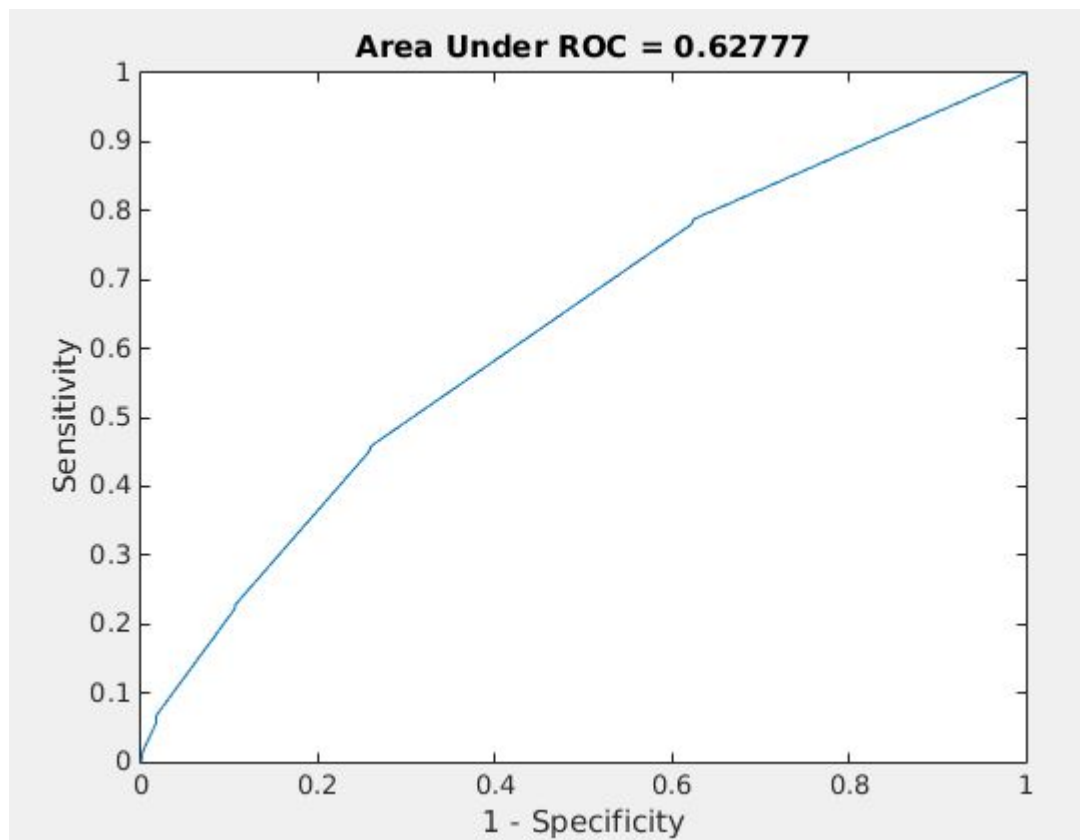


How many
votes are
enough?



How many
votes are
enough?

The Democracy Algorithm



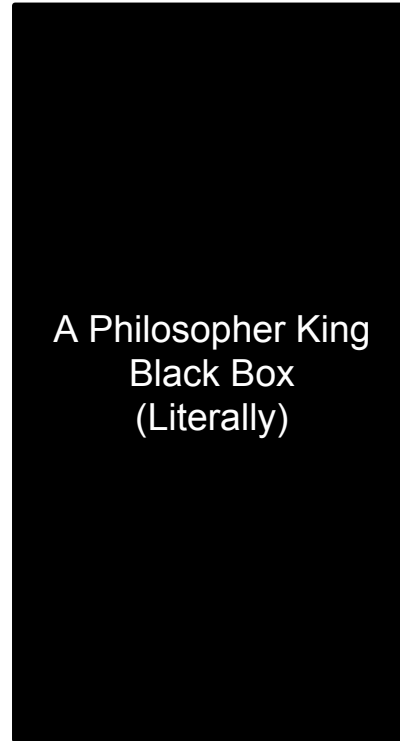
But can we do better...

Perhaps not all features are created equal.

Aristocracy?

Let's Make a Philosopher King Black Box

' ICUSTAYID'
' OUTCOME'
' Age'
' HeartRateMin' →
' HeartRateMax' →
' MeanBPMin'
' MeanBPMax'
' RespRateMin'
' RespRateMax'
' GCSMin'
' GCSMax'
' BILIRUBINmin'
' BILIRUBINmax'
' CREATININEmin
' CREATININEmax
' HEMOGLOBINmin
' HEMOGLOBINmax
' SODIUMmin'
' SODIUMmax'
' WBCmin' →
' WBCmax' →

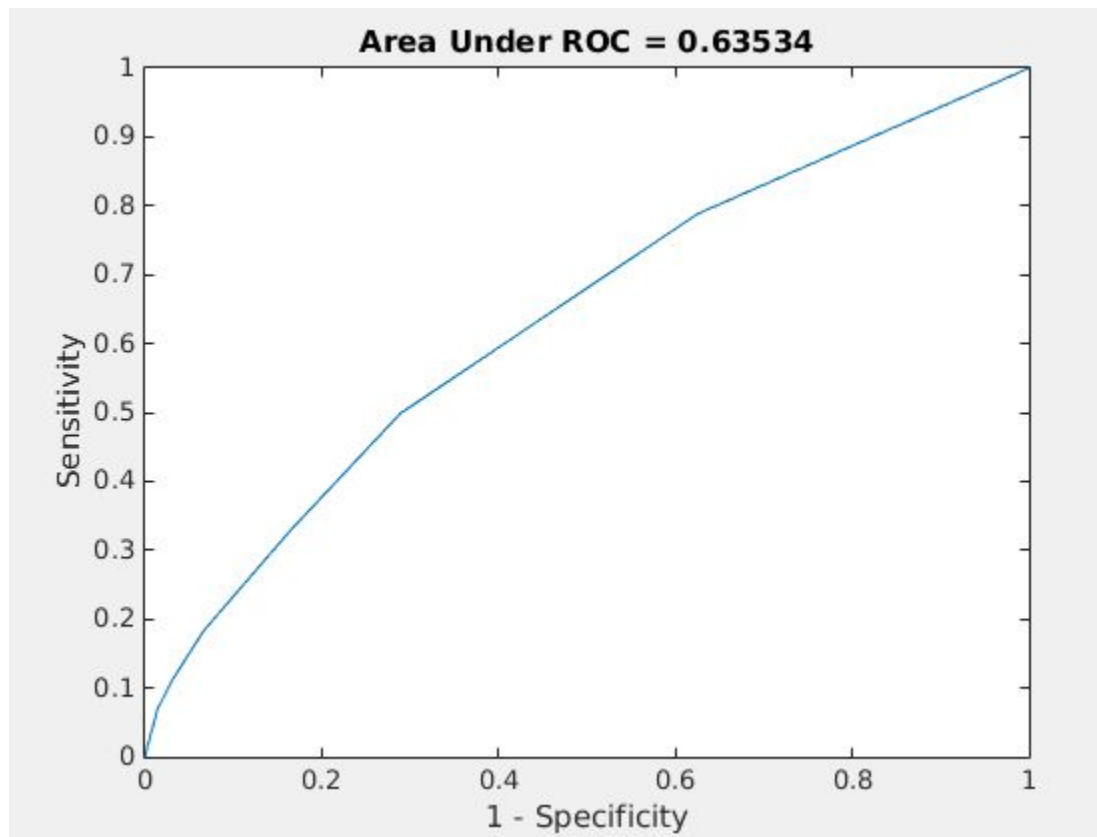


Inside The Philosopher King's Black Box

$$\begin{aligned} \text{votes} = & 4 * (\text{Age} < 30) + \\ & 1 * (\text{Age} > 75) + \\ & 1 * (\text{WBCmax} > 15) + \\ & 1 * (\text{WBCmin} > 15) + \\ & 1 * (\text{HeartRateMax} > 125) + \\ & 3 * (\text{HeartRateMin} > 100) + \\ & 3 * (\text{HeartRateMin} < 40)) / 14 \end{aligned}$$

prediction = 1 if votes > ...

The Philosopher King Algorithm... Plato was right...



What can we conclude

- The Optimist's Approach
 - Area under the curve = 0.50
- Our First Approach
 - Area under the curve = 0.51
- Democracy
 - Area under the curve = 0.62
- Aristocracy
 - Area under the curve = 0.64

What can we conclude

- The Optimist's Approach
 - Area under the curve = 0.50
- Our First Approach
 - Area under the curve = 0.51
- Democracy
 - Area under the curve = 0.62
- Aristocracy
 - Area under the curve = 0.64



Plato



Socrates



Aristotle

Before we get carried
away...

What Has Human Learning Taught Us?

- Features with distributions that separate the outcomes of interest do a good job in classification.
- If done correctly, combining multiple weak features can make for a strong classifier.
- Not all features are equally important, some should be considered more strongly than others.

Here's how it connects to Machine Learning

- ML Identifies features that separate the cases of interest, automatically!
- ML combines multiple weak features, automatically.
- ML considered some features more strongly than others, automatically!

Are we done?



Wait!
What about Jim!

Jim's ICUSTAY_ID - 290851

- Prediction is 0?
 - Could there be a bug?

- Let's Investigate