Bhaarat Sharma
bsharma30@gatech.edu

1 Logistic Regression

1.1 Batch Gradient Descent

a.

$NLL\,(w) = -\sum_{i=1}^{N}\Big[(1-y_i)\log(1-\sigma(\mathbf{w}^T\mathbf{x})) + y_i\log\sigma(\mathbf{w}^T\mathbf{x})\Big]$
$-\sum_{i=1}^{N}[y_i\mathbf{w}^T\mathbf{x} - \log(1+e^{\mathbf{w}^T\mathbf{x}})]$
$-\sum_{i=1}^{N}\mathbf{x}(y_i - \sigma(\mathbf{w}^T\mathbf{x}))$

1.2 Stochastic Gradient Descent

a.

$$= \Big[(1-y_t)\log(1-\sigma(\mathbf{w}^T\mathbf{x_t})) + y_t\log(\mathbf{w}^T\mathbf{x_t})\Big]$$

b.

$$w_t = w_t + \eta((y_t - \sigma(\mathbf{w}^T\mathbf{x_t}))x_t)$$

c.

$$log(n)$$

d.

Very large values of η make ROC curve closer to the middle whereas smaller values of η push it further away from the middle.

e.

$$w_t = w_t - \eta((y_t - \sigma(\mathbf{w}^T\mathbf{x_t}))x_t - \mu\|\mathbf{w}\|)$$

2

2.1

Metric	Deceased	Alive Patients
Event Count (Average)	682.64	1029.05
Event Count (Max)	12627	16829
Event Count (Min)	1	2
Encounter Count (Average)	18.66	24.86
Encounter Count (Max)	391	375
Encounter Count (Min)	1	1
Record Length (Average)	194.65	151.39
Record Length (Max)	3103	2601
Record Length (Min)	0	0
Common Diagnostics	DIAG320128 1019 DIAG319835 721 DIAG317576 719 DIAG42872402 674 DIAG313217 641	DIAG320128 415 DIAG319835 413 DIAG313217 374 DIAG197320 346 DIAG132797 297
Common Laboratory	LAB3009542 66910 LAB3000963 57733 LAB3023103 56967 LAB3018572 54667 LAB3007461 53548	LAB3009542 32747 LAB3023103 28376 LAB3000963 28288 LAB3018572 27364 LAB3016723 27041
Common Medication	DRUG19095164 12452 DRUG43012825 10388 DRUG19049105 9329 DRUG19122121 7586 DRUG956874 7294	DRUG19095164 6394 DRUG43012825 5446 DRUG19049105 4323 DRUG956874 3962 DRUG19122121 3908

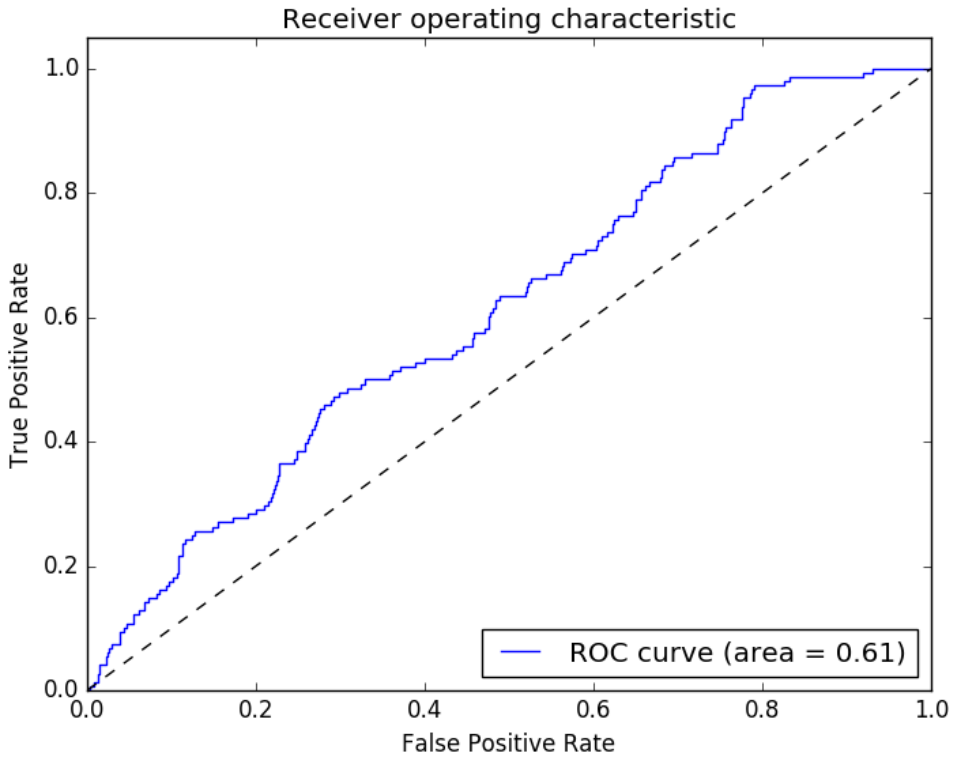
2.3 b.

See figures 1 - 5. A high η learning parameter leads to worse results. I believe this is because of overfitting. Tuning a low η and randomization produces decent results. However, it is interesting to note how the curve starts close to the dotted line, starts to shift away, and then finally merges with the dotted line again.

2.4.b

See figure 6

Figure 1: Learning 4.0 — Regularization .00005



2.4 c.

The two curves are very similar to each other. I believe this is because logistic regression is being used in both

Figure 2: Learning 9.0 — Regularization .00001

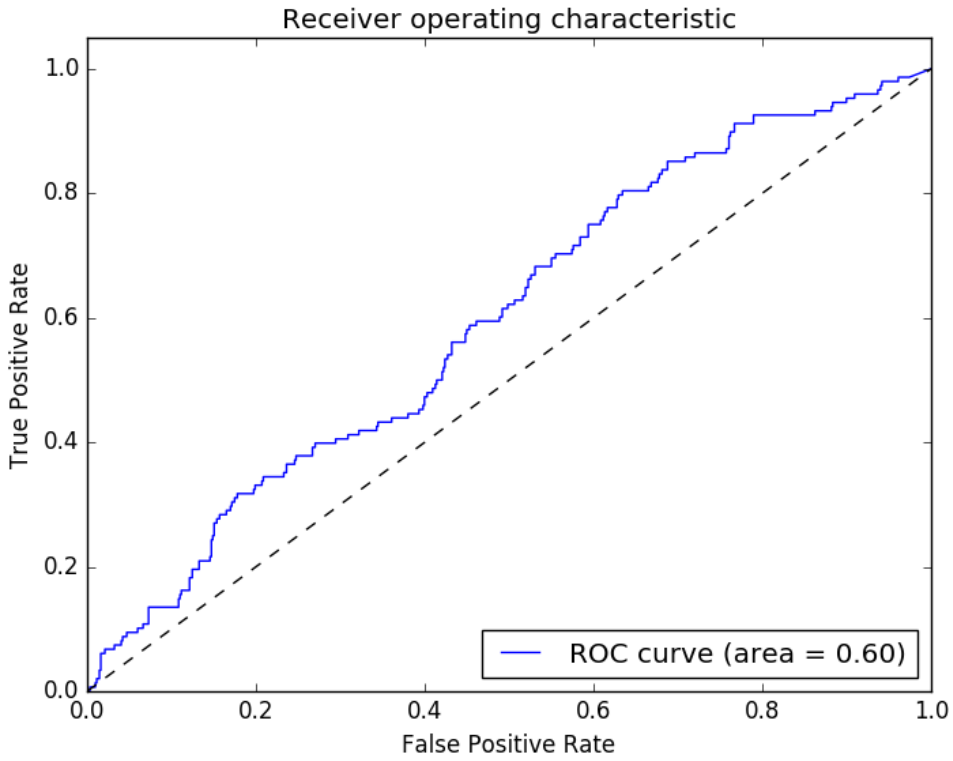


Figure 3: Learning .01 — Regularization .01

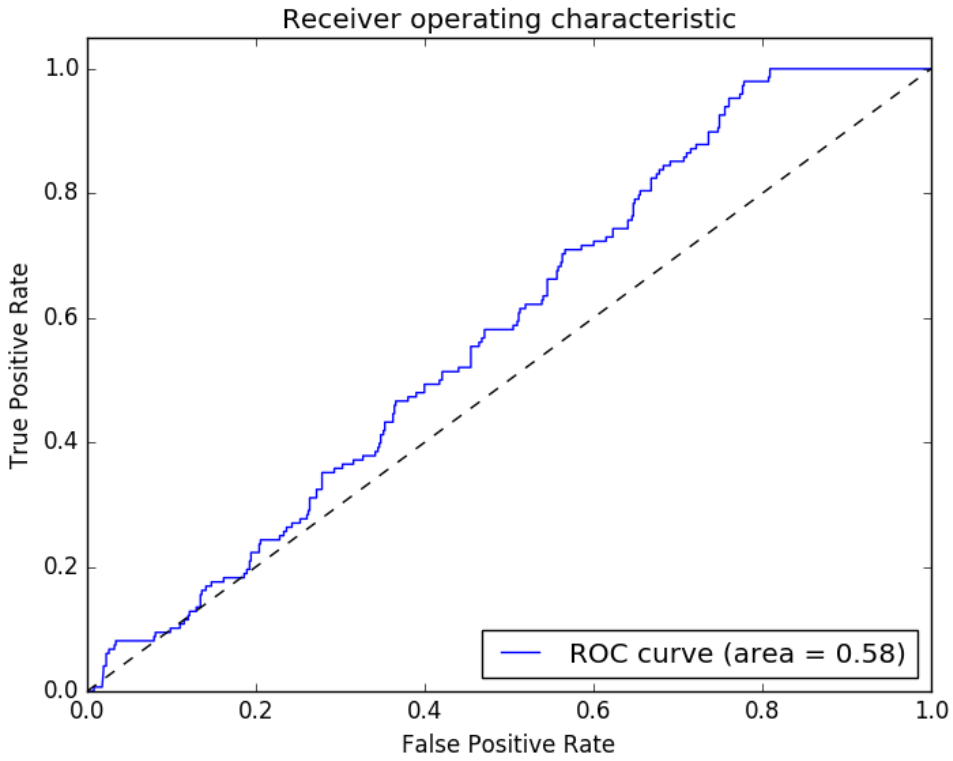


Figure 4: Learning .01 — Regularization .05

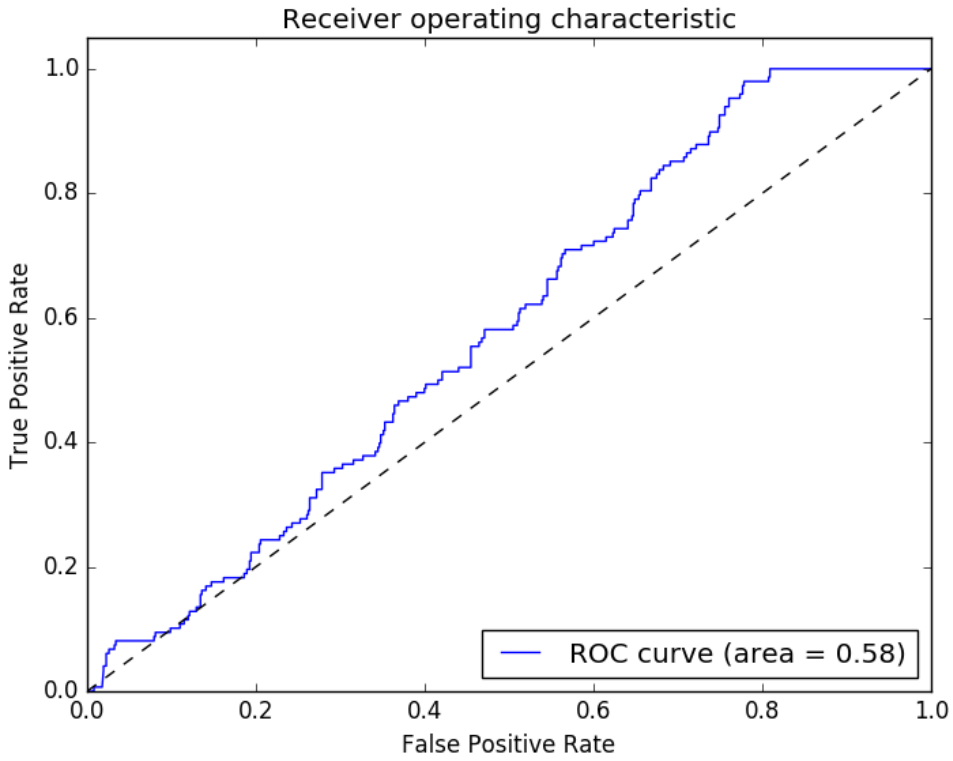


Figure 5: Learning .50 — Regularization .001

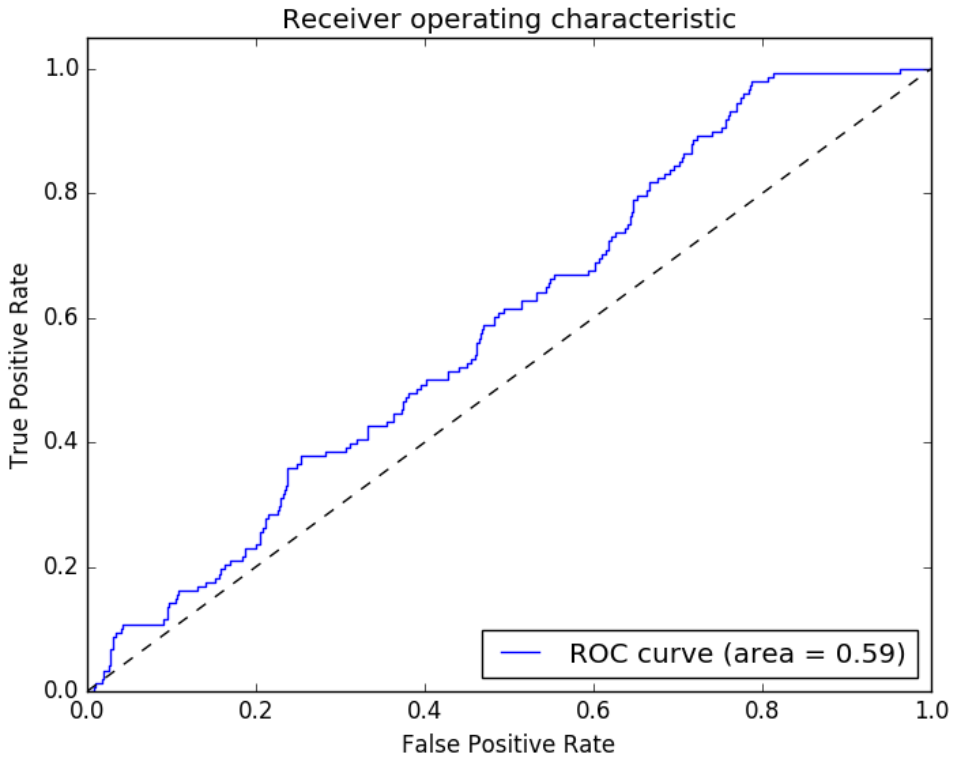


Figure 6: ROC Curve by testensemble.py