

Noah Breton 101232464
Marco Toito 101233884

1d)

Train

Model	L2 Loss	L1 Loss	Linf Loss
L2 Model	0.01662241	0.14495528	0.41709409
L1 Model	0.01883233	0.1367509	0.49488601
Linf Model	0.02195039	0.17581582	0.31620598

Test

Model	L2 Loss	L1 Loss	Linf Loss
L2 Model	0.02503494	0.17823322	0.75610414
L1 Model	0.02861608	0.19013652	0.80912286
Linf Model	0.03137868	0.19848521	0.85061708

1e)

All Linf losses are much larger, as they are skewed towards outliers

L2 loss is larger than L1 because more sensitive to outliers than L1, since it grows exponentially vs abs value

In training: Linf loss works the best with Linf model, due to model minimizing worst outlier and loss checking the worst outlier, this will match up better than L2 or L1 loss

In testing: Linf loss works worst with Linf model since it overfits to the largest errors, it will be pulled away from the rest of the data. but if that worst outlier is not as bad as in the training data, it will perform poorly.

Linf loss works well with the L1 and L2 Models compared to Linf loss since they won't overfit the error, but are still outperformed by the other combinations.

2d)

m	Train Accuracy	d	Train Accuracy	eta	Train Accuracy
10	0.9435	1	0.8422	0.1	0.9223
50	0.9227	2	0.92385	1.0	0.92625
100	0.924	4	0.98035	10	0.9217
200	0.9226	8	0.9994	100	0.89915

m	Test Accuracy	d	Test Accuracy	eta	Test Accuracy
10	0.901675	1	0.841425	0.1	0.91957
50	0.917045	2	0.920555	1.0	0.919735
100	0.91904	4	0.973945	10	0.920225
200	0.92144	8	0.997175	100	0.8943

2e)

for training, the more data points (the larger m is) accuracy plateaus around 0.92

for test, the more data points the more accurate prediction

having more data points helps the model generalize better

for training, the more features (d) allows it to separate the classes into more distinct groups

for test, the more features (d) also allows it to separate into more distinct groups

for small training step size (0.1, 1, 10), the model is fairly accurate

when training rate is too high, it starts to overshoot predictions

Q3.B

Train

Model	L2 Loss	L1 Loss	Linf Loss
L2 Model	5.78391296	2.60944142	13.02634622
L1 Model	6.16741665	2.5432599	14.80431039
Linf Model	10.32429758	3.71144018	9.85051766

Test

Model	L2 Loss	L1 Loss	Linf Loss
L2 Model	5.80625802	2.62708668	13.00012091
L1 Model	6.21148704	2.56804612	14.7868502
Linf Model	10.23087832	3.69124104	9.85072718

Q3.D

eta	Train Accuracy	Validation Accuracy	Test Accuracy
0.1	0.86746988	0.78313253	0.69047619
1.0	0.93975904	0.75903614	0.73809524
10	0.92771084	0.73493976	0.66666667
100	0.93975904	0.73493976	0.5952381

We recommend using 0.1 learning rate. In the validation data, 0.1 performs the best in terms of accuracy