COMP 3105 — Fall 2025 — Assignment 1

Pascal Law 101318299

Question 1:

Subquestion b1.

$$\underset{1\times(d+1)}{\mathbf{c}^T} \underset{(d+1)\times 1}{\mathbf{u}} = \delta$$

Since $\mathbf{u} = \begin{bmatrix} \mathbf{w} \\ \delta \end{bmatrix}$, we can make all the weights in $\mathbf{w}$ zero, and just have δ in the last entry of $\mathbf{u}$.

As such, we can set:

$$\mathbf{c} = \begin{bmatrix} 0 & 0 & \underset{(d+1)\times 1}{0} & \cdots & 0 & 1 \end{bmatrix}^T$$

Subquestion b2.

Change the first part of inequality of L_∞ to:

$$-\delta \leq 0$$

We can break $G^1\mathbf{u} \leq 0$ into two parts:

Subquestion c2.

Table 1: Different training losses for different models

Model	L_2 loss	L_∞ loss
L_2 model	0.09557022	1.58819212
L_∞ model	0.23302085	0.86917599

Table 2: Different test losses for different models

Model	L_2 loss	L_∞ loss
L_2 model	0.05291075	1.03171424
L_∞ model	0.29564879	2.1535331

Analysis of result:

When we compare our training data with our test data, we find that our L_2 loss has a smaller difference when compared to our L_∞ loss, and similarly our L_2 model also has a smaller difference when compared to our L_∞ model.

This means that our L_2 model using L_2 loss has the smallest difference and so is the most accurate prediction of the data. On the other hand, the L_∞ model has the biggest difference which means that L_∞ model using L_∞ loss is the most inaccurate prediction of the data.

The reason that the L_∞ loss is so inaccurate is because of the difference of size between the test data and the training data. We only generate 30 samples for the training while we generate 1000 samples for the test data. This means that our L_∞ loss is modelled on a very limited set of training data, and so it is heavily affected by outliers in our much larger test data.

This is in contrast with our L_2 loss, which is not as affected by outliers, which makes it much more accurate when the training data sample is small.

When considering the models, the L_2 model is more accurate because it has a smaller tolerance, which means it doesn't chase outliers like our L_∞ model.

These factors lead to the conclusion that our L_2 loss and L_2 model are more accurate than the L_∞ counterparts for this set of training and test data.

Subquestion d2.

Table 1: Different training losses for different models

Model	L_2 loss	L_∞ loss
L_2 model	53.18368801	32.21643524
L_∞ model	66.09571725	26.39169568

Table 2: Different test losses for different models

Model	L_2 loss	L_∞ loss
L_2 model	55.20252554	33.40449987
L_∞ model	68.04960819	36.9655386

Question 2:

Subquestion c1.

Table 5: Training accuracies with different hyper-parameters

m	**Train Accuracy**	**dim1**	**Train Accuracy**	**dim2**	**Train Accuracy**
10	0.9725	1	0.8467	1	0.9266
50	0.9252	2	0.9262	2	0.9275
100	0.9217	4	0.98285	4	0.9255
200	0.923125	8	0.99935	8	0.9349

Table 6: Testing accuracies with different hyper-parameters

m	**Test Accuracy**	**dim1**	**Test Accuracy**	**dim2**	**Test Accuracy**
10	0.882065	1	0.836885	1	0.91926
50	0.911655	2	0.91675	2	0.915435
100	0.91711	4	0.968755	4	0.915865
200	0.91836	8	0.996295	8	0.90904

Subquestion c2.