DATA7703 Tutorial 10

2021 Semester 2

1. Discuss the differences between an outlier and an adversarial example.

2. Describe three approaches for robustly learning an inlier model in the presence of outliers.

3. We consider robust regression on a small dataset in this question. We will work with the following small dataset.

x	1	2	4	5
y	2	3	9	6

(a) List the inliers and outliers in this dataset.

(b) Find the OLS model fitted on the entire dataset and the OLS model fitted using the inliers only.

(c) What are the ℓ_1 losses (that is, mean absolute errors) for the two models in (b)? Which model is better in terms of the ℓ_1 loss?

(d) What are the Huber losses for the two models in (b)? Do this for $\delta = 0.1$ and $\delta = 1$ respectively. Comment on which model is better in terms of Huber loss.

(e) Compute the Theil-Sen estimator for the following dataset. Use all pairs of points in your calculation. Compare the inlier OLS model and the Theil-Sen model.

4. We take a closer look at the RANSAC algorithm in this question.

(a) In lecture, we discussed about using R^2 to choose the best candidate inlier model. Is this the same as choosing the candidate inlier model with minimum MSE? Explain your answer.

(b) In practice, we usually consider a candidate inlier model only when it is trained on a sufficiently large number of inliers. Discuss how this can be helpful.