

HW1

Question 1 (20 points)

When missing value problem happens in the dataset, what are the methods to resolve this? For each method, please list the pros and cons.

Question 2 (15 points)

What are the three types of data shift? For each one, how could we resolve the problem.

Question 3 (30 points)

For small sample data problem, one problem would be due to imbalanced data. What are the approaches to resolve the imbalanced data problem? For each approach, please explain the general theory / method.

Question 4 (35 points)

For the two algorithms we learned, linear regression and random forest, please describe the theory, pros and cons of each method as well as come up a real-life application for each method.