

Quiz 2

1. **(2 points)** Which plot below can be used for exploring the correlation between two features?

- (a) Line plot
- (b) Hist plot
- (c) Scatter plot**

2. **(2 points)** Boxplot can show the feature's minimal value, median value, and maximal value.

- (a) True**
- (b) False

3. **(2 points)** Z-score normalization is to do $\frac{x - x_{\min}}{x_{\max} - x_{\min}}$.

- (a) True
- (b) False**

4. **(2 points)** Assume we have a categorical feature in a dataset, and there are three categories in this feature: [Apple, Orange, Banana]. Please use the One-Hot encoding to convert these three categories into numerical values.

- (a) Apple: [1, 0, 0]
- (b) Orange: [0, 1, 0]
- (c) Banana: [0, 0, 1]

5. **(2 points)** For the numerical feature, how to deal with its missing values?

Long tail distribution: fill in missing values with median value

Non-long tail distribution: fill in missing values with either median or mean value.