**Introduction to Machine Learning for Materials Science Lab**

**Quiz Questions**

1.1) Regression

1.2) Features/Descriptors

2.1) Missing data, Duplicate Data, mislabelled data.

Note: I’d accept other reasonable answers here. Definitely not limited to just these 3.

3.1) 10 percent is test data

4.1) sub-sampling data regions that are overly populated

Note: Other acceptable answers could be other strategies that remove bias from dataset like oversampling. Or weighting datapoints based on frequency

5.1) Citrination and MASTML