IBMS Dry Lab 答题总结

Data collection and preprocess

- Collect basic information
- feature engineering: identification of the traing features and predict label
- Preprocess: normalization, data type transformation
- If we need more data/ information, we could design a experiment to get them

Model selection

- Deep-learning based model
- For that you will need a test set to make sure your model is able to predict the outcome. Keep a subset (~20%) of the data separate to test your model (Test set). And other subset (~80%) of the data seperate to train set
- Can conduct several models, to choose the better model for this task
- You should be aware of not overfitting your model: including more and more variables to your models is not necessarily good. Your model may be robust to variables-associated noise.

Wet lab verification

- Use wet lab test to verify our model