

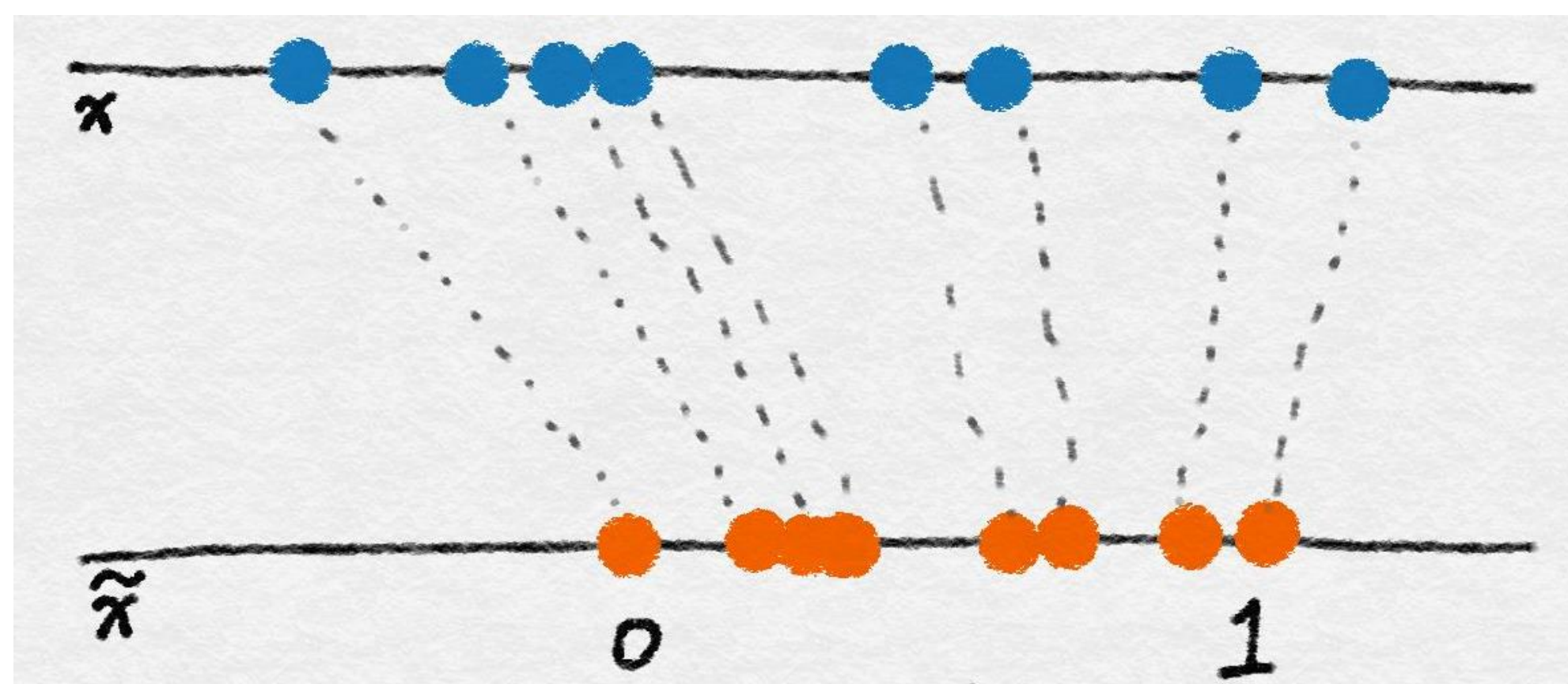
第八课 机器学习简单介绍

第十节 数据预处理及特征工程

主 讲: Robin

数据预处理及特征工程

- 数值型特征，如：长度、宽度、像素值等
 - **数值范围归一化** (feature normalization) **可能会**提高模型的性能，如：
线性回归，kNN，SVM，神经网络等
- 最大最小归一化：将原始数据变换映射到**0-1**之间，**消除量纲**的影响



➡ 归一化前

➡ 归一化后

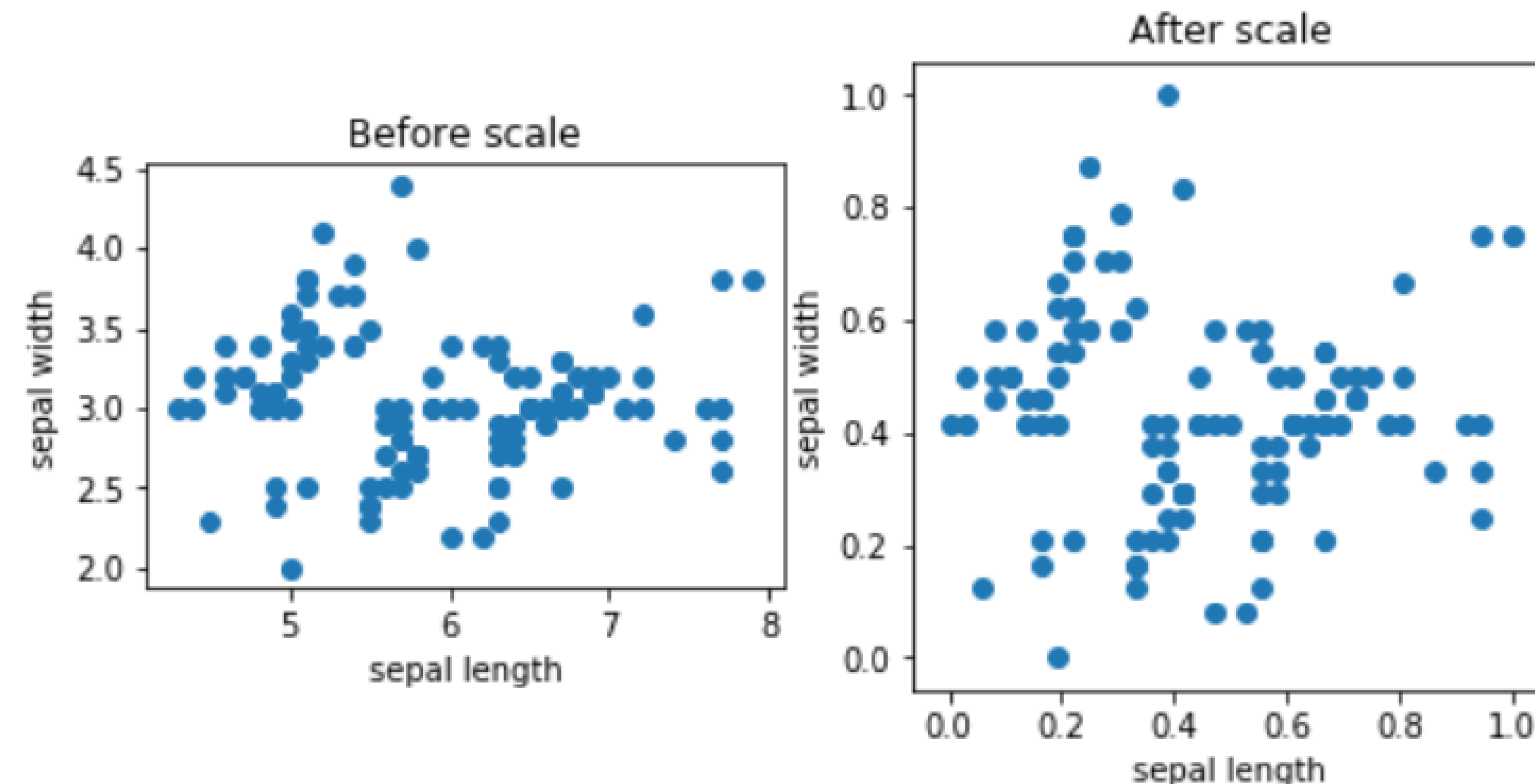
数据预处理及特征工程

- 如何归一化

- $$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$
- `sklearn.preprocessing.MinMaxScaler()`

- 注意

- 先对训练集进行归一化，然后对测试集进行归一化
- 在测试集上的scaler和训练集上的scaler要保持一致；不要在训练集和测试集分别使用不同的scaler



只为遇见明天更优秀的你！