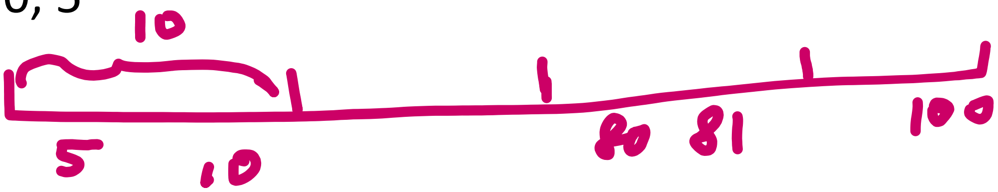
Association Rule: example: the layout of stuff in Starbucks

Supervised learning (labeled data):

- Classification (labels: a set of items)
- Regression (labels: real numbers)

Predict the humidity in Las Vegas (%): 80, 81, 100, 10, 5



Simple rules:

- < 20 low
- 20 <= < 70 medium
- >= 70 high

Unsupervised learning (no labels):

Dataset for supervised / unsupervised learning:

Training Dataset (historical data) => build up a model

Testing Dataset (given new data (never used in training)) => use the model to predict

Measures:

1. Accuracy
   100 data points: 98 of them labelled with yes, and 2 labelled with no. – imbalanced problem (common interview question)

   Accuracy 99%? (99 of them are labelled correctly, 1 was labeled wrong). If the wrong labeled one is the data with original label no, that is, the accuracy for "no" class is only 50%.

2. Precision and recall (common interview question)
   F1 score is the combination of precision and recall.

- Supervised Learning (i.i.d)
- Unsupervised learning
- Reinforcement learning (a sequence decision making process, which violates the i.i.d assumption of supervised learning)

100 datapoints: split them into 70 datapoints in training and 30 datapoints in testing.
100 datapoints: 80 are labeled as yes; 20 are labeled as no. When I split data into training / testing:

Train data:70 datapoints are labeled as yes;
Testing: 30 (20 no + 10 yes).

How to calculate support count, support. How to find frequent itemset.

If we set up the mini_support with a larger values, then we will have less frequent itemsets.

-
-
- Clustering