

Machine Learning 1

Yuwen Deng

1 Problem 1

1.What is the target variable of this problem? What are the potential features?

Target value:we want to predict with machine learning algorithm.

Potential feature:features which are invisible or overlooked in train set

2.If we do not have these data now, where can we find them or how can we collect them?

scrap a website and extract data; RSS feed or API; some device.

3. Does this problem solved comprehensively? How does people solve it usually?

No,some common algorithms are divided in two categories, supervised learning and unsupervised learning, the former contains k-Nearest Neighbors, Naive Bayes, Support vector machines, Decision trees, Linear, Locally weighted linear, Ridge and Lasso; The latter contains k-Means, DBSCAN, Expectation maximization and Parzen window.

2 Problem 2

```
1
2 from numpy import random as rp
3 import random as rd
4 n=200#sample number
5 p=70#feature
6 Data=rp.rand(200,70)
7
```

```
8
9  tvar=rp.rand(200)
10
11
12  li=[i for i in range(200)]
13  train=rd.sample(li,140)
14
15  train_set=[]
16  test_set=[]
17
18  for i in range(200):
19      if i in train:
20          train_set.append(Data[i])
21      else:
22          test_set.append(Data[i])
```