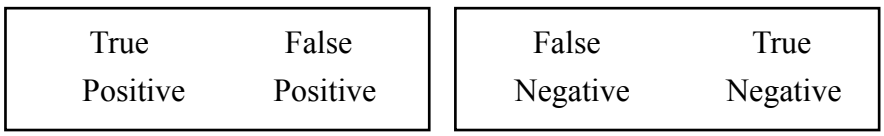
SC201 Lecture 8

Decision Tree

How to choose split?

1. 

2. Gini Impurity1 = _____

   Gini Impurity2 = _____
3. _____ (Ginni Impurity1, Gini Impurity2)
4. 愈 _____ 的 Impurity，愈 _____ 的 split

Random Forest

Covert for _____ habit of _____ to their Dtrain

```
from sklearn import ensemble

_____

forest.fit(x_train, y)

print('Acc:', forset_score(x_train, y))
```

Bootstrapping

_____ ➡ _____

- Create new dataset _____
- Get a sense of _____ if we redid the experiments

Bagging

_____ + _____

Bagging Classifier

Fit classifiers each on random subsets of the original dataset and them aggregate their individual predictions.

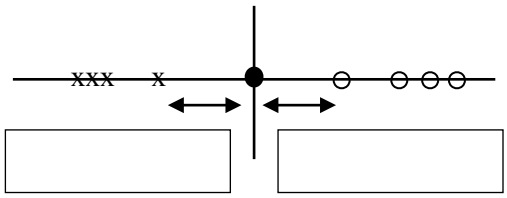
```
from sklearn import ensemble

_____
```

Ensemble Learning

Use _____ to obtain better prediction.

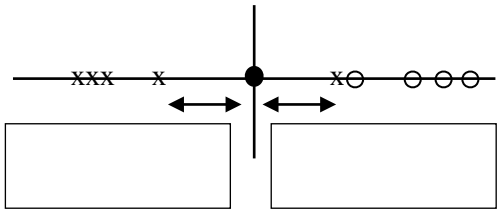
Super VectorMachines

How to choose the threshold splitting up x and o?



① _____ (邊界data 2點之 _____)

_____==_____ ➡ 不允許 _____



② _____ (SVC)

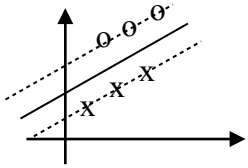
(soft margin _____)

When data is 1D → SVC is a _____

2D → SVC is a _____

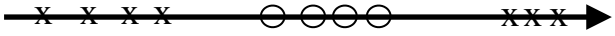
3D → SVC is a _____

4D → SVC is a _____



③ _____ (SVM)

_____



- Start with data in _____
- Move data into a _____
- Find a _____ to _____

How to transform data ?

- _____Kernel (_____)
- _____Kernel (_____)
- ⋮

_____

_____

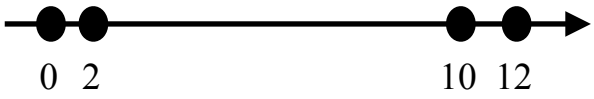
print('Acc:', svc.score(x_train, y))

k-means Clusteing

- Cluster data into k groups
- _____ (_____)
- Used in _____
- _____ is called _____, x is data

➡ Cost Function: _____

例1



① 2-means
② $\mu_0 = 2, \mu_1 = 12$

< algorithm >

① choose k (number of groups)

② randomly pick k centroids

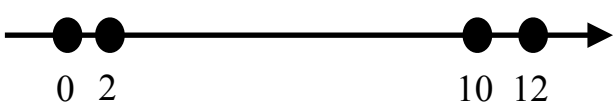
③ _____ = [-1]*len(data)
每筆資料給哪一個centroid

④ for i in range(len(data)):
assignments[i] = argmin(|data[i] - centroid[k]|)

⑥ iteration

⑤ re-assign centroids to the _____ of its group

例2



① 2-means
② $\mu_0 = 0, \mu_1 = 2$