

Label Propagation and Spreading

April 19, 2022

0.0.1 Semi Supervised Learning - label propagation and spreading

```
[1]: from sklearn.datasets import load_digits

[2]: digits=load_digits()

[3]: digits.data.shape

[3]: (1797, 64)

[4]: digits.target

[4]: array([0, 1, 2, ..., 8, 9, 8])

[5]: import matplotlib.pyplot as plt

[6]: import numpy as np

[7]: # plot a value

[8]: from sklearn.model_selection import train_test_split

[9]: #split the dataset using train_test_split function, pass train data, labels,
    ↪and test data ratio
x_train, x_test, y_train, y_test = train_test_split(digits.data,digits.
    ↪target,test_size=0.2)

[10]: x_train.shape

[10]: (1437, 64)

[11]: x_test.shape

[11]: (360, 64)

[12]: #target value
    print(y_train)
```

```
[1 8 8 ... 6 5 5]
```

```
[13]: #first get a list of indices of data  
indexList=np.arange(len(x_train))
```

```
[14]: print(indexList)
```

```
[ 0    1    2 ... 1434 1435 1436]
```

```
[15]: #split training set to labelled and non-labelled  
x_labelled=x_train[indexList[:300]]
```

```
[16]: #split training set to labelled and non-labelled  
y_labelled=y_train[indexList[:300]]
```

```
[17]: nonLabelledIndices=indexList[300:]
```

```
[18]: #remove the labels for non labelled indices in training data  
y_train_nonlabel = np.copy(y_train)
```

```
[19]: #setting -1 as non labelled  
y_train_nonlabel[nonLabelledIndices] = -1
```

```
[20]: from sklearn.semi_supervised import LabelPropagation
```

```
[21]: lp=LabelPropagation(gamma=.30)  
#gamma is measure of kernel func,
```

```
[22]: lp.fit(x_train, y_train_nonlabel)
```

```
[22]: LabelPropagation(gamma=0.3)
```

```
[23]: lp.score(x_test,y_test)
```

```
[23]: 0.9638888888888889
```

```
[24]: #capture labels set by Label Propagataion on the set of non labelled data  
y_labelled=lp.transduction_[nonLabelledIndices]
```

```
[25]: from sklearn.metrics import confusion_matrix
```

```
[26]: conf = confusion_matrix(y_train[nonLabelledIndices], y_labelled, labels=lp.  
    ↪classes_)
```

```
[27]: conf
```

```
[27]: array([[109,  0,  0,  0,  0,  0,  0,  0,  0,  0],
           [ 0, 115,  2,  1,  0,  0,  0,  0,  0,  0],
           [ 0,  0, 102,  0,  0,  0,  0,  1,  0,  0],
           [ 0,  0,  1, 95,  0,  2,  0,  2,  4,  0],
           [ 0,  4,  0,  0, 121,  0,  0,  2,  0,  0],
           [ 0,  0,  0,  0,  1, 115,  1,  0,  0,  1],
           [ 0,  3,  0,  0,  0,  0, 122,  0,  0,  0],
           [ 0,  0,  0,  0,  0,  0,  0, 110,  0,  2],
           [ 0,  9,  0,  2,  0,  0,  0,  0, 97,  0],
           [ 0,  2,  0,  2,  1,  3,  0,  1,  3, 101]])
```

```
[28]: from sklearn.semi_supervised import LabelSpreading
```

```
[29]: ls=LabelSpreading(gamma=0.3)
      #gamma is measure of kernel func,
```

```
[30]: ls.fit(x_train, y_train_nonlabel)
```

```
[30]: LabelSpreading(gamma=0.3)
```

```
[31]: ls.score(x_test, y_test)
```

```
[31]: 0.9861111111111112
```

```
[32]: #capture labels set by Label Propagataion on the set of non labelled data
      y_labelled=lp.transduction_[nonLabelledIndices]
```

```
[33]: conf = confusion_matrix(y_train[nonLabelledIndices], y_labelled, labels=lp.
      ↪classes_)
```

```
[34]: conf
```

```
[34]: array([[109,  0,  0,  0,  0,  0,  0,  0,  0,  0],
           [ 0, 115,  2,  1,  0,  0,  0,  0,  0,  0],
           [ 0,  0, 102,  0,  0,  0,  0,  1,  0,  0],
           [ 0,  0,  1, 95,  0,  2,  0,  2,  4,  0],
           [ 0,  4,  0,  0, 121,  0,  0,  2,  0,  0],
           [ 0,  0,  0,  0,  1, 115,  1,  0,  0,  1],
           [ 0,  3,  0,  0,  0,  0, 122,  0,  0,  0],
           [ 0,  0,  0,  0,  0,  0,  0, 110,  0,  2],
           [ 0,  9,  0,  2,  0,  0,  0,  0, 97,  0],
           [ 0,  2,  0,  2,  1,  3,  0,  1,  3, 101]])
```

```
[ ]:
```