Dimensionality_reduction_visualisation

January 21, 2020

1 Dimensionality reduciton on the iris data set

[1]:
```
from sklearn import datasets
from sklearn.utils import shuffle
import pandas as pd
import numpy as np

#Load data in matrix X for easier reference
iris = datasets.load_iris()
X = iris.data

#Normalize Data to be in range [0,1]
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
scaler.fit(X)
X = scaler.transform(X)

#Create Pandas Dataframe for table display
df = pd.DataFrame(X[:,0:3])
df.columns = iris.feature_names[0:3]
df['class'] = iris.target

#label class names
df['class_name'] = [iris.target_names[i] for i in df['class']]

#shuffle data
df = shuffle(df)
X = np.array(df.loc[:,'sepal length (cm)':'petal length (cm)'])


#Visualize data in table
df
```

[1]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	class	class_name
108	0.666667	0.208333	0.813559	2	virginica
25	0.194444	0.416667	0.101695	0	setosa
147	0.611111	0.416667	0.711864	2	virginica

16	0.305556	0.791667	0.050847	0	setosa
69	0.361111	0.208333	0.491525	1	versicolor
77	0.666667	0.416667	0.677966	1	versicolor
42	0.027778	0.500000	0.050847	0	setosa
14	0.416667	0.833333	0.033898	0	setosa
12	0.138889	0.416667	0.067797	0	setosa
148	0.527778	0.583333	0.745763	2	virginica
115	0.583333	0.500000	0.728814	2	virginica
59	0.250000	0.291667	0.491525	1	versicolor
83	0.472222	0.291667	0.694915	1	versicolor
34	0.166667	0.458333	0.084746	0	setosa
41	0.055556	0.125000	0.050847	0	setosa
120	0.722222	0.500000	0.796610	2	virginica
17	0.222222	0.625000	0.067797	0	setosa
79	0.388889	0.250000	0.423729	1	versicolor
125	0.805556	0.500000	0.847458	2	virginica
131	1.000000	0.750000	0.915254	2	virginica
4	0.194444	0.666667	0.067797	0	setosa
63	0.500000	0.375000	0.627119	1	versicolor
1	0.166667	0.416667	0.067797	0	setosa
55	0.388889	0.333333	0.593220	1	versicolor
140	0.666667	0.458333	0.779661	2	virginica
126	0.527778	0.333333	0.644068	2	virginica
92	0.416667	0.250000	0.508475	1	versicolor
37	0.166667	0.666667	0.067797	0	setosa
75	0.638889	0.416667	0.576271	1	versicolor
110	0.611111	0.500000	0.694915	2	virginica
..	...	...	...	...	...
88	0.361111	0.416667	0.525424	1	versicolor
67	0.416667	0.291667	0.525424	1	versicolor
124	0.666667	0.541667	0.796610	2	virginica
38	0.027778	0.416667	0.050847	0	setosa
135	0.944444	0.416667	0.864407	2	virginica
43	0.194444	0.625000	0.101695	0	setosa
31	0.305556	0.583333	0.084746	0	setosa
10	0.305556	0.708333	0.084746	0	setosa
89	0.333333	0.208333	0.508475	1	versicolor
61	0.444444	0.416667	0.542373	1	versicolor
33	0.333333	0.916667	0.067797	0	setosa
80	0.333333	0.166667	0.474576	1	versicolor
137	0.583333	0.458333	0.762712	2	virginica
78	0.472222	0.375000	0.593220	1	versicolor
85	0.472222	0.583333	0.593220	1	versicolor
45	0.138889	0.416667	0.067797	0	setosa
73	0.500000	0.333333	0.627119	1	versicolor
76	0.694444	0.333333	0.644068	1	versicolor
111	0.583333	0.291667	0.728814	2	virginica

```
93          0.194444          0.125000          0.389831      1  versicolor
23          0.222222          0.541667          0.118644      0      setosa
142         0.416667          0.291667          0.694915      2   virginica
8           0.027778          0.375000          0.067797      0      setosa
57          0.166667          0.166667          0.389831      1  versicolor
109         0.805556          0.666667          0.864407      2   virginica
18          0.388889          0.750000          0.118644      0      setosa
82          0.416667          0.291667          0.491525      1  versicolor
48          0.277778          0.708333          0.084746      0      setosa
72          0.555556          0.208333          0.661017      1  versicolor
122         0.944444          0.333333          0.966102      2   virginica

[150 rows x 5 columns]
```

[3]:
```
from mpl_toolkits import mplot3d
import matplotlib.pyplot as plt
%matplotlib notebook

#Iniatiate figure
plt.figure(figsize = (10,10))
ax = plt.axes(projection = '3d')

#Filter dataframe for different classes
versicolor = df[df['class_name'].str.match('versicolor')]
virginica = df[df['class_name'].str.match('virginica')]
setosa = df[df['class_name'].str.match('setosa')]

#Plot different classes
ax.scatter3D(versicolor['sepal length (cm)'],versicolor['sepal width␣
 ↪(cm)'],versicolor['petal length (cm)'],color = 'red', marker = '^')
ax.scatter3D(virginica['sepal length (cm)'],virginica['sepal width␣
 ↪(cm)'],virginica['petal length (cm)'],color = 'blue', marker = 'o')
ax.scatter3D(setosa['sepal length (cm)'],setosa['sepal width␣
 ↪(cm)'],setosa['petal length (cm)'],color = 'green', marker = '*')
ax.set_title('Three dimensional plot of reduced data set')
ax.set_xlabel('sepal length [cm]')
ax.set_ylabel('sepal width [cm]')
ax.set_zlabel('petal length [cm]')
plt.show()
```

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.HTML object>
```

[ ]:
```
```