comparison

May 1, 2020

```python
In [42]: import pandas as pd
         import matplotlib.pyplot as plt
         %matplotlib inline
```

```python
In [43]: # Square Euclidean and Manhattan Accuracy computed in terms of # neighbors
         df = pd.read_csv("comparison.txt", header=None, sep=" ")
         df.columns = ['neighbors','euclid','manhattan']
         df.head()
```

```
Out[43]:    neighbors    euclid   manhattan
         0           1    0.0000      0.0000
         1           2    0.0000      0.0000
         2           3   66.6667     83.3333
         3           4   77.7778     88.8889
         4           5   77.7778     88.8889
```

```python
In [44]: df.describe()
```

```
Out[44]:             neighbors       euclid     manhattan
         count    213.000000   213.000000    213.000000
         mean     107.000000    34.898290     37.245708
         std       61.631972    20.380634     22.728650
         min        1.000000     0.000000      0.000000
         25%       54.000000    16.666700     16.666700
         50%      107.000000    27.777800     27.777800
         75%      160.000000    44.444400     50.000000
         max      213.000000    77.777800     88.888900
```

```python
In [45]: # Plot accuracy with respect to number of neighbors
         plt.plot(df['neighbors'],df['euclid'], label='Square Euclidean')
         plt.plot(df['neighbors'],df['manhattan'], label='Manhattan')
         plt.legend()
         plt.title('% Accuracy vs Neighbors computed')
```

```
Out[45]: <matplotlib.text.Text at 0x1202bc3c8>
```

% Accuracy vs Neighbors computed
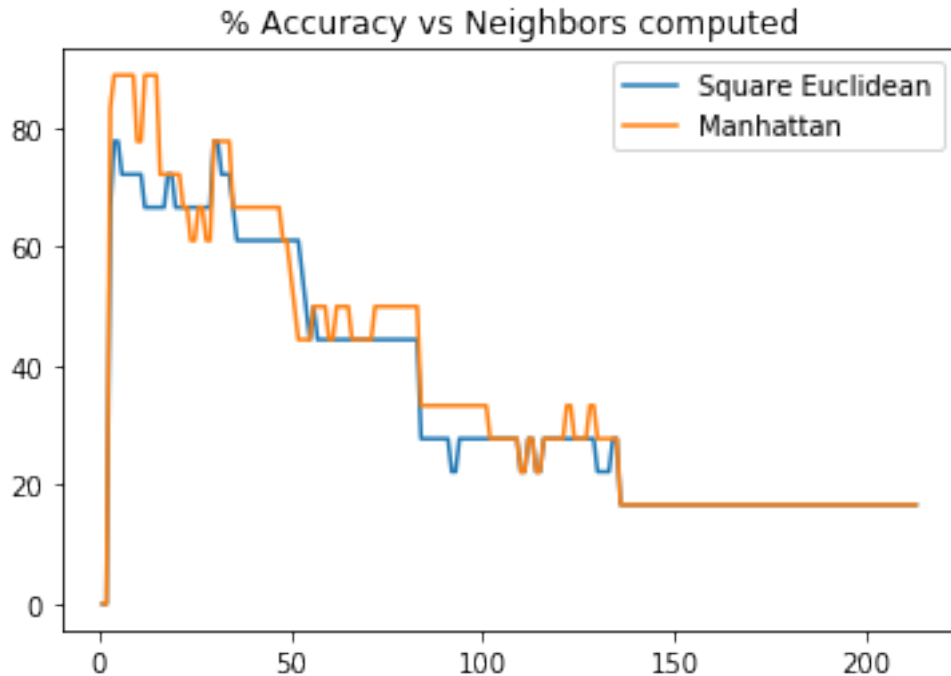
In [46]: `# Find where square euclid accuracy peaked`
```
temp = df[df['euclid'] == max(df['euclid'])]
temp
```

Out[46]:
```
     neighbors   euclid   manhattan
3            4  77.7778     88.8889
4            5  77.7778     88.8889
29          30  77.7778     77.7778
30          31  77.7778     77.7778
```

In [47]: `# Plot neighborhood of peak euclid accuracy`
```
temp = df[df['neighbors'] <= 31]
temp = temp[temp['neighbors'] >= 4]
plt.plot(temp['neighbors'],temp['euclid'], label='Square Euclidean')
plt.plot(temp['neighbors'],temp['manhattan'], label='Manhattan')
plt.legend()
plt.title('Peak Square Euclidean % Accuracy Neighborhood')
```

Out[47]: <matplotlib.text.Text at 0x1203c8240>

Peak Square Euclidean % Accuracy Neighborhood
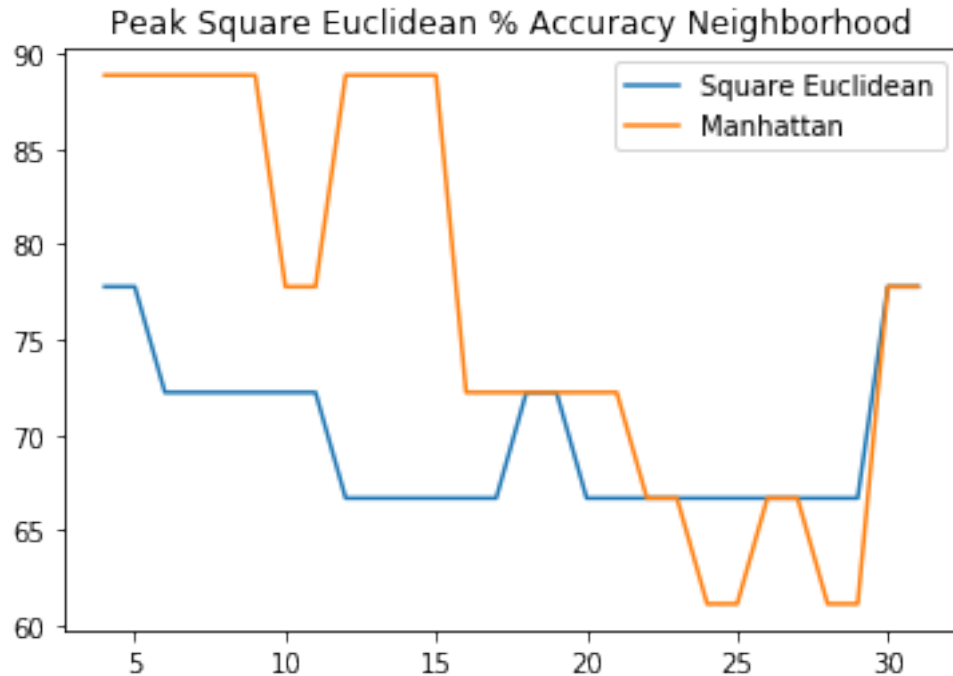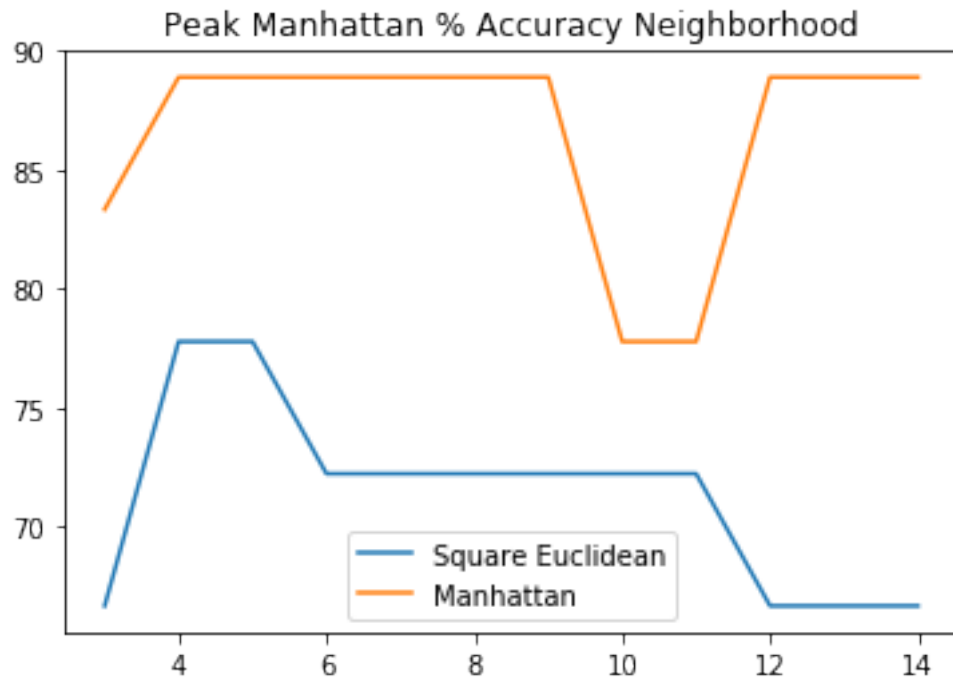


```
In [48]: # Find where
         temp = df[df['manhattan'] == max(df['manhattan'])]
         temp
```

```
Out[48]:      neighbors    euclid   manhattan
         3            4   77.7778    88.8889
         4            5   77.7778    88.8889
         5            6   72.2222    88.8889
         6            7   72.2222    88.8889
         7            8   72.2222    88.8889
         8            9   72.2222    88.8889
         11          12   66.6667    88.8889
         12          13   66.6667    88.8889
         13          14   66.6667    88.8889
         14          15   66.6667    88.8889
```

```
In [49]: temp = df[df['neighbors'] <= 14]
         temp = temp[temp['neighbors'] >= 3]
         plt.plot(temp['neighbors'],temp['euclid'], label='Square Euclidean')
         plt.plot(temp['neighbors'],temp['manhattan'],label='Manhattan')
         plt.legend()
         plt.title('Peak Manhattan % Accuracy Neighborhood')
```

```
Out[49]: <matplotlib.text.Text at 0x12056fe80>
```

Peak Manhattan % Accuracy Neighborhood

In [50]: *# Find where Square Euclidean and Manhattan both peaked*
         df[(df['euclid'] == max(df['euclid'])) & (df['manhattan'] == max(df['manhattan']))]

Out[50]:   neighbors   euclid   manhattan
         3   4 77.7778   88.8889
         4   5 77.7778   88.8889