train

October 20, 2019

```python
[100]: import pandas as pd
       import math
       import csv
```

```python
[101]: path = './data_tp/'
       AODc = pd.read_csv(path+'AOD.csv')
       PMc = pd.read_csv(path+'PM2.5_all.csv')
```

```python
[102]: AODd = { AODc["date"][i] : AODc["data"][i][1:-1].split(', ') for i in␣
       ↪range(AODc.shape[0]) }
       for key in AODd:
           AODd[key] = [float( AODd[key][x]) for x in range(len(AODd[key]))]
       PMd = {PMc["date"][i] : float(PMc["PM2.5"][i]) for i in range(PMc.shape[0])}

       #AODd = {date : [features]}
       #PMd = {date : PM2.5@12pm}
```

```python
[103]: X = []
       y = []

       for k in AODd:
           if k in PMd:
               X.append(AODd[k])
               y.append(PMd[k])
```

```python
[104]: len(X)
```

```
[104]: 576
```

```python
[105]: import numpy as np
       import matplotlib.pyplot as plt
       from sklearn.linear_model import LinearRegression
       from sklearn.model_selection import train_test_split


       X = np.asarray(X, dtype=np.float32)
```

```
#print(X.shape)
#X.reshape(-1, X.shape[0])
print(X.shape)
y = np.asarray(y, dtype=np.float32)
print(y.shape)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 →random_state=0)


model = LinearRegression()
model.fit(X, y)
```

```
(576, 38)
(576,)
```

[105]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

[106]:
```
r_sq = model.score(X, y)
print(r_sq)
```

```
0.5395696836596655
```

[108]:
```
model.coef_
```

[108]:
```
array([ 3.36093354e+00, -8.70689106e+00, -4.98759699e+00, -1.39064728e+02,
        4.81518677e+02,  1.58627949e+01, -6.09708008e+02, -2.64280304e+02,
       -1.54929337e+02, -8.46729431e+01, -7.60400295e-01, -1.45099243e+02,
       -1.22962456e+02,  3.01127411e+02,  1.87292328e+01, -1.41671257e+01,
        9.43958759e-04, -7.93389320e+00,  0.00000000e+00,  0.00000000e+00,
       -2.53420520e+00,  7.37782431e+00, -1.35253060e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  9.76683807e+00, -4.59541750e+00,
       -5.55706215e+00,  1.34128046e+00, -9.96624827e-02,  9.51182266e+04,
        6.24165535e-02,  5.17616272e-02], dtype=float32)
```

[ ]:

[ ]:

[ ]:

[109]:
```
#0.54 104-108
#0.45 105-106
```

[ ]:
```
```