一、前期准备工作

1.导入数据

In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
"""
Cement:                          水泥   （单位：千克）
Blast Furnace Slag:              矿渣   （单位：千克）
Fly Ash:                         煤灰   （单位：千克）
Water:                           水     （单位：千克）
Superplasticizer:                塑化剂（单位：千克）
Coarse Aggregate:                粗颗粒（单位：千克）
Fine Aggregate:                  细颗料（单位：千克）
Age:                             天数   （已经使用的天数）
Concrete compressive strength: 1立方米混凝土抗压强度
"""
dataframe = pd.read_excel('../data/Concrete_Data.xls')
dataframe.head(5)
```

Out[1]:

	Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Co comp st
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.986
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.887
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.269
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.052
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.296

2.整理数据

In [2]:

```
# """
# 将所有数据中水泥的质量调整为100千克
# """
# dataframe["Blast Furnace Slag"]=dataframe["Blast Furnace Slag"]/dataframe["Cement"]*100
# dataframe["Fly Ash"]=dataframe["Fly Ash"]/dataframe["Cement"]*100
# dataframe["Water"]=dataframe["Water"]/dataframe["Cement"]*100
# dataframe["Superplasticizer"]=dataframe["Superplasticizer"]/dataframe["Cement"]*100
# dataframe["Coarse Aggregate"]=dataframe["Coarse Aggregate"]/dataframe["Cement"]*100
# dataframe["Fine Aggregate"]=dataframe["Fine Aggregate"]/dataframe["Cement"]*100
# dataframe["Cement"]=dataframe["Cement"].map(lambda x: x/x*100)
# dataframe.head()
```

3.探究水泥质量与各个成分含量之间的关系

In [4]:

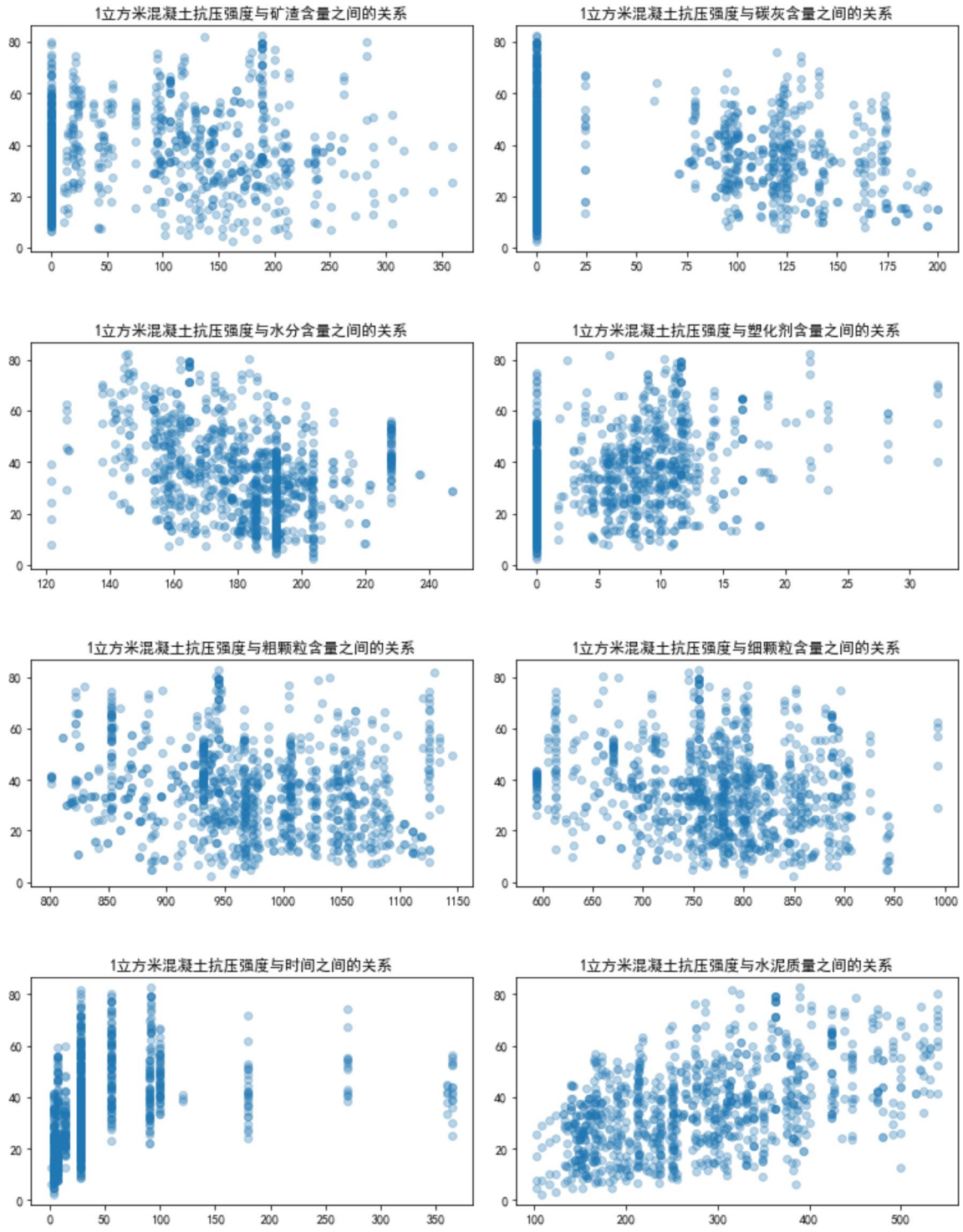
```
import numpy as np

#设置中文显示
from pylab import mpl
mpl.rcParams['font.sans-serif'] = ['SimHei']

fig = plt.figure(figsize = (13,17))
#调整子图间距
plt.subplots_adjust(wspace =0.1, hspace =0.4)
ax1 = fig.add_subplot(421)
ax2 = fig.add_subplot(422)
ax3 = fig.add_subplot(423)
ax4 = fig.add_subplot(424)
ax5 = fig.add_subplot(425)
ax6 = fig.add_subplot(426)
ax7 = fig.add_subplot(427)
ax8 = fig.add_subplot(428)
ax1.scatter(dataframe['Blast Furnace Slag'], dataframe['Concrete compressive strength'], alpha=
0.3)
ax1.set_title('1立方米混凝土抗压强度与矿渣含量之间的关系')
ax2.scatter(dataframe['Fly Ash'], dataframe['Concrete compressive strength'], alpha=0.3)
ax2.set_title('1立方米混凝土抗压强度与碳灰含量之间的关系')
ax3.scatter(dataframe['Water'], dataframe['Concrete compressive strength'], alpha=0.3)
ax3.set_title('1立方米混凝土抗压强度与水分含量之间的关系')
ax4.scatter(dataframe['Superplasticizer'], dataframe['Concrete compressive strength'], alpha=0.3
)
ax4.set_title('1立方米混凝土抗压强度与塑化剂含量之间的关系')
ax5.scatter(dataframe["Coarse Aggregate"], dataframe['Concrete compressive strength'], alpha=0.3
)
ax5.set_title('1立方米混凝土抗压强度与粗颗粒含量之间的关系')
ax6.scatter(dataframe["Fine Aggregate"], dataframe['Concrete compressive strength'], alpha=0.3)
ax6.set_title('1立方米混凝土抗压强度与细颗粒含量之间的关系')
ax7.scatter(dataframe["Age"], dataframe['Concrete compressive strength'], alpha=0.3)
ax7.set_title('1立方米混凝土抗压强度与时间之间的关系')
ax8.scatter(dataframe["Cement"], dataframe['Concrete compressive strength'], alpha=0.3)
ax8.set_title('1立方米混凝土抗压强度与水泥质量之间的关系')
```

Out[4]:

Text(0.5, 1.0, '1立方米混凝土抗压强度与水泥质量之间的关系')



二、多元线性回归

1.训练模型

In [5]:

```
#Cement Blast Furnace Slag      Fly Ash Water   Superplasticizer        Coarse Aggregate
Fine Aggregate  Age     Concrete compressive strength
from sklearn.linear_model import LinearRegression

#初始化模型
mul_LR_model = LinearRegression()


#拟合模型
mul_LR_model.fit(dataframe[['Cement','Blast Furnace Slag','Fly Ash','Water','Superplasticizer',
'Coarse Aggregate','Fine Aggregate','Age']], dataframe['Concrete compressive strength'])
#预测
dataframe['预测值'] = mul_LR_model.predict(dataframe[['Cement','Blast Furnace Slag','Fly Ash','W
ater','Superplasticizer','Coarse Aggregate','Fine Aggregate','Age']])
#显示
dataframe.head(5)
```

Out[5]:

	Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Co comp s
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.986
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.887
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.269
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.052
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.296

2.计算得分

In [6]:

```
mul_score = mul_LR_model.score(dataframe[['Cement','Blast Furnace Slag','Fly Ash','Water','Super
plasticizer','Coarse Aggregate','Fine Aggregate','Age']], dataframe['Concrete compressive streng
th'])
mul_score
```

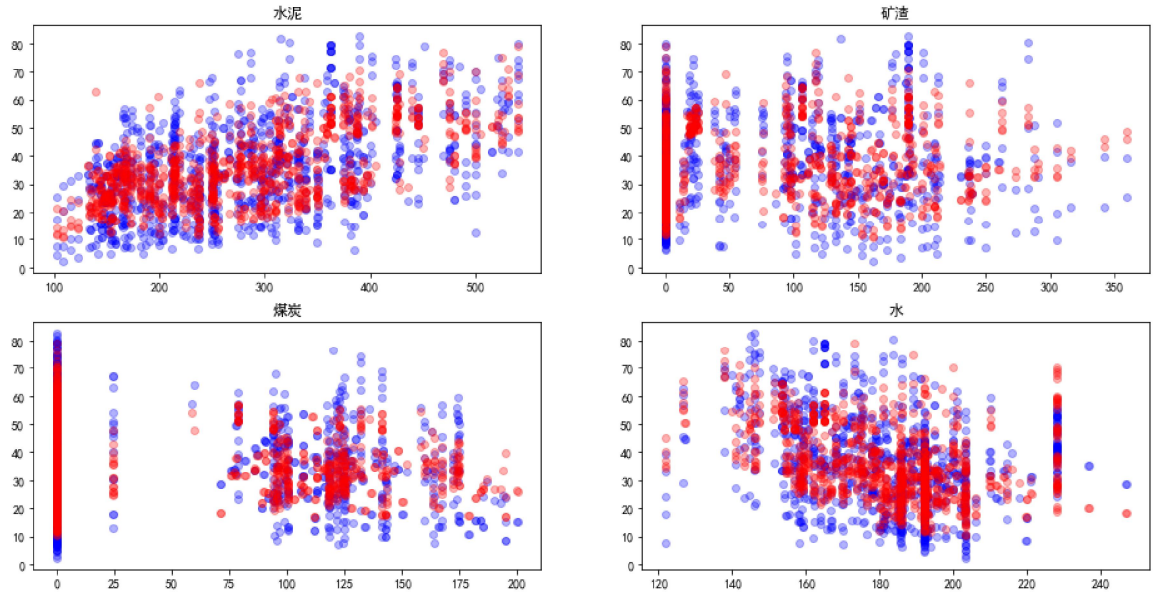
Out[6]:

0.6154647342687214

3.可视化预测结果

In [7]:

```
fig = plt.figure(figsize = (16,8))
ax1 = fig.add_subplot(2,2,1)
ax2 = fig.add_subplot(2,2,2)
ax3 = fig.add_subplot(2,2,3)
ax4 = fig.add_subplot(2,2,4)
ax1.scatter(dataframe['Cement'], dataframe['Concrete compressive strength'], c='blue', alpha=0.3
)
ax1.scatter(dataframe['Cement'], dataframe['预测值'], c='red', alpha=0.3)
ax1.set_title('水泥')
ax2.scatter(dataframe["Blast Furnace Slag"], dataframe['Concrete compressive strength'], c='blu
e', alpha=0.3)
ax2.scatter(dataframe["Blast Furnace Slag"], dataframe['预测值'], c='red', alpha=0.3)
ax2.set_title('矿渣')
ax3.scatter(dataframe['Fly Ash'], dataframe['Concrete compressive strength'], c='blue', alpha=0.
3)
ax3.scatter(dataframe['Fly Ash'], dataframe['预测值'], c='red', alpha=0.3)
ax3.set_title('煤炭')
ax4.scatter(dataframe['Water'], dataframe['Concrete compressive strength'], c='blue', alpha=0.3)
ax4.scatter(dataframe['Water'], dataframe['预测值'], c='red', alpha=0.3)
ax4.set_title('水')
plt.show()
```



In [ ]: