class 5

房价数据

```
import pandas as pd
data = pd.read_csv('load_kc_house_data.csv')
```

bathroom 转整数

```
def trans_int(x):
    if x-int(x) <= 0.5:
        return int(x)
    else:
        return int(x)+1

data_new['bathrooms'].apply(trans_int)

data_new1 = data_new.copy()
data_new1['bathrooms'] = data_new['bathrooms'].apply(trans_int)

data_new1.groupby('bathrooms').mean()['price'].plot.bar()
```

线性回归

```
from sklearn.linear_model import LinearRegression

X = data_new[['bathrooms','sqft_living']]
y = data_new['price']

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

测试模型准确度

```
from sklearn.metrics import r2_score, mean_absolute_error

price_pred = model.predict(X)
r2_score(data_new['price'],price_pred)
```

决策树

```
# 决策树
from sklearn.tree import DecisionTreeRegressor
```

```
X = data_new[['bathrooms','sqft_living']]
y = data_new['price']

model = DecisionTreeRegressor()
model.fit(X,y)
from sklearn.metrics import r2_score, mean_absolute_error

price_pred = model.predict(X)
r2_score(data_new['price'],price_pred)
```

作业

- 使用至少3个特征对房价进行预测，尝试找到一种模型可以将房价预测的准确度提升70%以上
- 将上述分析和建模过程按照七步建模法，整理在word中

将代码和word文档在下节课上课前一天发送到 learningmm@163.com 主题为“宋校+姓名+第5次课作业”