Thực hành Nguyên Lý máy học
Bổ sung Buổi 2: CÂY HỒI QUY

1. HƯỚNG DẪN THỰC HÀNH

1. **Cho tập dữ liệu housing_RT.csv có dạng:**

	price	lotsize	bedrooms	bathrms	stories
1	38500.0	4000	2	1	1
2	49500.0	3060	3	1	1
3	60500.0	6650	3	1	2
4	61000.0	6360	2	1	1

2. **Đọc dữ liệu vào biến "dulieu"**

```
import pandas as pd
dulieu = pd.read_csv("housing_RT.csv", index_col=0)
dulieu.iloc[1:5,]
```

3. Sử **dụng nghi thức hold-out Phân chia tập dữ liệu huấn luyện**

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split( dulieu.ix[:,1:5],dulieu.ix[:,0],
test_size=1/3.0, random_state=100)
X_train[1:5,]
X_test[1:5]
```

4. **Xây dựng mô hình**

```
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(X_train, y_train)
```

5. **Dự báo và đánh giá mô hình**

```
y_pred = regressor.predict(X_test)
y_test[1:5]
y_pred[1:5]
```

Đánh giá thông sai số giữa dự báo và thực tế

- Chỉ số MSE

```
from sklearn.metrics import mean_squared_error
err = mean_squared_error(y_test, Y_pred)
err
```

- Chỉ số RMSE

```
from math import sqrt
sqrt(err)
```