Regression

Data import

```
1 import pandas as pd
2 import os
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import torch
6 import torch.optim as optim
7 from torch import nn as nn
8 from torch.nn import functional as F
```

```
1 dataset_url = 'https://bitbucket.org/hyuk125/lg_dic/raw/889649d1bc273bf53967cb
2 data = pd.read_csv(dataset_url)
```

```
1 data.head()
```

데이터 label encoding

```
1 from sklearn.preprocessing import LabelEncoder
2 le = LabelEncoder()
3 #sex
4 le.fit(data['sex'])
5 data['sex'] = le.transform(data['sex'])
6 # smoker or not
7 le.fit(data['smoker'])
8 data['smoker'] = le.transform(data['smoker'])
9 #region
10 le.fit(data['region'])
11 data['region'] = le.transform(data['region'])
```

```
1 data.head(3)
```

Data setup

```
1 smoke_data = data[data['smoker'] == 1]
```

```
1 X = torch.from_numpy(smoke_data['bmi'].values)
2 y = torch.from_numpy(smoke_data['charges'].values)
```

```
X = X.reshape(-1, 1).float()
y = y.reshape(-1, 1).float()
```

```
plt.scatter(X, y, s=10)
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

과제1. 위 데이터를 nn.Module을 활용하여 linear regression하는 코드를 작성하시오

단, epoch=10000, learning rate=0.0001로 하시오

```
# 모델
```

```
hx = (model(X_train)).detach().numpy()
```

```
plt.figure(figsize=[6, 6])
plt.scatter(X_train, y_train, s=10)
plt.scatter(X_train, hx, s=20, c='r')
```

```
1
```