
Machine Learning Assignment

비만을 예측

- 과목: 기초웹개발론
- 전공: 컴퓨터공학전공
- 학번: 2017108242
- 이름: 강대렬

Table of contents

+

○

●

1. 개요
2. 관련 연구
3. 내용 요약
4. 지도 학습 과정
5. 결론

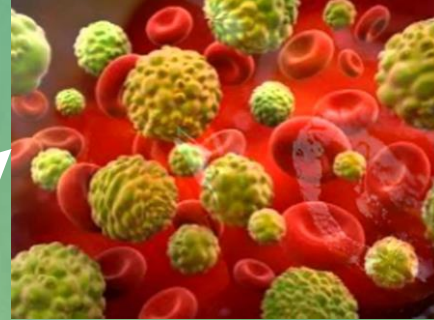
개요



비만 – obesity

비만은 지방이 정상보다 더 많이 축적된 상태를 말합니다.

개요



혈관 질환

+



당뇨



심장 질환

개요



자신의 몸 상태를 아는 것 = 중요한 것

관련 연구



Bioimpedence - 전류

Skin fold - 피부 두께

수중 체중 측정

DEXA - X선

Bod pot - 공기변위 체질량

+

o

•

관련 연구



비만(체지방)을 <-> 신체 조건

내용 요약



신체조건과 비만율은 상관 관계를 가집니다.

지도 학습 과정

+

○

●

1. 데이터셋 로드

2. 데이터 분석

3. 모델 훈련

4. 예측 결과

지도 학습 - Dataset load

```
Body_fat = pd.read_csv(file_path)
```

```
Body_fat.head()
```

	Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
0	1.0708	12.3	23	154.25	67.75	36.2	93.1	85.2	94.5	59.0	37.3	21.9	32.0	27.4	17.1
1	1.0853	6.1	22	173.25	72.25	38.5	93.6	83.0	98.7	58.7	37.3	23.4	30.5	28.9	18.2
2	1.0414	25.3	22	154.00	66.25	34.0	95.8	87.9	99.2	59.6	38.9	24.0	28.8	25.2	16.6
3	1.0751	10.4	26	184.75	72.25	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4	18.2
4	1.0340	28.7	24	184.25	71.25	34.4	97.3	100.0	101.9	63.2	42.2	24.0	32.2	27.7	17.7

지도 학습 - Dataset load

```
Body_fat = pd.read_csv(file_path)
```

```
Body_fat.info()
```

Data columns (total 15 columns):

#	Column	Non-Null	Count	Dtype
0	Density	252	non-null	float64
1	BodyFat	252	non-null	float64
2	Age	252	non-null	int64
3	Weight	252	non-null	float64
4	Height	252	non-null	float64
5	Neck	252	non-null	float64
6	Chest	252	non-null	float64
7	Abdomen	252	non-null	float64
8	Hip	252	non-null	float64
9	Thigh	252	non-null	float64
10	Knee	252	non-null	float64
11	Ankle	252	non-null	float64
12	Biceps	252	non-null	float64
13	Forearm	252	non-null	float64
14	Wrist	252	non-null	float64

Density - 수중 체밀도

Body fat - 체지방률

Age - 나이

Weight - 몸무게(lbs)

Height - 키(inches)

Neck - 목 둘레

Chest - 가슴 둘레

Abdomen - 복부 둘레

Hip - 엉덩이 둘레

Thigh - 허벅지 둘레

Knee - 무릎 둘레

Ankle - 발목 둘레

Biceps - 이두근 둘레

Forearm - 전완근 둘레

Wrist - 손목 둘레

지도 학습 – Dataset Analysis

```
plt.figure(figsize=(15,10))
sns.heatmap(Body_fat.corr(),annot=True)
plt.title('Heatmap of Variable Correlations')
plt.show()
```



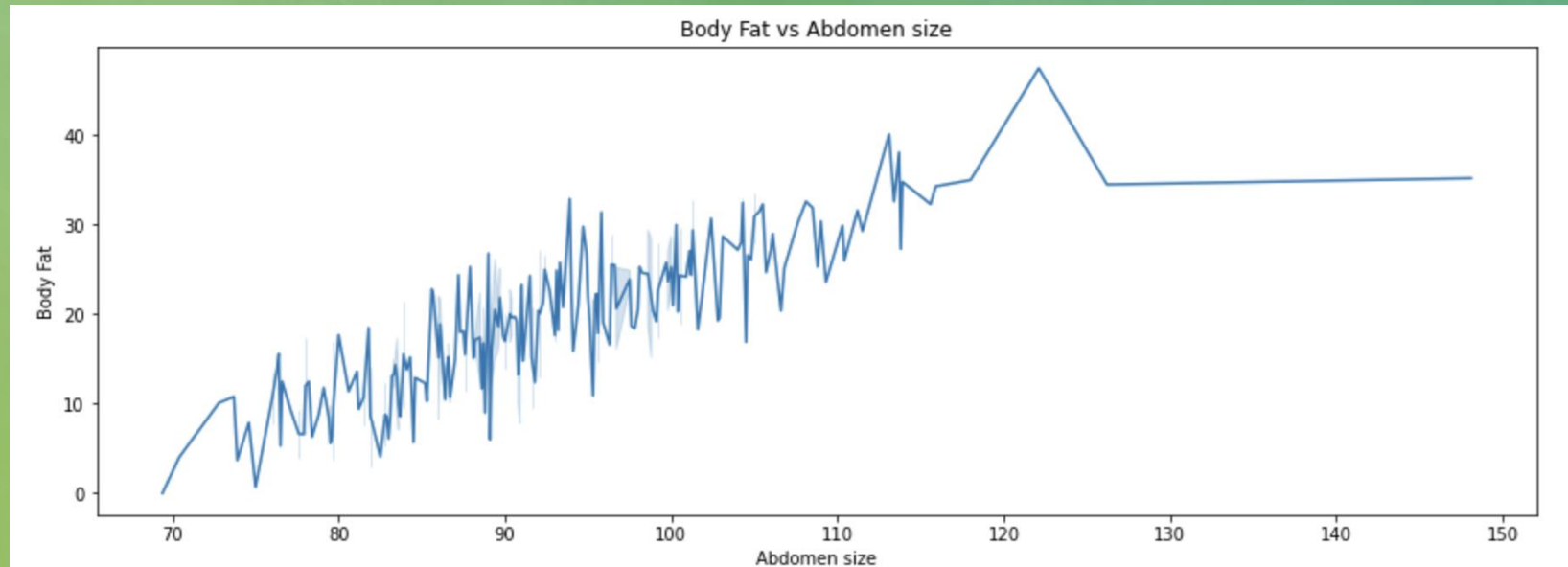
- Positive Correlations
Abodiem(복부 둘레)
/ Chest (가슴 둘레)
/ Hip과(힙 둘레)와 상관관계

- Negative Correlations
density 와 부정적 상관 관계

지도 학습 – Dataset Analysis

```
plt.figure(figsize=(15,5))
sns.lineplot(x='Abdomen',y='BodyFat',data=Body_fat)
plt.title('Body Fat vs Abdomen size')
plt.xlabel('Abdomen size')
plt.ylabel('Body Fat')
plt.show()
```

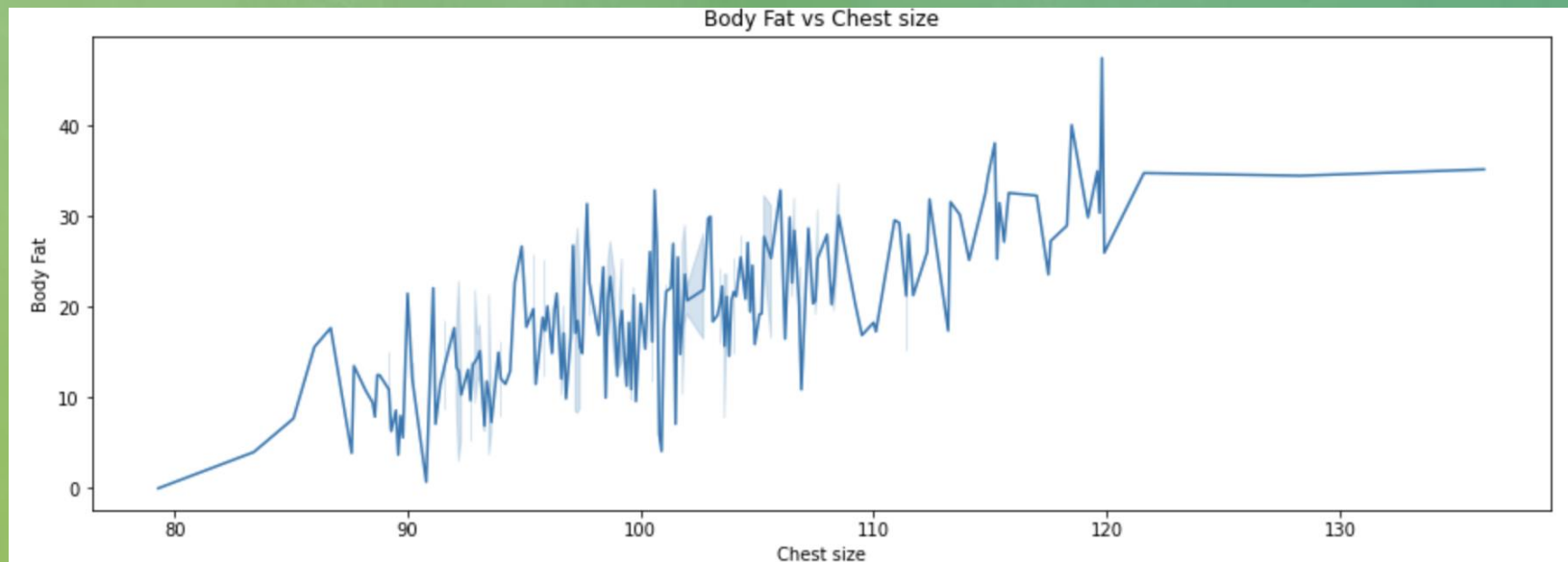
Body fat – Abodiem(복부)



지도 학습 – Dataset Analysis

```
plt.figure(figsize=(15,5))
sns.lineplot(x='Chest',y='BodyFat',data=Body_fat)
plt.title('Body Fat vs Chest size')
plt.xlabel('Chest size')
plt.ylabel('Body Fat')
plt.show()
```

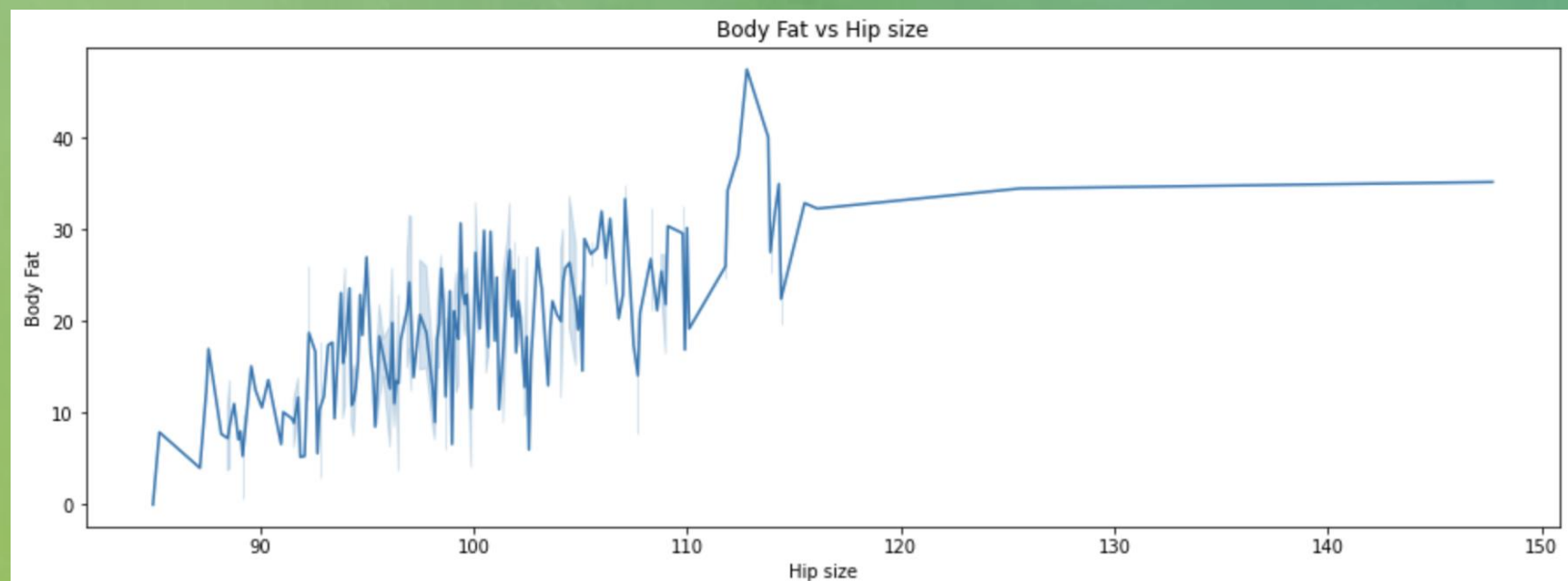
Body fat – Chest(가슴)



지도 학습 – Dataset Analysis

```
plt.figure(figsize=(15,5))
sns.lineplot(x='Hip',y='BodyFat',data=Body_fat)
plt.title('Body Fat vs Hip size')
plt.xlabel('Hip size')
plt.ylabel('Body Fat')
plt.show()
```

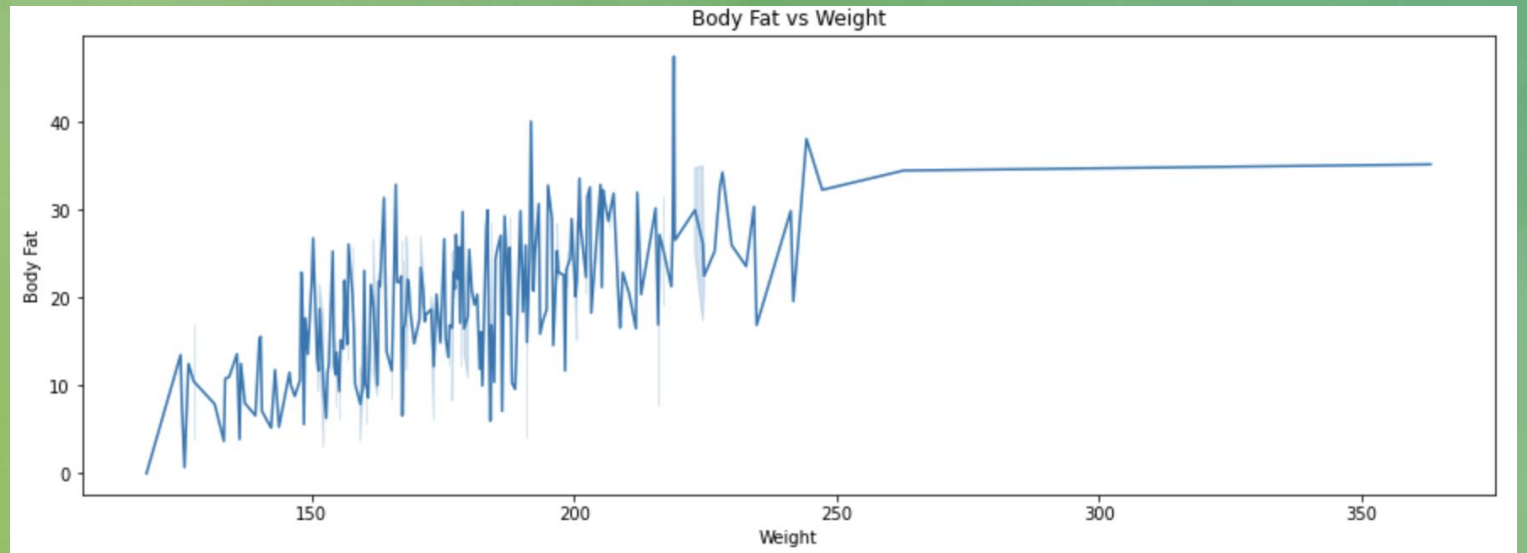
Body fat – Hip(힙)



지도 학습 – Dataset Analysis

```
plt.figure(figsize=(15,5))
sns.lineplot(x='Weight',y='BodyFat',data=Body_fat)
plt.title('Body Fat vs Weight')
plt.xlabel('Weight')
plt.ylabel('Body Fat')
plt.show()
```

Body fat – Weight(몸무게)



지도 학습 – Dataset Analysis

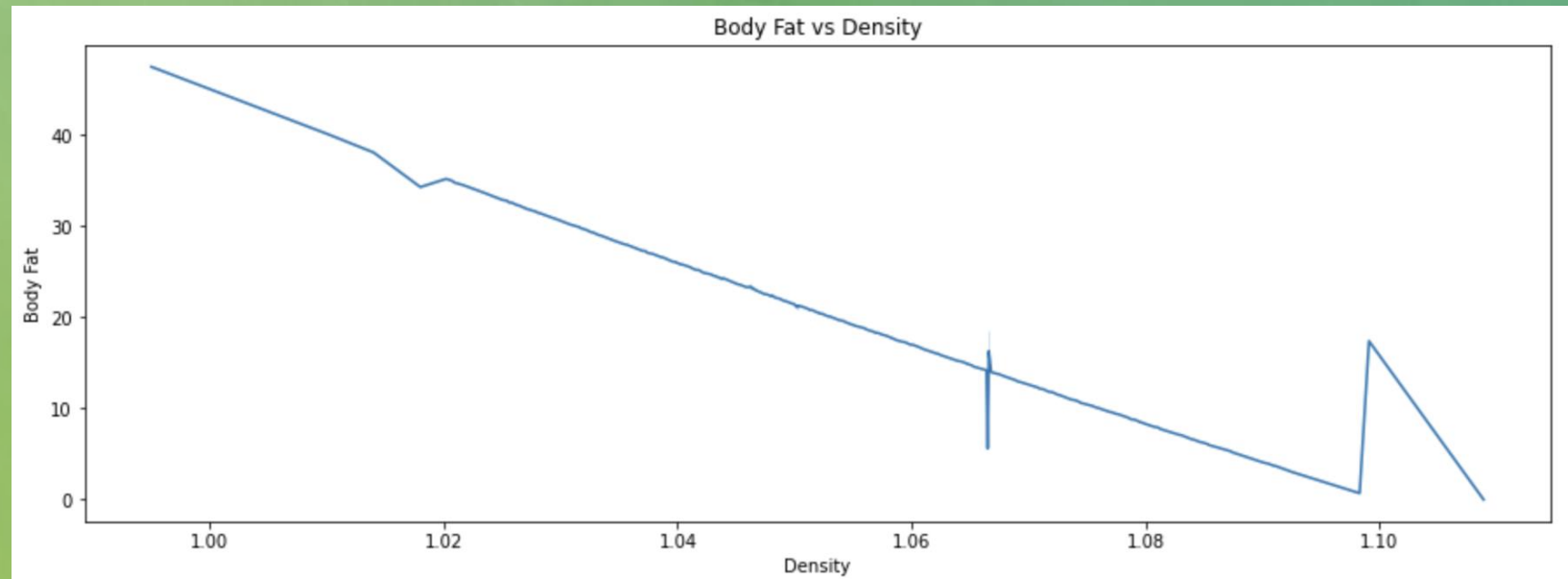
+

○

●

```
plt.figure(figsize=(15,5))
sns.lineplot(x='Density',y='BodyFat',data=Body_fat)
plt.title('Body Fat vs Density')
plt.xlabel('Density')
plt.ylabel('Body Fat')
plt.show()
```

Body fat – 수중 체밀도
(Density)

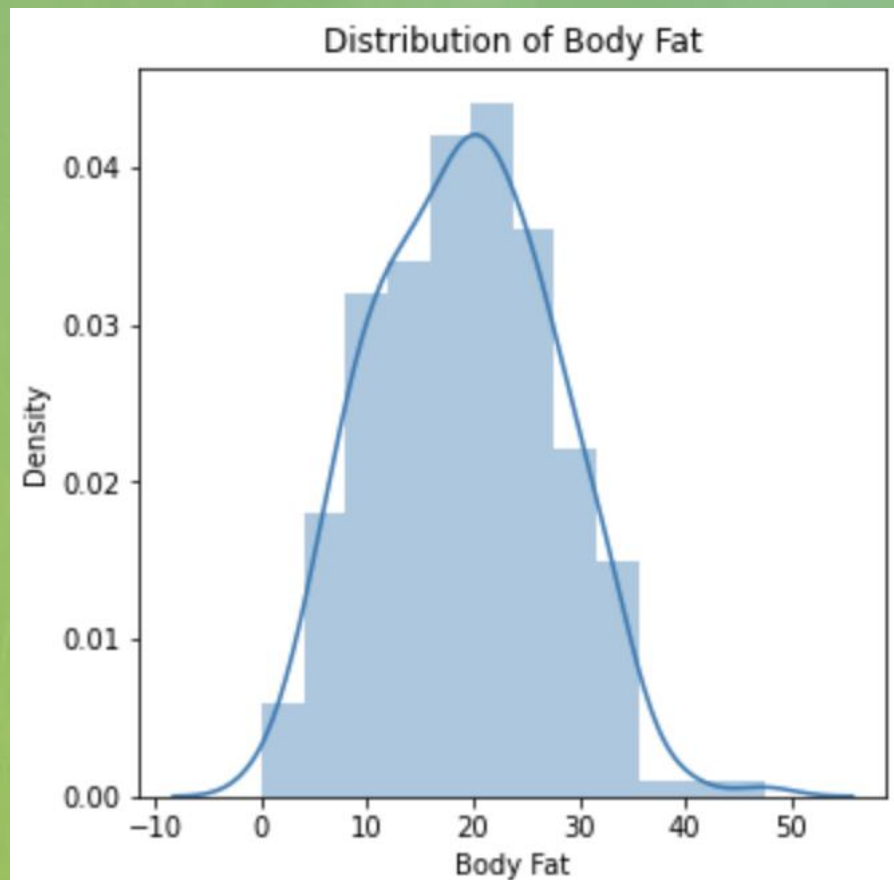


지도 학습 - Dataset Analysis

```
plt.figure(figsize=(5,5))
sns.distplot(Body_fat['BodyFat'])
plt.title('Distribution of Body Fat')
plt.xlabel('Body Fat')
plt.show()
```

체지방률과 긍정적 상관관계를 보여주는
변수들은 모두 비슷한 패턴을 보였음

지도 학습 – Dataset Analysis



정규 분포표와 유사한 형태의 데이터셋
-> 선형 모델에 적합

```
plt.figure(figsize=(5,5))  
sns.distplot(Body_fat['BodyFat'])  
plt.title('Distribution of Body Fat')  
plt.xlabel('Body Fat')  
plt.show()
```

지도 학습 - Model Training

```
X = Body_fat.drop('BodyFat',axis=1)
y = Body_fat['BodyFat']
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.7,random_state=100)
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
col = X_train.columns
X_train = scaler.fit_transform(X_train[col])
X_train = pd.DataFrame(X_train,columns=col)
X_train.head()
```

X축 타겟변수 - 나머지
Y축 타겟 변수 - Body fat
train_test_split -> 훈련세트 /
테스터세트 분할
StandardScaler - 데이터 표준화

	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
0	1.252498	-1.258781	-1.753094	-0.645491	-0.922137	-0.898873	-0.997424	-0.786300	-1.480711	-1.370083	-1.034330	-0.870232	-1.615847
1	0.395968	-0.796605	-1.379440	-0.131256	-0.126958	-0.347837	-1.093501	-0.671803	-1.312226	-1.614455	0.396390	0.269063	-1.072913
2	-0.071230	1.382221	-0.258479	1.094997	2.281300	1.865342	0.937843	0.740328	0.667481	-0.820245	0.805167	0.962547	-0.964326
3	-0.071230	0.911793	0.955895	1.490562	0.463747	0.275467	0.237853	0.702162	0.456874	0.584897	1.656786	-2.802081	1.207412
4	-0.071230	0.094733	1.329549	-0.091699	-0.081519	-0.374937	0.320205	0.186925	0.119902	0.523804	0.157937	0.219528	0.121543

지도 학습 - Prediction Result

```
from sklearn.linear_model import Ridge
from sklearn.model_selection import GridSearchCV
Model = Ridge(random_state=100)

params = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100, 1000]}

grid_search = GridSearchCV(estimator=Model, param_grid=params, verbose=1, n_jobs=-1, scoring='r2')
grid_search.fit(X_train, y_train)
```

Linear_model -> Ridge

alpha -> grid search 통해 탐색할
하이퍼 파라미터의 후보값들 셋팅

fit -> 그리드 서치를 실행하여 최적의
alpha 값 찾기

지도 학습 - Prediction Result

```
Model_best = grid_search.best_estimator_  
  
from sklearn.metrics import r2_score  
y_train_pred = Model_best.predict(X_train)  
print(r2_score(y_train, y_train_pred))  
  
y_test_pred = Model_best.predict(X_test)  
print(r2_score(y_test, y_test_pred))
```

0.7634335602965887

0.6824226929347279

훈련 데이터에 대한 R2 값 출력

테스트 데이터에 대한 R2 값 출력

결과 - 테스트 데이터에 대한 R2 값의 차이가 크지 않음.

최종 결론

신체 조건 과 비만율 은 예측 가능하며
높은 상관 관계가 있음.

소감

- 탐구 주제 안에서 데이터 시각화를 통해 분석하는 과정 안에서 흥미를 찾게 됨
- 이미 만들어진 AI 라이브러리를 활용하며, 라이브러리의 중요성을 깨달음

감사합니다

+

o

•