

week 1

1. 머신러닝, 딥러닝 이진 분류

스케일링이 필요한 알고리즘: 로지스틱 회귀, k 최근접, pca 등등

필요없는 알고리즘: 트리 형태의 알고리즘 예를들어 결정 트리 랜덤 포레스트, 부스팅 형식 등등

하지만 스케일링이 필요 없는 알고리즘에서 스케일링을 했다고 해서 문제될 거 없음

그래서 결론: 그냥 스케일링 박자

```
dataset=pd.read_csv('diabetes.csv')
x=dataset.iloc[:, :-1].values
y=dataset.iloc[:, -1].values
```

[::-1] 앞이 행, 뒤애가 열 해석하면

모든 행 선택, 열은 마지막 행 제외 모두 선택

→ x는 특성들

[::-1] 그럼 애는

모든 행 선택, 마지막 행만 선택

→y는 레이블, 타깃값

2. 딥러닝 머신러닝 다중 분류

```
x=dataset.iloc[:,1:-1].values
```

```
y=dataset.iloc[:, -1].values
```

[::1:-1] 행은 모두 선택, 열은 첫번째 마지막 열 빼고 모두 선택

왜? id 값은 필요 없으므로 드롭하기 위해

[::-1] 모든 행 선택, 열은 마지막 행만 선택

```
# Encode "Sex" column
```

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder

ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [0])],
remainder='passthrough')
x = np.array(ct.fit_transform(x))
print(x)
```

원핫 인코딩은 범주형 데이터를 이진 벡터로 변환하는 것

ring은 범주형 데이터이므로 딥러닝 이진분류 시행하는것은 바람직하지 않음

→ 딥러닝 회귀를 시행해야함

3. 머신러닝 딥러닝 타깃값이 0과 1인 데이터를 회귀

```
dataset=pd.read_csv('diabetes.csv')
x=dataset.iloc[:,[0,1,2,3,4,6,7]]
y=dataset.iloc[:,5]
```

회귀 진행하기 위해 0,1,2,3,4,6,7번째 열을 데이터 값으로 선택

나머지 5번째 열을 타깃 데이터로 선택

활성화 함수의 선택

- **회귀 문제:** 보통 출력층에 활성화 함수를 사용하지 않거나 선형 함수를 사용합니다
- **이진 분류:** 출력층에 **시그모이드 함수**를 사용해 확률 값을 출력합니다.
- **다중 클래스 분류:** 출력층에 **소프트맥스 함수**를 사용해 각 클래스에 속할 확률을 구합니다.
- **중간층:** 보통 **ReLU**나 **Leaky ReLU**가 많이 사용되며, 이를 통해 비선형성을 추가하고 더 복잡한 패턴을 학습할 수 있습니다.

```
model.compile(optimizer="adam", loss="mse", metrics=["mse"])
```

여기서 mse란 평균 제곱오차로 주로 회귀 문제에서 사용되며 예측값(y_{pred})과 실제값(y_{test})의 차이를 제곱하여 평균을 구한다,

4. 머신러닝 딥러닝 다중 회귀