실무 데이터 분석
(로지스틱회귀 & MLP)

인공지능 기반 스마트 설계
컴퓨터 AI공학부
천세진

DATA SCIENCE LABS

데이터

품　　명: PARKING SPRAG(8속)_<열전>　　　품　　번: 45926-4G100
측정시간: 2023.10.05. 22:40:08　　　측 정 자: 양정훈
특기사항: 231005_일상검사_야_초_1-2-1_OK

번호	항 목	측정값	기준값	상한공차	하한공차	편 차	판 정
3	평면1						
	평면도	0.002	0.100			Total	|+
	SMmf	4P	0.001	0.001	-0.001	0.002	
5	원1(I) <상>						
	D	16.496	16.485	0.030	0.000	0.011	-|
	SMmf	4P	0.000	0.000	0.000	0.001	
6	원2(I) <중>						
	D	16.500	16.485	0.030	0.000	0.015	|+
	SMmf	4P	0.002	0.002	-0.002	0.004	
7	원3(I) <하>						
	D	16.502	16.485	0.030	0.000	0.017	|+
	SMmf	4P	0.000	0.000	0.000	0.001	
8	원통1(I) <- 원1, 원2, 원3의 측정점 병합						
	D	16.499	16.485	0.030	0.000	0.014	-|
	원통도	0.005	0.000				
	직각도	0.021	0.050		평면1		|++
	SMmf	12P	0.002	0.003	-0.002	0.005	
14	점2 <- 점1의 되부름 <열전 관리치수(Spec : 116.6±0.1)>						
	X	116.644	116.600	0.100	0.000	0.044	-|
	Y	-10.904	10.900	0.100	-0.100	0.004	|+
16	각도1 <- 각도[XYPLAN]:직선2와 직선3						
	Ang	57.129	57.000	0.333	-0.333	0.129	|++
17	직선4 <27° 소재>						
	Y/X	27.226	27.000	0.500	-0.500	0.226	|++

데이터 처리 단계

1. 데이터 이해
(목적, 구성, 특징)

2. 데이터 전처리
(결측값, 이상치, 중복값)

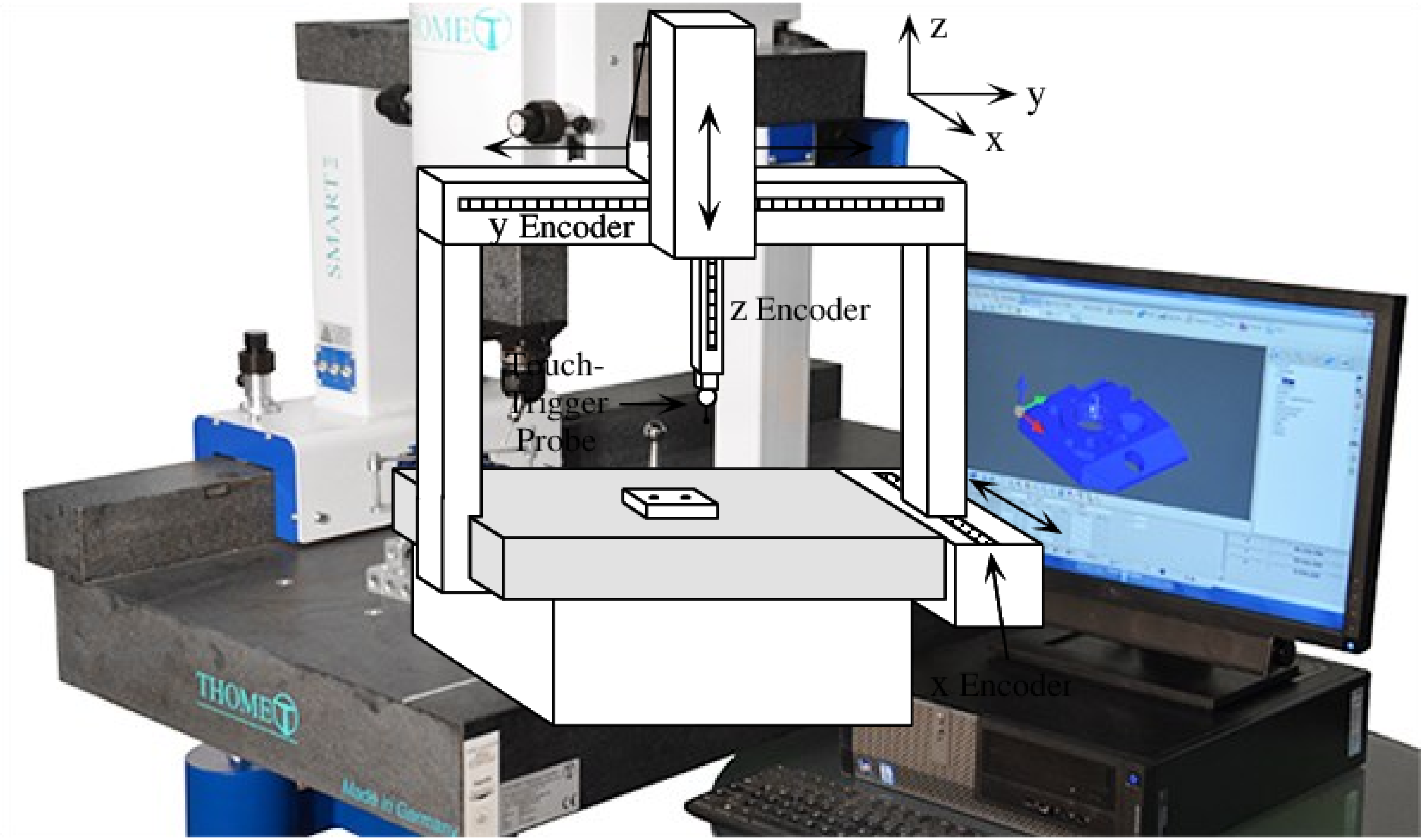
3. 데이터 탐색
(데이터의 분포, 상관관계, 이상치 탐색)

4. 통계적인 분석
(Aggregation/Summarization)

5. 시각화

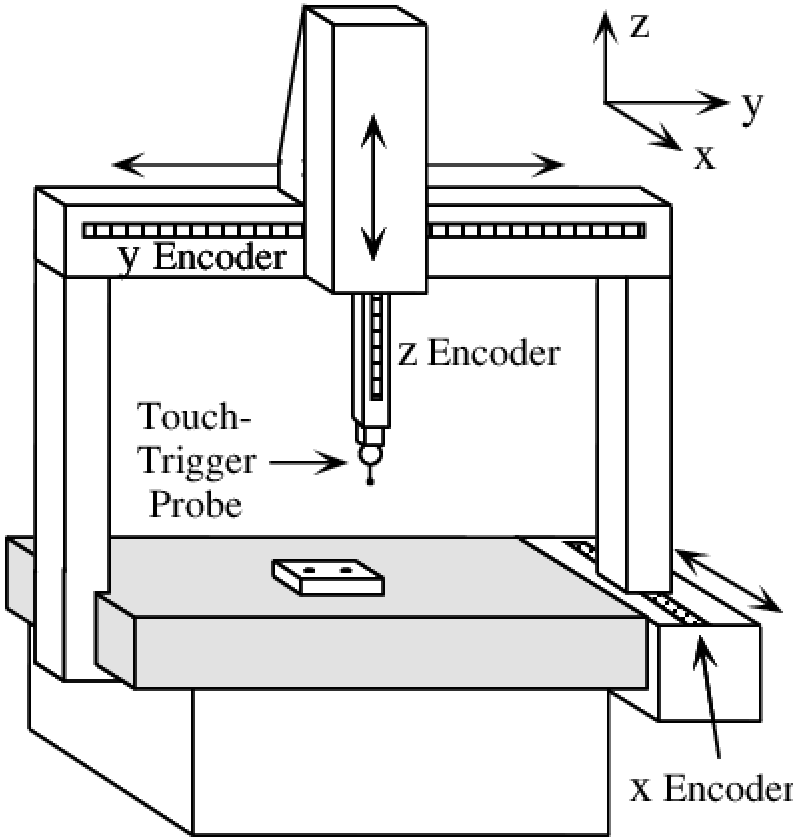
6. 결론 도출

1. 데이터 이해



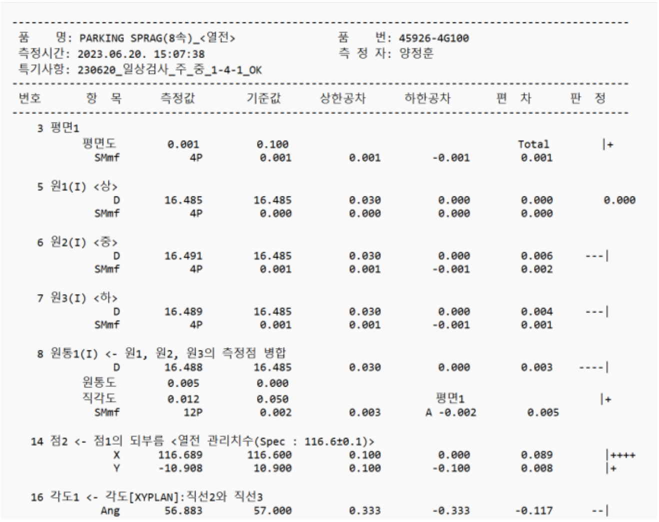
DATA SCIENCE LABS

1. 데이터 이해



DATA SCIENCE LABS

Traditional workflow



```
품    명: PARKING SPRAG(8속)_<열전>                 품    번: 45926-4G100
측정시간: 2023.06.20. 15:07:38                      측 정 자: 양정훈
특기사항: 230620_일상검사_주_중_1-4-1_OK
-------------------------------------------------------------------------------
번호      항 목      측정값      기준값     상한공차     하한공차      편 차      판 정
-------------------------------------------------------------------------------
  3 평면1
       평면도      0.001       0.100                                Total       |+
        SMmf          4P       0.001      0.001      -0.001       0.001

  5 원1(I) <상>
           D      16.485      16.485      0.030       0.000       0.000       0.000
        SMmf          4P       0.000      0.000       0.000       0.000

  6 원2(I) <중>
           D      16.491      16.485      0.030       0.000       0.006    ---|
        SMmf          4P       0.001      0.001      -0.001       0.002

  7 원3(I) <하>
           D      16.489      16.485      0.030       0.000       0.004    ---|
        SMmf          4P       0.001      0.001      -0.001       0.001

  8 원통1(I) <- 원1, 원2, 원3의 측정점 병합
           D      16.488      16.485      0.030       0.000       0.003   ----|
       원통도      0.005       0.000
       직각도      0.012       0.050                  평면1                  |+
        SMmf         12P       0.002      0.003    A -0.002       0.005

 14 점2 <- 점1의 되부름 <열전 관리치수(Spec : 116.6±0.1)>
           X     116.689     116.600      0.100       0.000       0.089       |++++
           Y     -10.908      10.900      0.100      -0.100       0.008       |+

 16 각도1 <- 각도[XYPLAN]:직선2와 직선3
         Ang      56.883      57.000      0.333      -0.333      -0.117     --|
```

3D측정기1
(CMM: Coordinate Measuring Machine)

작업자1

CMM data

데이터분석자1

3D측정기2

작업자2

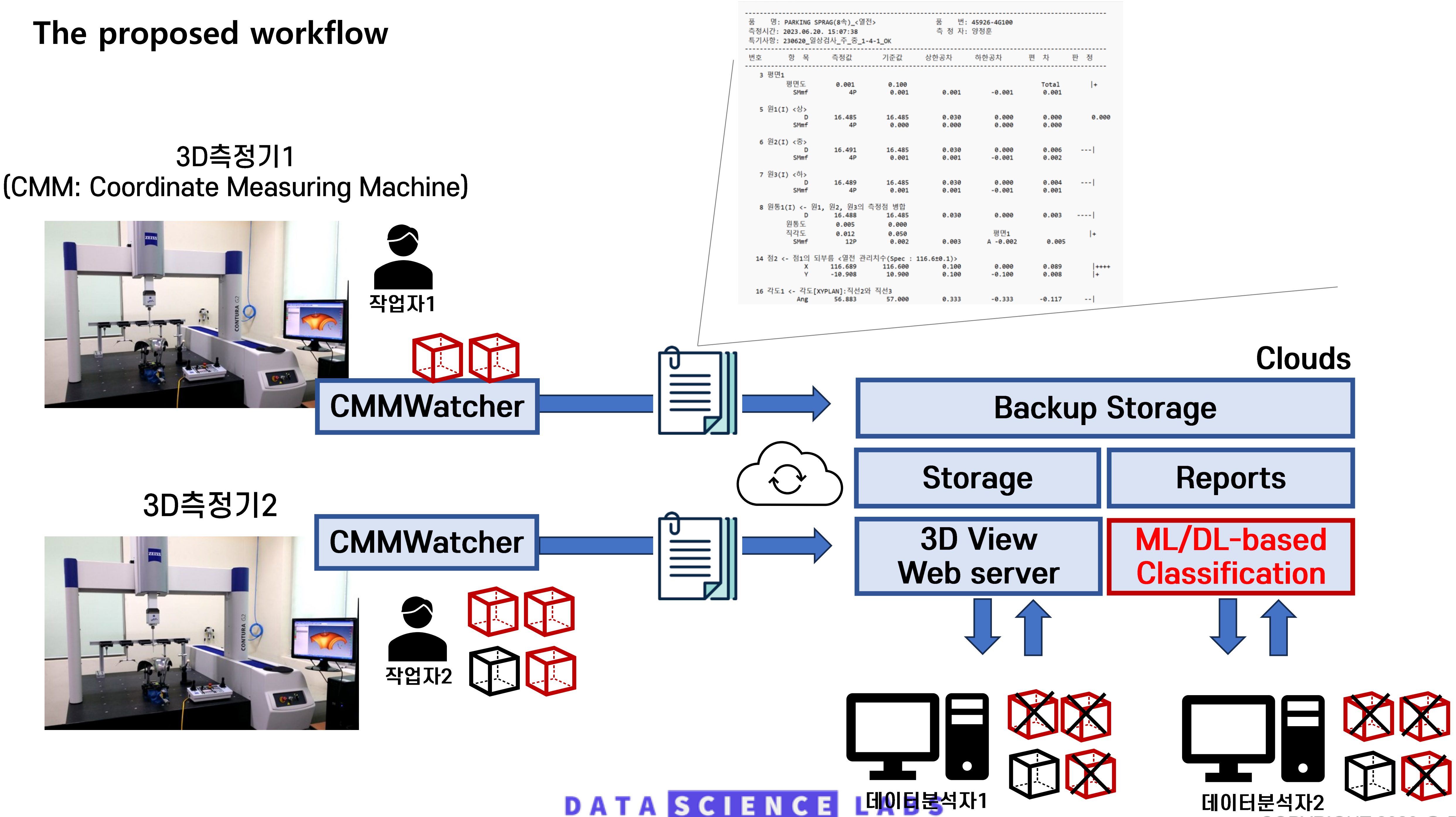
수작업: Copy 및 데이터 확인

데이터분석자2

수기 기록: 문서 증가

DATA SCIENCE LABS

The proposed workflow



1 데이터 이해

```
----------------------------------------------------------------------------------------
품    명: PARKING SPRAG(8속)_<열전>                 품    번: 45926-4G100
측정시간: 2023.10.05. 22:40:08                      측 정 자: 양정훈
특기사항: 231005_일상검사_야_초_1-2-1_OK
----------------------------------------------------------------------------------------
번호        항  목       측정값        기준값      상한공차      하한공차       편  차     판  정
----------------------------------------------------------------------------------------
  3 평면1
         평면도        0.002        0.100                                   Total      |+
           SMmf           4P        0.001       0.001        -0.001        0.002

  5 원1(I) <상>
              D       16.496
           SMmf           4P

  6 원2(I) <중>
              D       16.500
           SMmf           4P

  7 원3(I) <하>
              D       16.502       16.485       0.030         0.000        0.017      |+
           SMmf           4P        0.000       0.000         0.000        0.001

  8 원통1(I) <- 원1, 원2, 원3의 측정점 병합
              D       16.499       16.485       0.030         0.000        0.014     -|
         원통도        0.005        0.000
         직각도        0.021        0.050                     평면1                    |++
           SMmf          12P        0.002       0.003        -0.002        0.005

 14 점2 <- 점1의 되부름 <열전 관리치수(Spec : 116.6±0.1)>
              X      116.644      116.600       0.100         0.000        0.044     -|
              Y      -10.904       10.900       0.100        -0.100        0.004      |+

 16 각도1 <- 각도[XYPLAN]:직선2와 직선3
            Ang       57.129       57.000       0.333        -0.333        0.129      |++

 17 직선4 <27° 소재>
            Y/X       27.226       27.000       0.500        -0.500        0.226      |++
```


```
----------------------------------------------------------------------------------------
번호        항  목       측정값        기준값      상한공차      하한공차       편  차     판  정
----------------------------------------------------------------------------------------
  3 평면1
         평면도        0.002        0.100                                   Total      |+
           SMmf           4P        0.001       0.001        -0.001        0.002
```

1. 데

```
--------------------------------------------------------------------------------
품      명: PARKING SPRAG(8속)_<열전>              품      번: 45926-4G100
측정시간: 2023.10.05. 22:40:08                     측 정 자: 양정훈
특기사항: 231005_일상검사_야_초_1-2-1_OK
```

- **데이터 변환**

품명	품번	측정시간	측정자	특기사항	품질상태

```
---------------------------------------------------------------------------------------
번호        항  목      측정값       기준값      상한공차      하한공차       편  차      판  정
---------------------------------------------------------------------------------------
   3 평면1
          평면도        0.002        0.100                                  Total       |+
           SMmf           4P        0.001       0.001        -0.001        0.002
```

번호	항목	측정값	기준값	상한공차	하한공차	편차	판정

DATA SCIENCE LABS

1. 데이터 이해

• **데이터 피봇팅(Pivot)**

품명	품번	측정시간	측정자	특기사항	품질상태
PARKING SPRAG(8속)_<열전>	45926-4G100	2023.10.05. 22:40:08	양정훈	231005_일상검사_야_초_1-2-1_OK	OK

번호	항목	측정값	기준값	상한공차	하한공차	편차	판정
3 평면1	평면도	0.002	0.100				|+
3 평면1	SMmf	4P	0.001			0.002	
5 원1(I) <상>	D	16.496	16.485	0.030	0.000	0.011	-|
5 원1(I) <상>	SMmf	4P	0.000	0.000	0.000	0.001	

품명	품번	측정시간	측정자	특기사항	품질상태
PARKING SPRAG(8속)_<열전>	45926-4G100	2023.10.05. 22:40:08	양정훈	231005_일상검사_야_초_1-2-1_OK	OK

편차_3 평면1_평면도	편차_3 평면1_SMmf	편차_5 원1(I) <상>_D	편차_5 원1(I) <상>_SMmf
	0.002	0.011	0.001

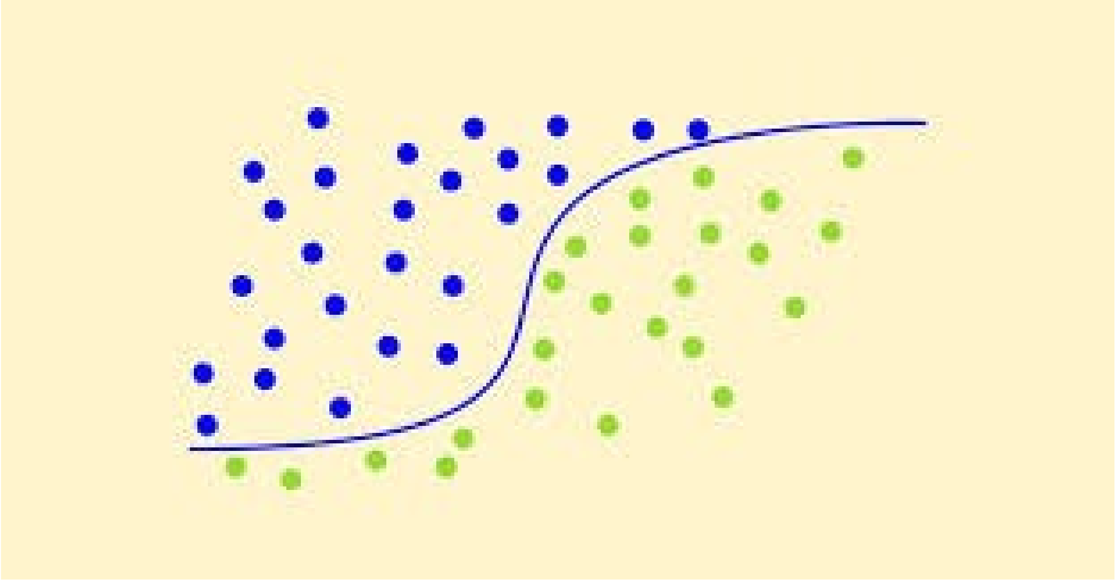
번호	항목	측정값	기준값	상한공차	하한공차	편차	판정
3 평면1	평면도	0.002	0.100				|+
3 평면1	SMmf	4P	0.001			0.002	
5 원1(I) <상>	D	16.496	16.485	0.030	0.000	0.011	-|
5 원1(I) <상>	SMmf	4P	0.000	0.000	0.000	0.001	

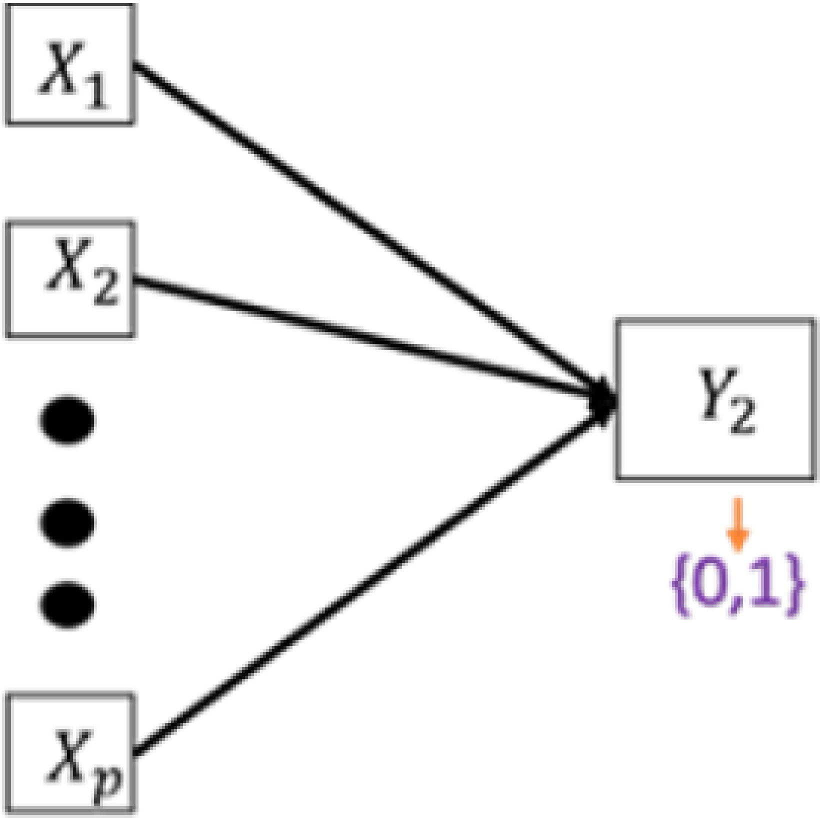
1. 데이터 이해

- 적절한 X | y 형태의 데이터를 구성

```
품명,편차_각도1 <- 각도[XYPLAN]:직선2와 직선3 Ang,편차_각도2 <- 각도[XYPLAN]:직
PARKING SPRAG(8속)_<열전>,0.303,-0.023,0.037,0.016,-0.001,0.037,0.010,0.014,0
PARKING SPRAG(8속)_<열전>,0.239,-0.009,0.046,0.021,0.030,0.046,0.010,0.014,0.
PARKING SPRAG(8속)_<열전>,0.129,0.067,0.046,0.051,0.040,0.046,0.011,0.015,0.0
PARKING SPRAG(8속)_<열전>,0.256,-0.148,0.053,0.017,0.022,0.053,0.012,0.018,0.
PARKING SPRAG(8속)_<열전>,0.205,-0.133,0.037,0.007,0.024,0.037,0.009,0.015,0.
PARKING SPRAG(8속)_<열전>,0.336,-0.014,0.034,-0.004,-0.006,0.034,0.010,0.013,
PARKING SPRAG(8속)_<열전>,0.242,0.013,0.050,0.030,0.008,0.050,0.012,0.017,0.0
```

4. 로지스틱 회귀





4. 로지스틱 회귀 vs. 선형회귀

- **선형회귀**
  - 연속적인 값을 예측하는 데 사용됩니다. 예를 들어, 집의 크기, 위치 등의 특성을 바탕으로 집값을 예측하는 경우에 사용됩니다.
  - 실수 값을 직접 출력합니다. 예를 들어 집값, 온도 등이 될 수 있습니다.
- **로지스틱 회귀**
  - 확률을 출력하며, 이 확률은 보통 특정 클래스에 속할 확률로 해석됩니다 (예: 스팸 메일일 확률).
  - **로지스틱 회귀: 범주형 결과(대개 이진 분류)를 예측하는 데 사용됩니다. 예를 들어, 이메일이 스팸인지 아닌지, 특정 질병의 유무 등을 예측할 때 사용합니다.**

DATA SCIENCE LABS

4. 로지스틱 회귀

```python
# 독립 변수와 종속 변수 분리
X = data_imputed.drop(columns=['품질상태'])
y = data_imputed['품질상태']

# 데이터를 훈련 세트와 테스트 세트로 분할
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

✓ 0.0s

```python
# 로지스틱 회귀 모델 생성 및 훈련
logistic_model = LogisticRegression()
logistic_model.fit(X_train, y_train)

# 테스트 세트에서 예측
y_pred = logistic_model.predict(X_test)

# 모델 평가
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)
```

4. 로지스틱 회귀

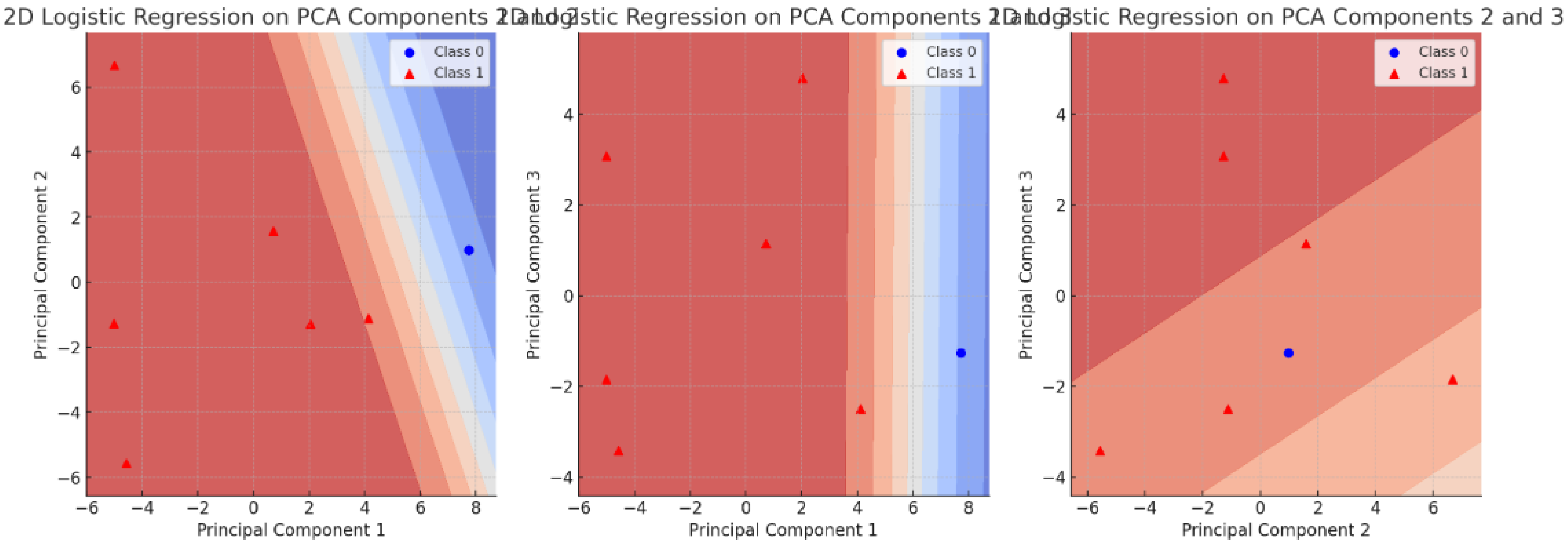
```
1.0                precision    recall  f1-score   support

         1.0       1.00      1.00      1.00         2

    accuracy                           1.00         2
   macro avg       1.00      1.00      1.00         2
weighted avg       1.00      1.00      1.00         2
```

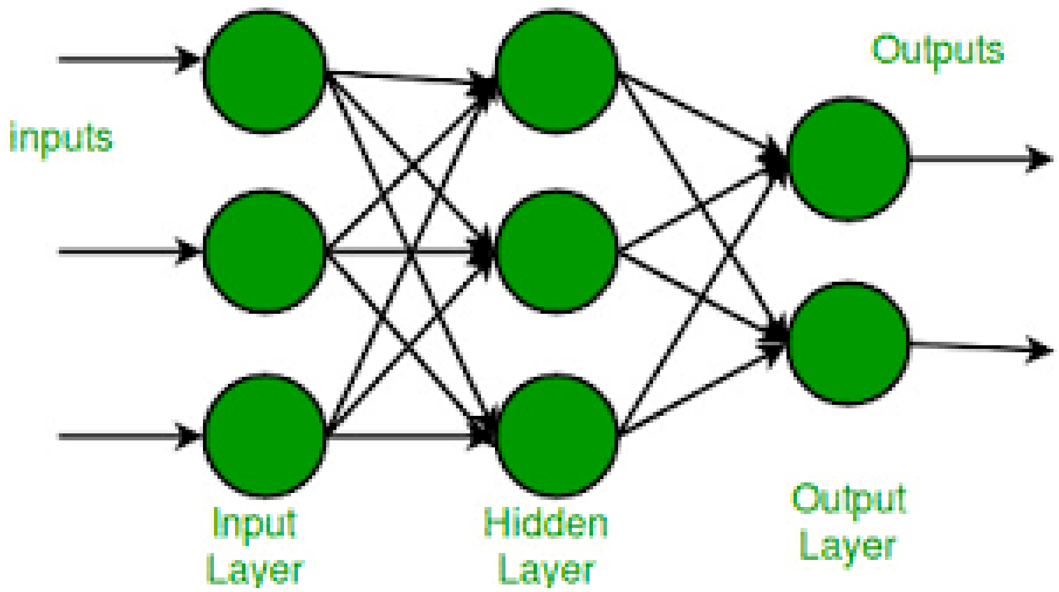
4. 로지스틱 회귀

• 차원축소를 통해 얻어진 주요 특징이 어디에 속하는지 보여줌



4. MLP

- **Multilayer perceptron**



데이터준비 → 모델 초기화 → 순전파 → 손실계산 → 역전파 → 가중치 갱신

4. MLP

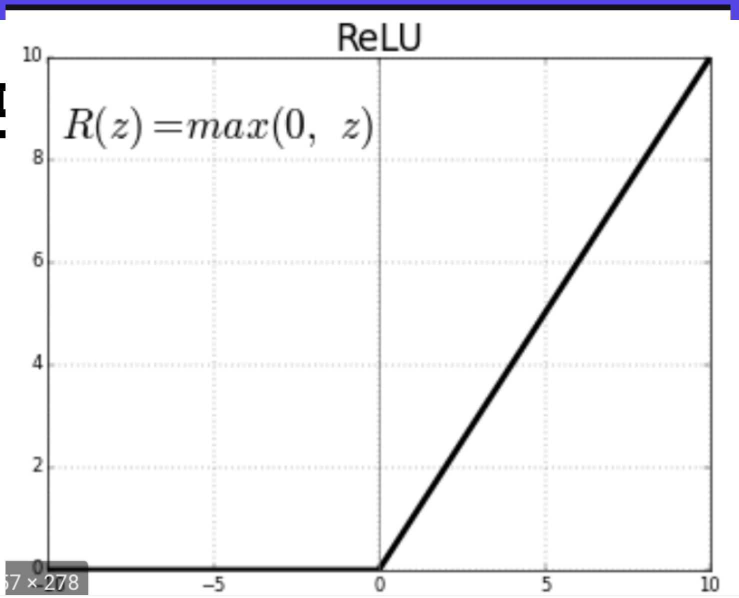
- 모델 초기화

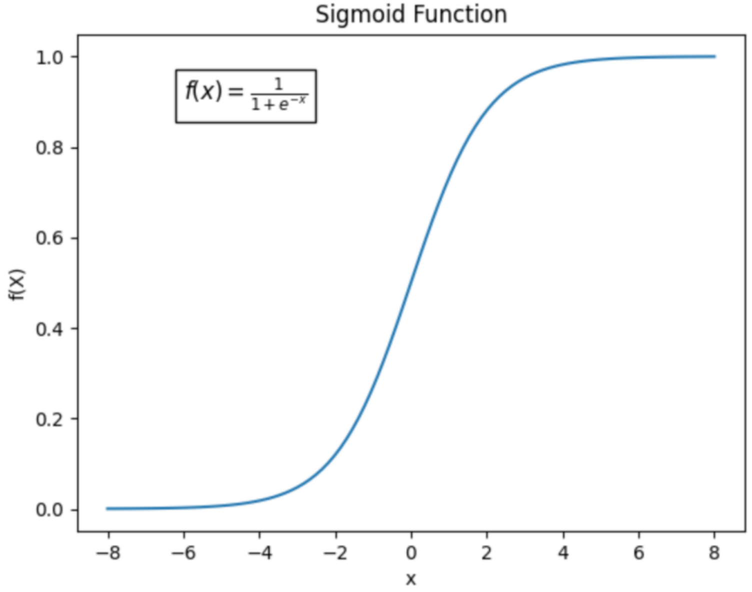
- 순전파

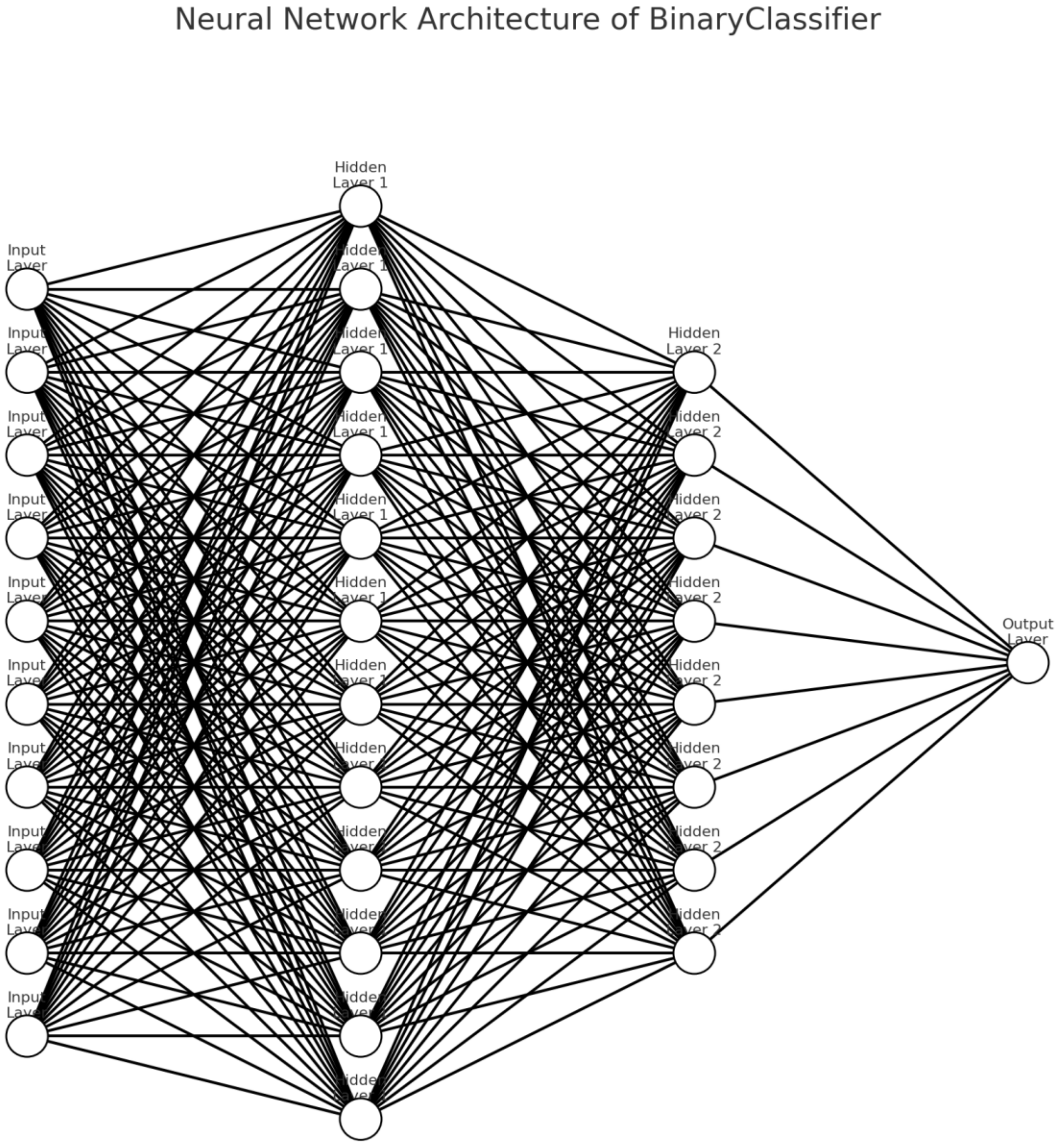
```python
# 모델 클래스 정의
class BinaryClassifier(nn.Module):
    def __init__(self):
        super(BinaryClassifier, self).__init__()
        self.fc1 = nn.Linear(X_train.shape[1], 12)
        self.fc2 = nn.Linear(12, 8)
        self.fc3 = nn.Linear(8, 1)
        self.sigmoid = nn.Sigmoid()

    def forward(self, x):
        x = torch.relu(self.fc1(x))
        x = torch.relu(self.fc2(x))
        x = self.sigmoid(self.fc3(x))
        return x
```

4. MLP


ReLU
$R(z) = max(0, z)$
10
8
6
4
2
0
−5
0
5
10


Sigmoid Function
$f(x) = \frac{1}{1+e^{-x}}$
f(x)
x
1.0
0.8
0.6
0.4
0.2
0.0
−8
−6
−4
−2
0
2
4
6
8


Neural Network Architecture of BinaryClassifier
Input Layer
Hidden Layer 1
Hidden Layer 2
Output Layer

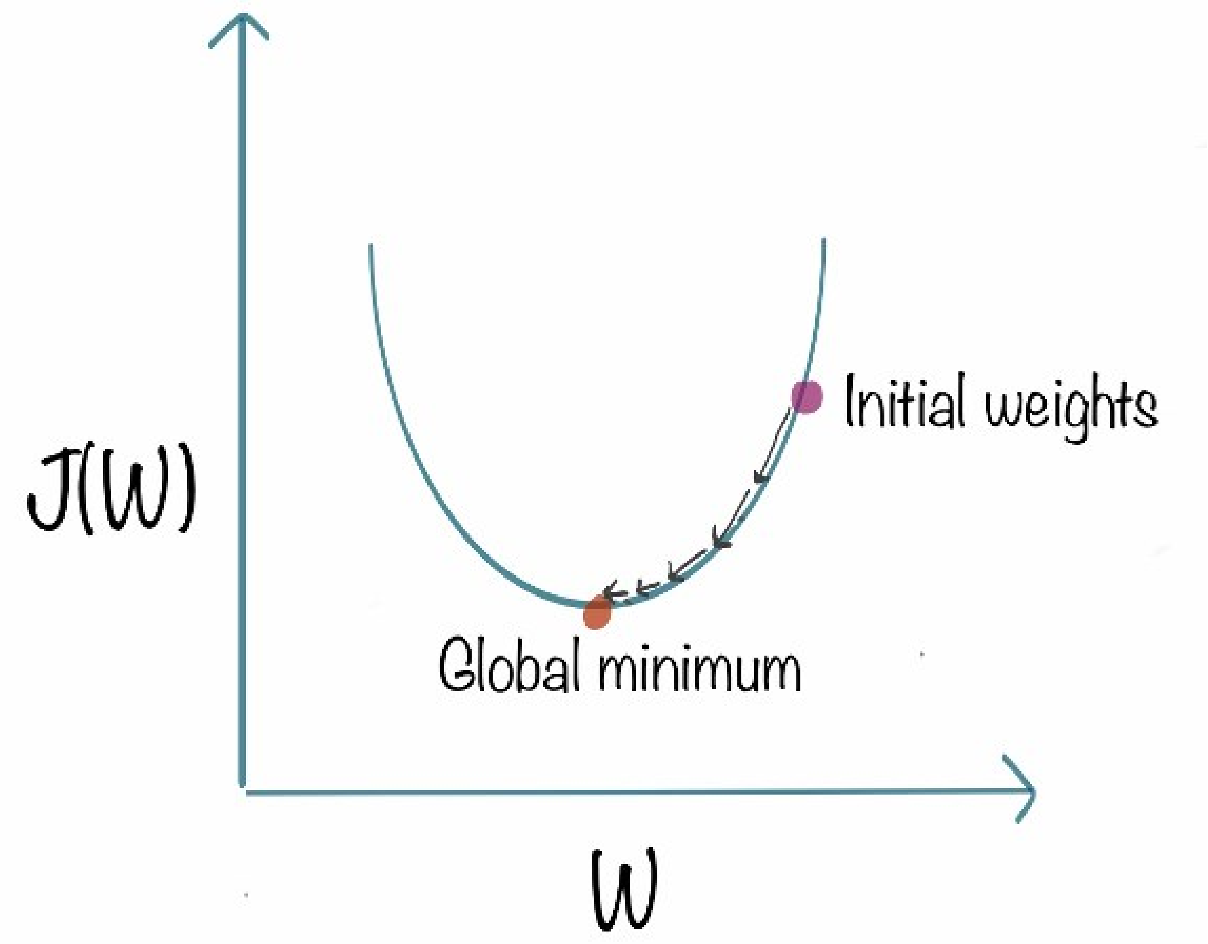
4. MLP

- **손실계산/ 역전파**

```
# 훈련 루프
for epoch in range(50):
    total_loss = 0
    correct = 0
    total = 0
    for data, target in train_loader:
        optimizer.zero_grad()
        output = model(data)
        loss = criterion(output, target.view(-1, 1))
        total_loss += loss.item()
        predicted = output.round()
        correct += (predicted == target.view(-1, 1)).sum().item()
        total += target.size(0)
        loss.backward()
        optimizer.step()

    epoch_loss = total_loss / len(train_loader)
    epoch_accuracy = correct / total
    epoch_losses.append(epoch_loss)
    epoch_accuracies.append(epoch_accuracy)

    print(f'Epoch {epoch+1}: Loss = {epoch_loss:.4f}, Accuracy = {epoch_accur
```



DATA SCIENCE LABS

4. MLP

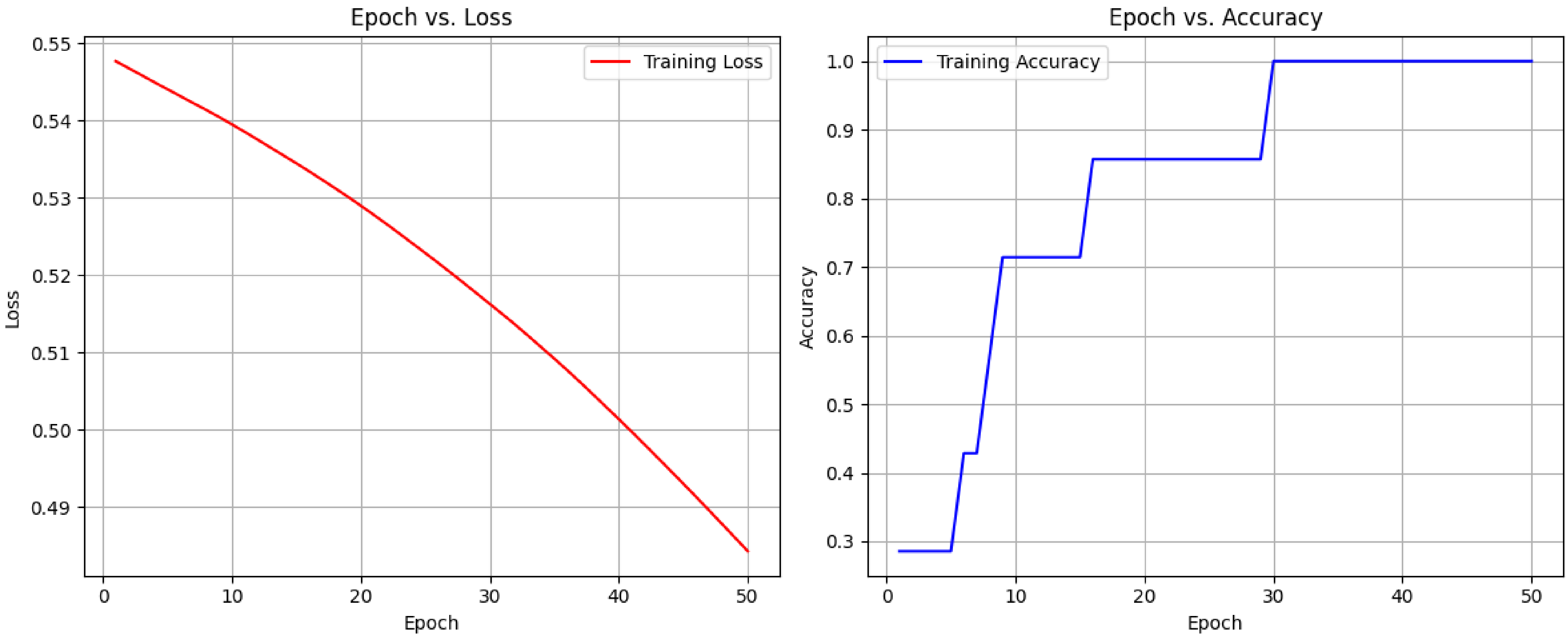
- **모델 평가**

```
# 모델 평가
with torch.no_grad():
    output = model(X_test)
    predicted = output.round()  # 0.5 이상을 1로, 미만을 0으로
    accuracy = (predicted == y_test.view(-1, 1)).sum().item() / len(y_test)
    print(f'Test Accuracy: {accuracy}')
```

```
Test Accuracy: 1.0
```

4. MLP

• **모델 평가**



DATA SCIENCE LABS