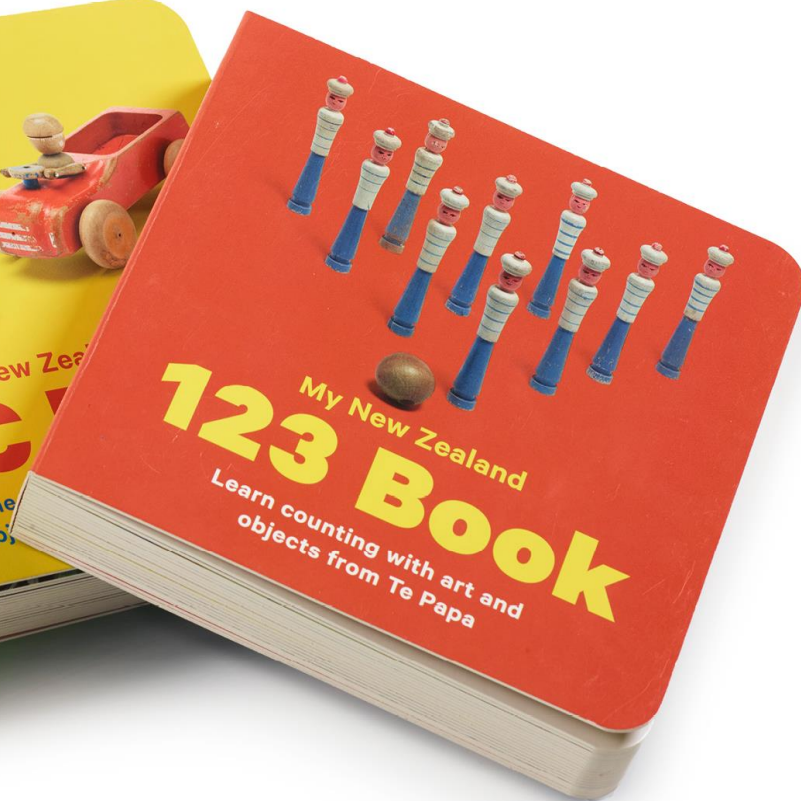
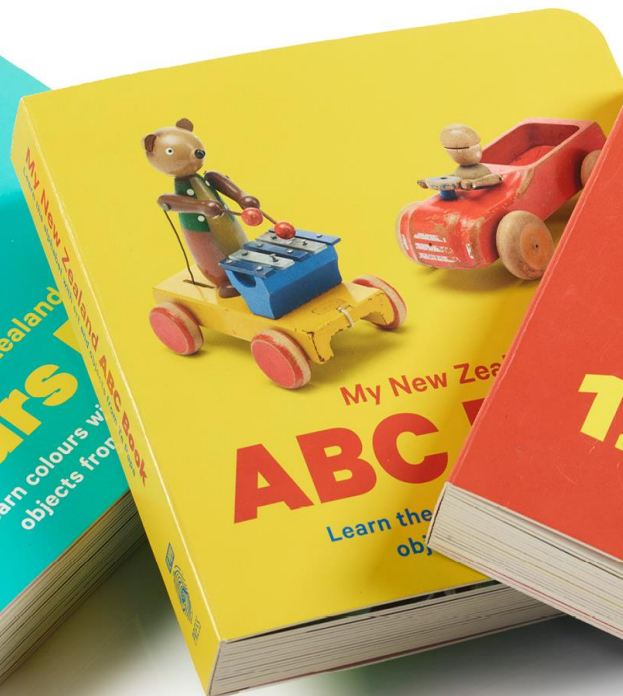


# AI와 머신러닝 분류(Classification)

제주대학교 변영철





개





토끼



돼지





오리



# 지도학습

## Supervised Learning



# 키 몸무게 발 크기 학년 성별



번호, 키, 몸무게, 발 크기, 학년, 성별

1,166,57,240,1,0

2,178,92,265,1,1

3,167,80,270,1,1

4,168,52,245,2,1

5,155,60,235,2,0

6,163,45,230,2,0

7,160,53,235,3,0

8,180,77,260,4,1

9,167,71,260,2,1

10,160,51,245,2,0

11,162,53,240,2,0

12,180,82,280,6,1

13,172,90,255,6,1

14,160,51,245,5,0

15,155,66,245,5,0

16,163,54,242,5,0

17,177,88,263,5,1

18,166,82,268,6,1

19,170,53,247,6,1

20,154,59,234,1,0

21,164,47,232,1,0

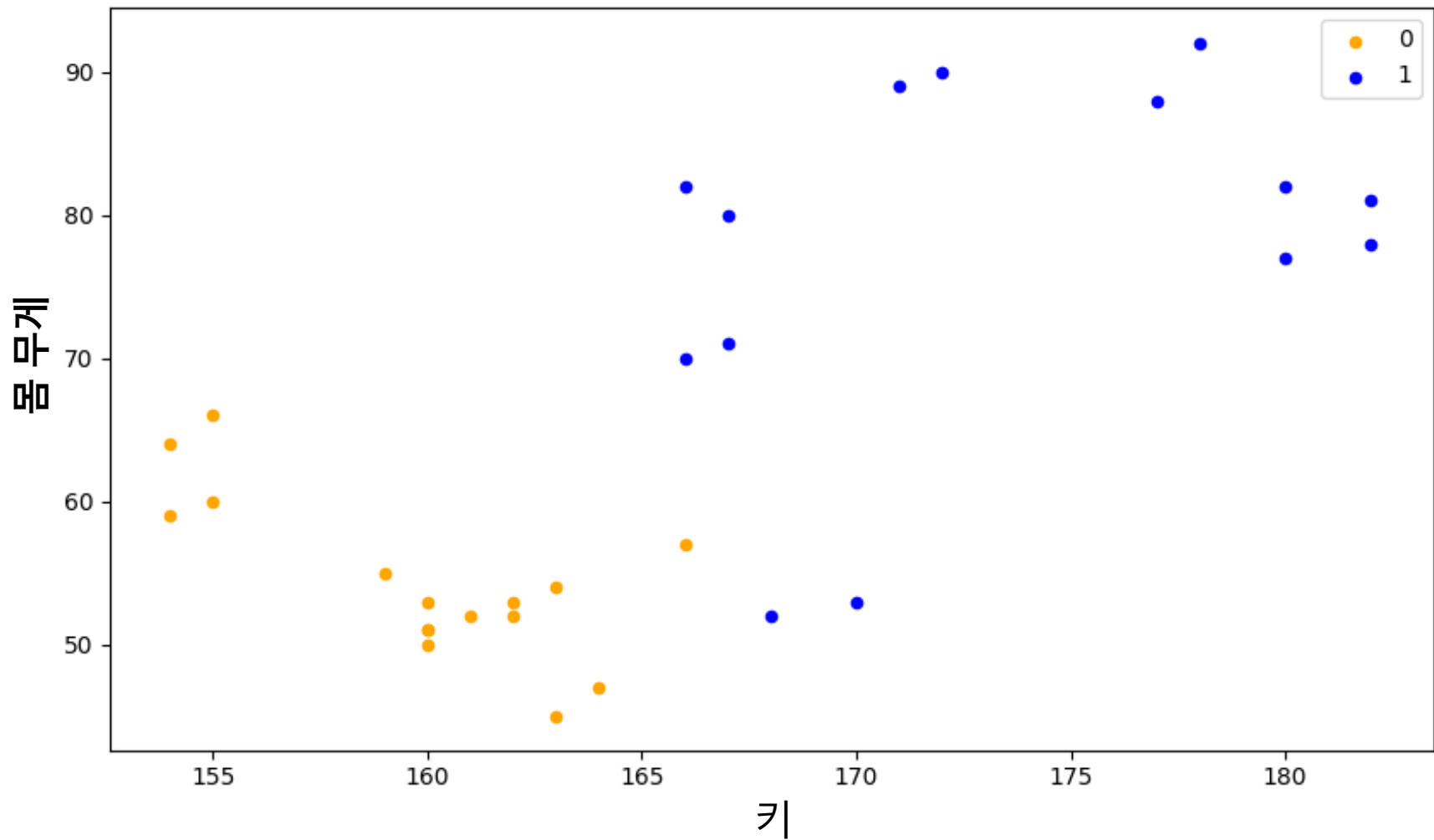


키에 따라  
몸무게는 어떻게 변할까?  
(성별에 따라 다른 색으로 표시)

**plot**(df, '키', '몸무게', '성별')

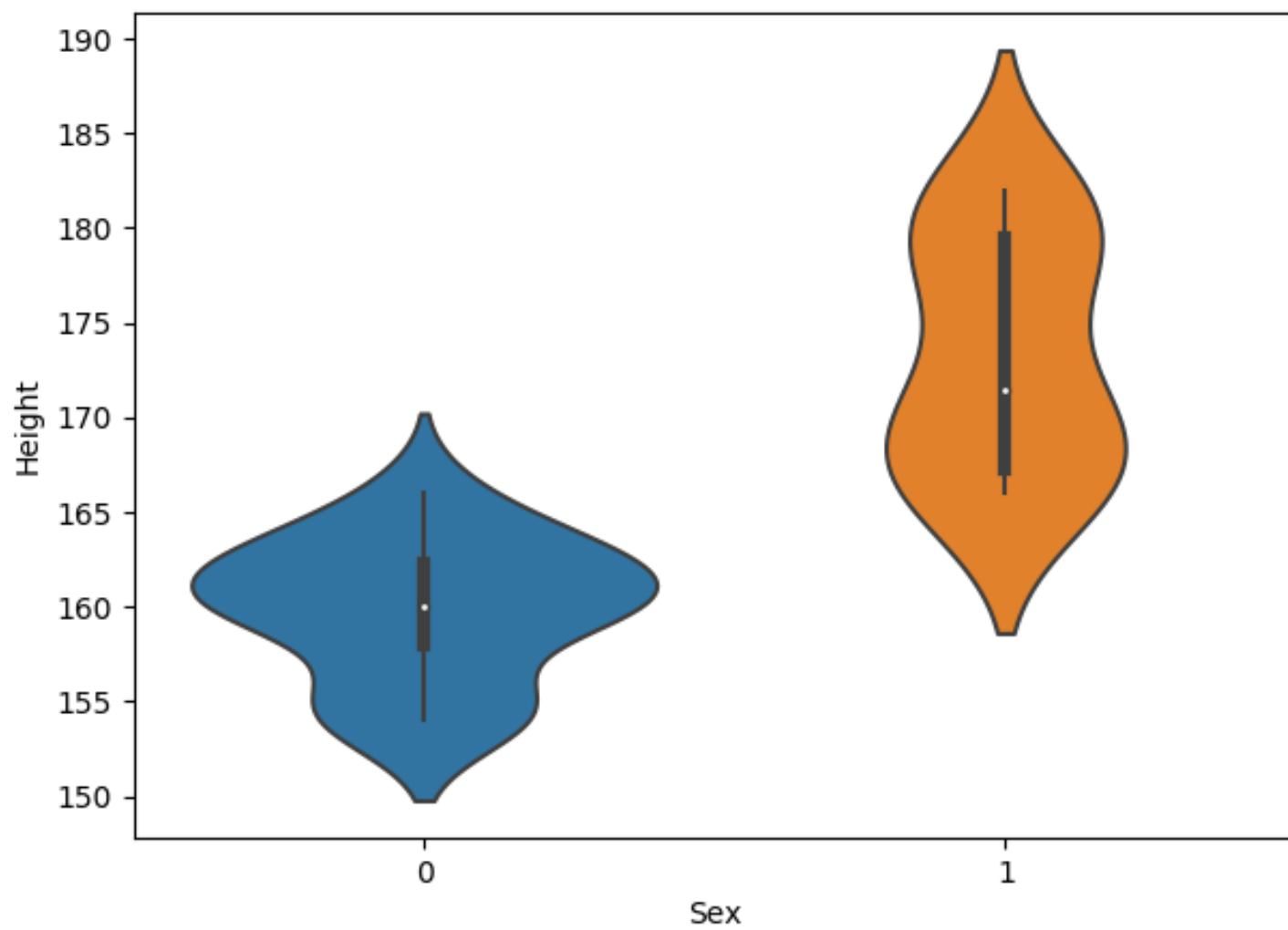


키 vs. 몸무게

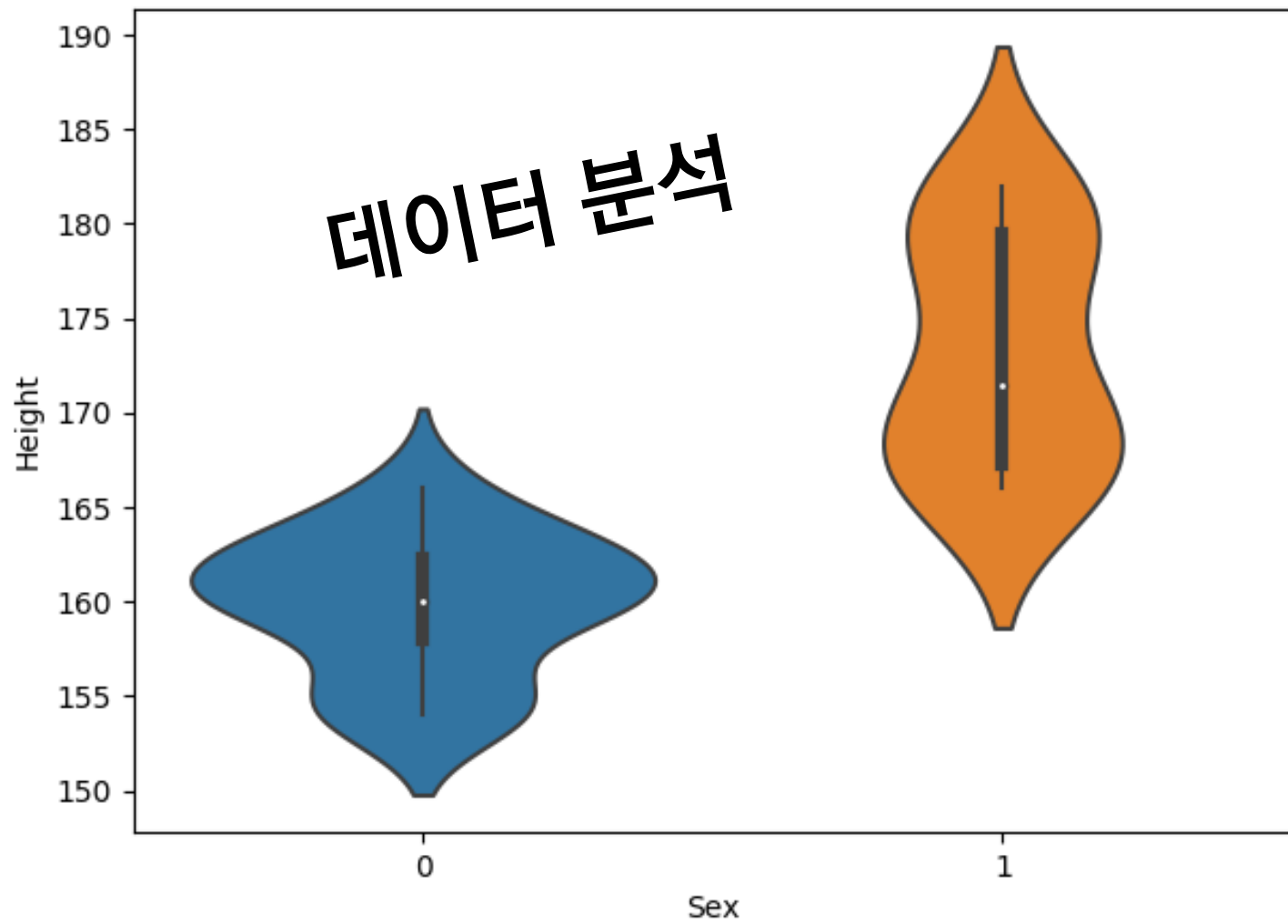


```
violinplot(df, '성별', '키')
```

성별에 따라 키가 어떻게 변하는지  
바이올린 모양으로 표시해보라!

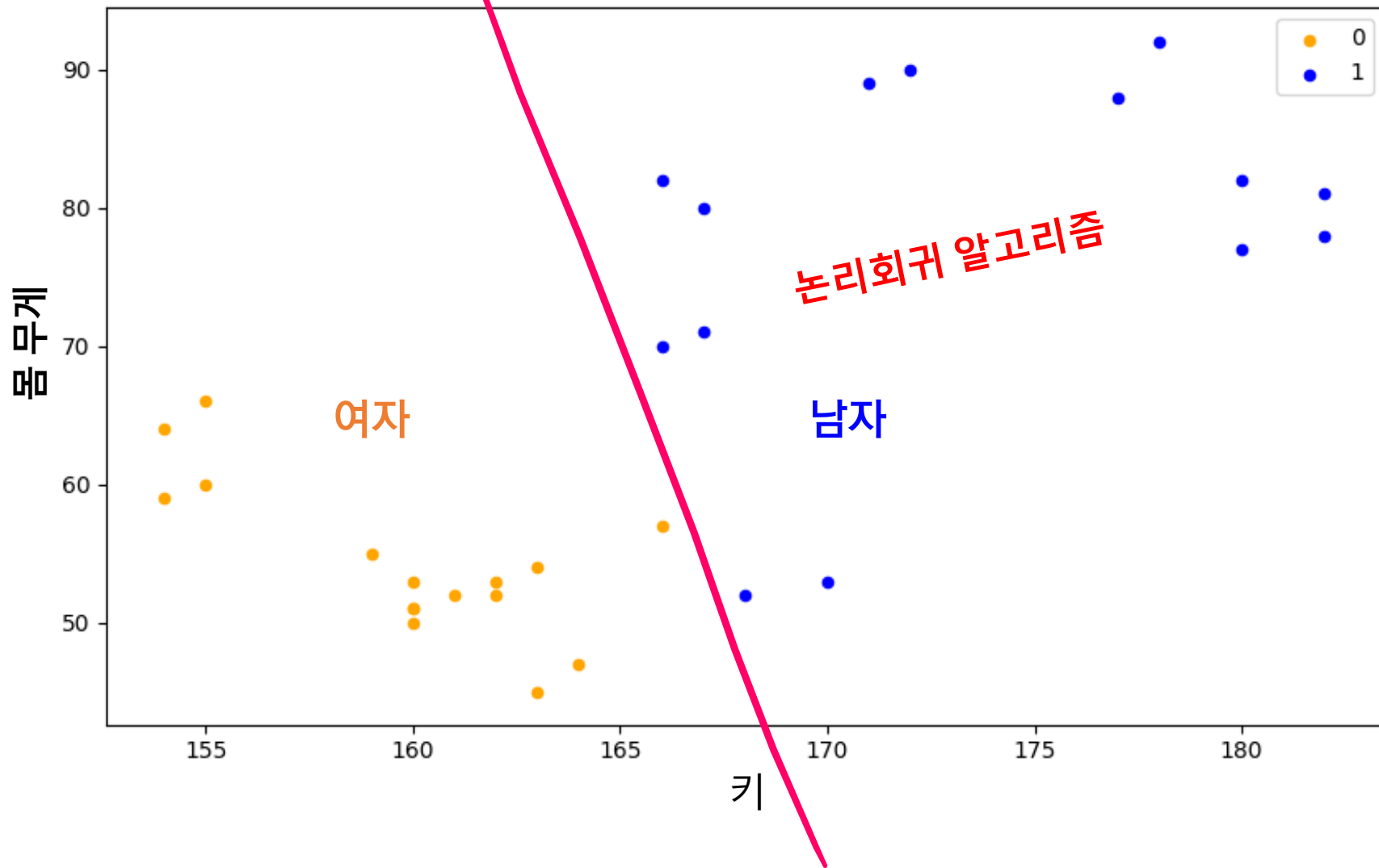






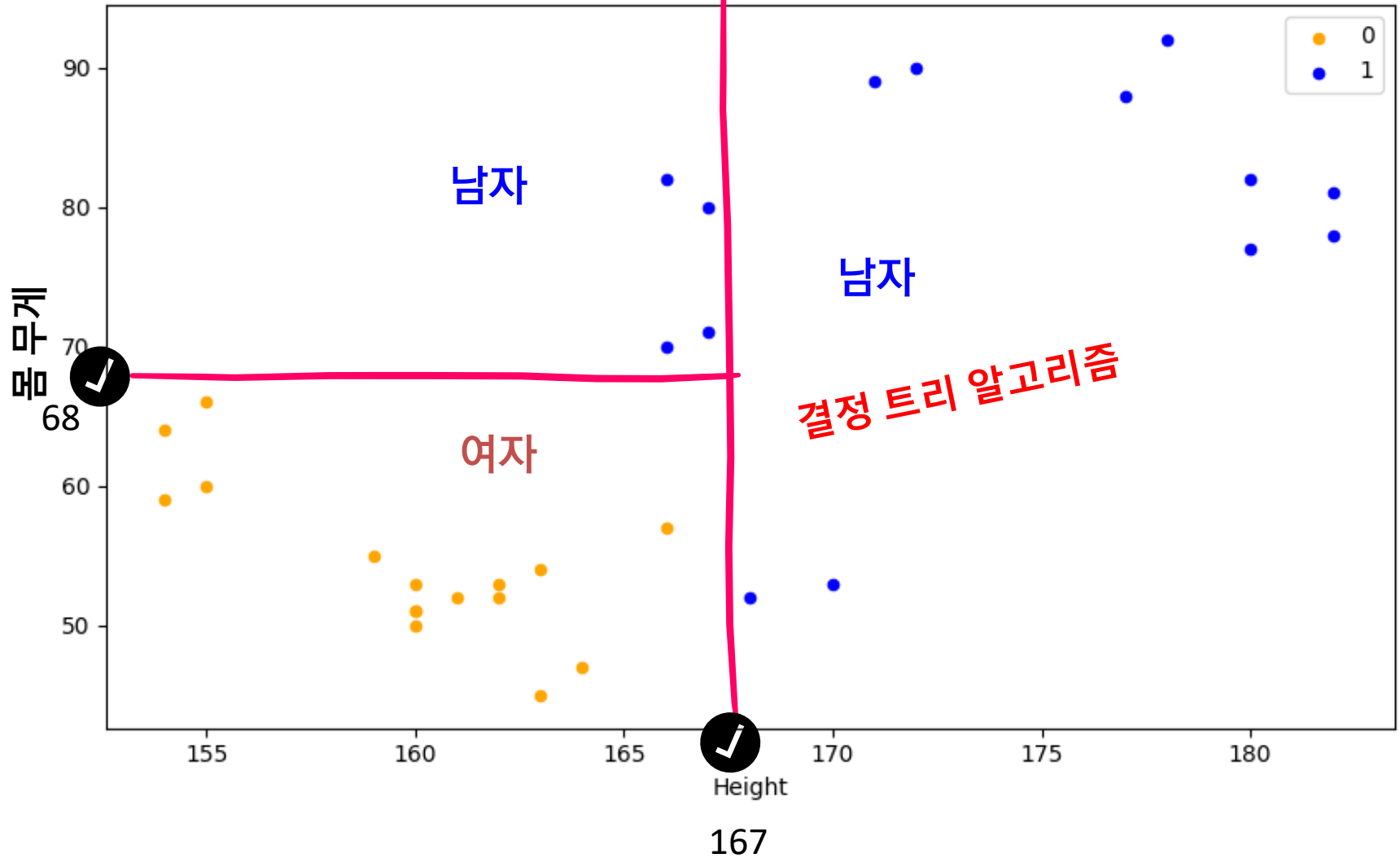
분류

Classification  
키 vs. 몸무게



# 분류

Classification  
키 vs. 몸무게







키, 몸무게, 발 크기, 학년, 성별



학습용  
문제

166,57,240,1,0  
178,92,265,1,1  
167,80,270,1,1  
168,52,245,2,1  
155,60,235,2,0  
163,45,230,2,0  
160,53,235,3,0  
180,77,260,4,1  
167,71,260,2,1  
160,51,245,2,0  
162,53,240,2,0  
180,82,280,6,1  
172,90,255,6,1  
160,51,245,5,0

정답

테스트용  
문제

155,66,245,5,0  
163,54,242,5,0  
177,88,263,5,1  
166,82,268,6,1  
170,53,247,6,1  
154,59,234,1,0  
164,47,232,1,0

정답

키, 몸무게, 발 크기, 학년, 성별

학습용  
문제

166,57,240,1,0  
178,92,265,1,1  
167,80,270,1,1  
168,52,245,2,1  
155,60,235,2,0  
163,45,230,2,0  
160,53,235,3,0  
180,77,260,4,1  
167,71,260,2,1  
160,51,245,2,0  
162,53,240,2,0  
180,82,280,6,1  
172,90,255,6,1  
160,51,245,5,0

정답

테스트용  
문제

155,66,245,5,0  
163,54,242,5,0  
177,88,263,5,1  
166,82,268,6,1  
170,53,247,6,1  
154,59,234,1,0  
164,47,232,1,0

정답

```
youngJa = svm.SVC()  
youngJa.fit('학습용문제', '정답')  
prediction=youngJa.predict('테  
스트용 문제')
```

# 분류 알고리즘

Machine Learning

- SVC (서포트벡터머신)
- DecisionTreeClassifier (결정트리)
- RandomForestClassifier (랜덤포레스트)
- XGBClassifier (XGBoost, eXtreme Gradient Boosting, Boosting or Additive Training) (부스팅)

- LogisticRegression (논리회귀)
- Multilayer Neural Networks
- CNN/RCNN/GCNN

Deep Learning



# 코드 추상화와 재사용

# 머신러닝 절차

1. CSV 파일 로드
2. 데이터 시각화와 분석
3. 데이터 전처리 (정규화, 데이터 균형)
4. 특징 상관관계 분석
5. 데이터 분할
6. ML 알고리즘을 이용한 분류/예측
7. 평가 및 결론