

# 선수 지식 - 통계

## 편향과 오차

편향과 오차 | 딥러닝의 기초가 되는 확률 개념 알아보기

강사 나동빈

# 선수 지식 - 통계

## 편향과 오차

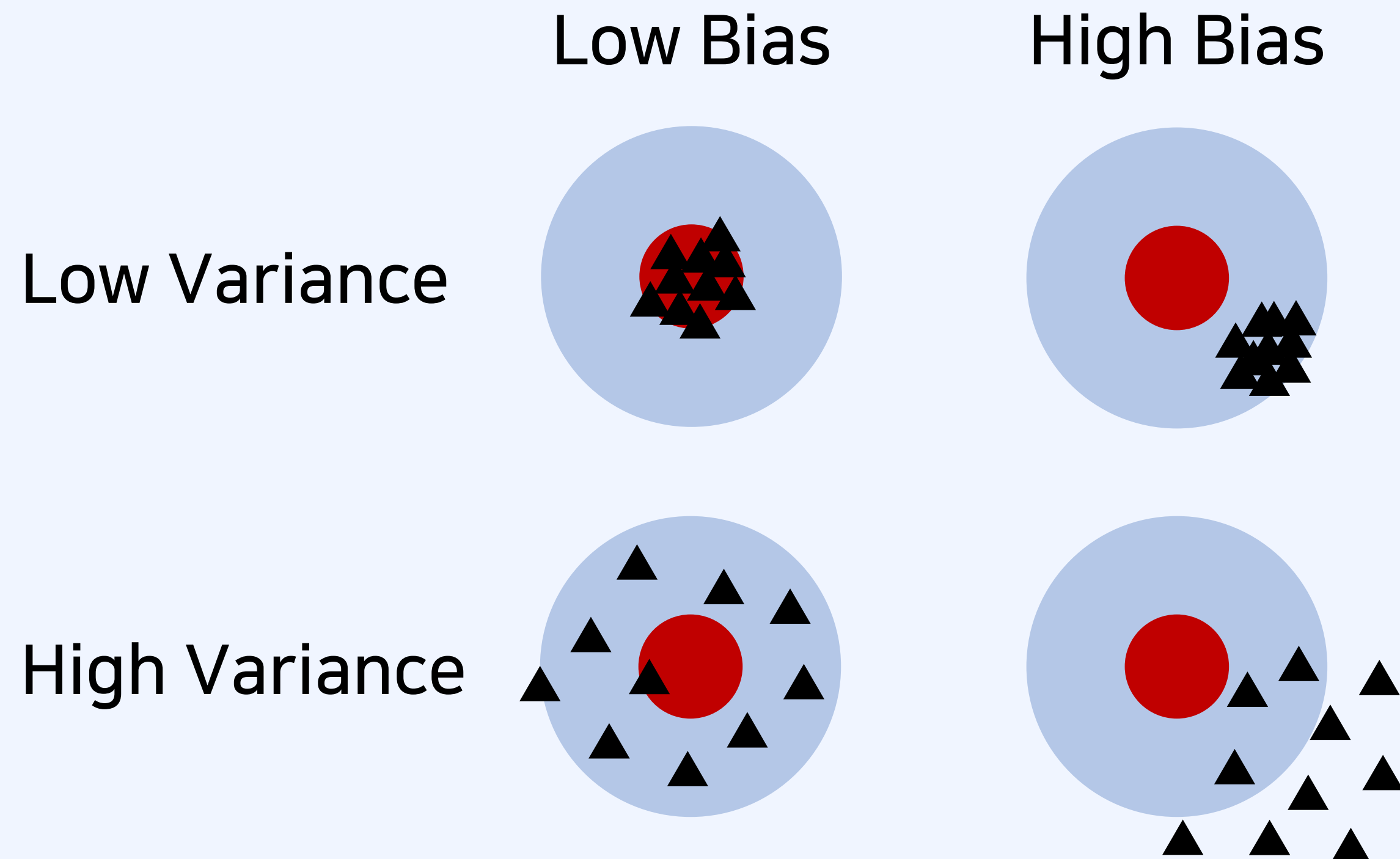
- 편향된 데이터(biased data): 실제 데이터를 반영하지 못 하고, 편향된 데이터  
**예시)** 한 유명한 해외 얼굴 데이터 세트는 대부분 서양인 얼굴로 구성되어 있다.



“해당 데이터 세트로 학습하여, 한국인을 대상으로 하는 서비스를 만든다면?”

## 편향(Bias)과 분산(Variance)

- 편향(bias)이 높을 때: 모델이 예측한 값이 정답과 멀리 떨어져 있는 경우
- 분산(variance)이 높을 때: 모델이 예측한 값이 서로 멀리 떨어져 있는 경우



## 오차(Error)

- 기계 학습 모델의 성능을 평가하기 위해 오차를 계산하는 과정이 필요하다.
  - 현재 학습 중인 모델이 얼마나 잘못되었는지 측정할 필요가 있다.
- 오차(error): 실제 정답과 우리 모델이 예측한 값의 차이
  - 오차는 내 모델이 얼마나 잘못되었는지 알려주는 수치화된 값으로 이해할 수 있다.
  - 비용(cost) 혹은 손실(loss)이라고도 부른다.

## 평균 제곱 오차(Mean Squared Error)

- 대표적인 오차 계산 함수 중 하나가 평균 제곱 오차(mean squared error)다.
- 평균 제곱 오차는 말 그대로 오차(error)를 제곱한 값의 평균을 의미한다.
- 각 데이터가 (입력  $x$ , 정답  $y$ )로 구성될 때, **MSE 공식**은 다음과 같다.

$$MSE = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2$$



## 평균 제곱 오차 계산 예시

- **나이 예측 모델** *parameter*  $\theta$ 를 고려해 보자.
- 이는  $f(x; \theta)$ 로 표현 가능하다.
- 이때 제곱 오차(squared error)는 다음과 같이 표현 가능하다.

$$(f(x; \theta) - y)^2$$

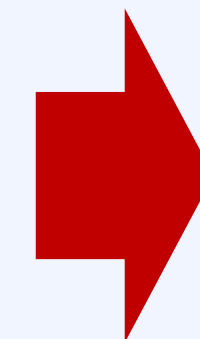
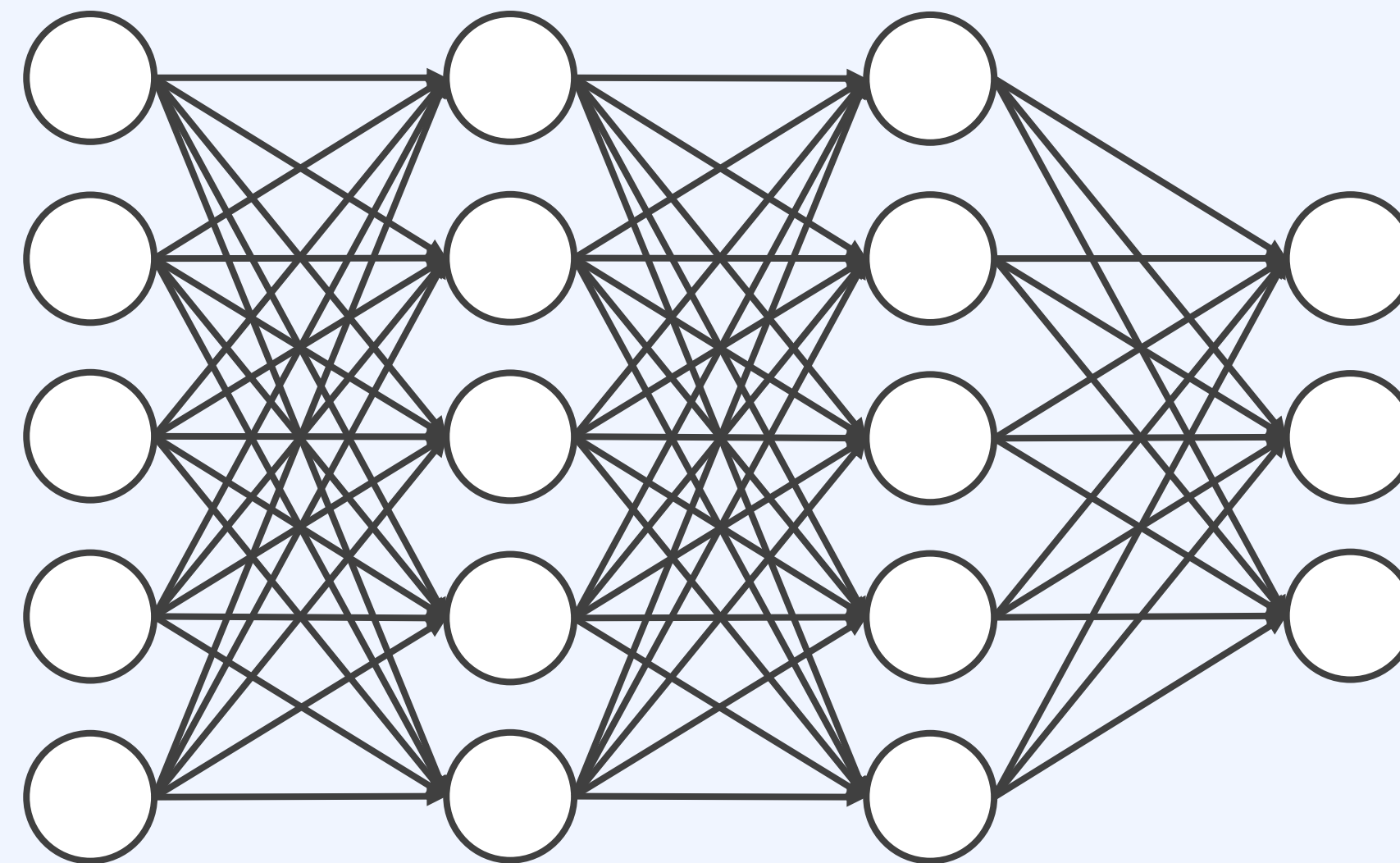
- 실제 나이가 27세( $y = 27$ )인 한 명의 사람 얼굴 이미지  $x$ 가 주어졌다고 가정하자.
- 모델의 예측 결과가 29세라면, 제곱 오차는  $(27 - 29)^2 = 4$ 이다.
- 모델의 예측 결과가 27세라면, 정확히 예측하여 제곱 오차는 0이다.

## 평균 제곱 오차 계산 예시

- 제곱 오차(squared error)는 다음과 같이 계산된다.



이미지  $x$



추론 결과  
 $f(x) = 78$



정답  
81

- 제곱 오차:  $(78 - 81)^2 = 9$