

# Feedback

## (Robust Regression, KNN)

2022.05.14

최종욱

# 1. Robust Regression

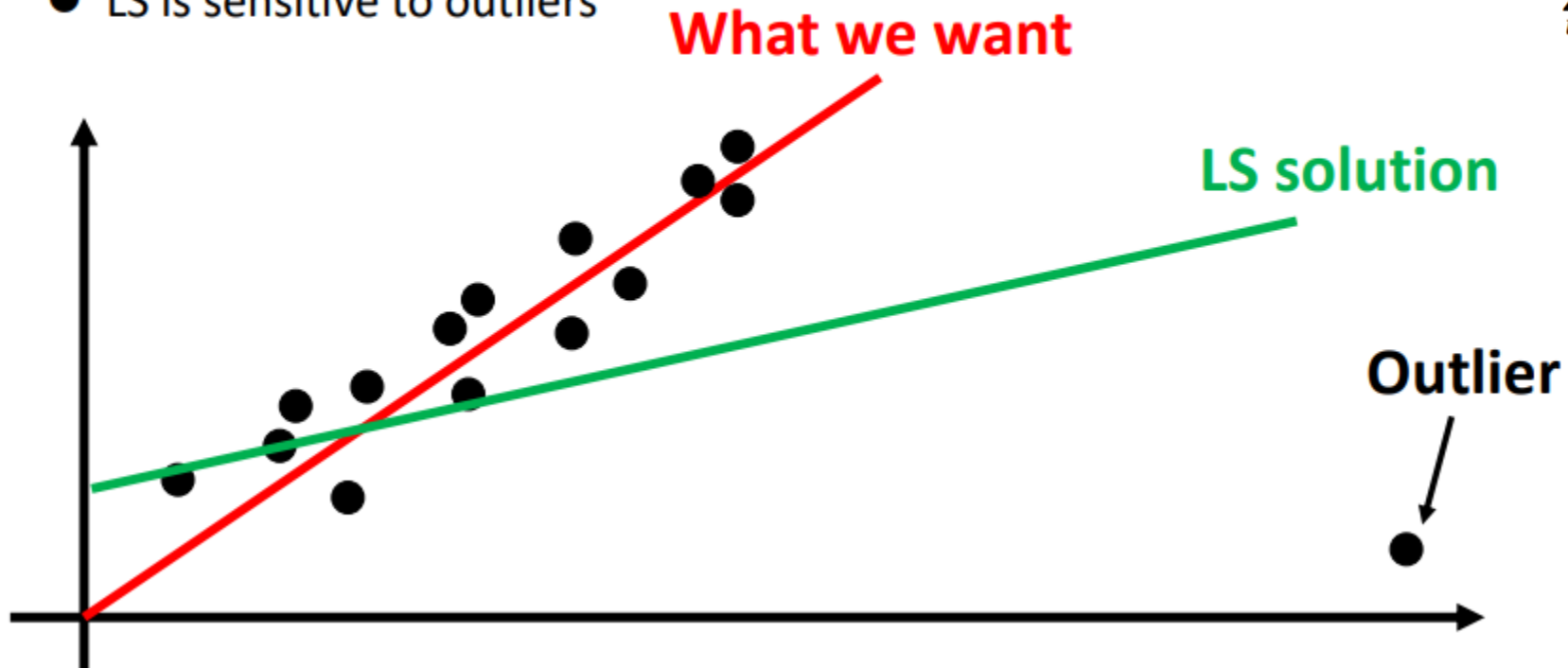
# Robust Regression

- Least Square (L2 loss) with Outliers

- Let's consider least square problem with outliers in 'y':

- LS is sensitive to outliers

$$f(w) = \sum_{i=1}^n (wx_i - y_i)^2$$



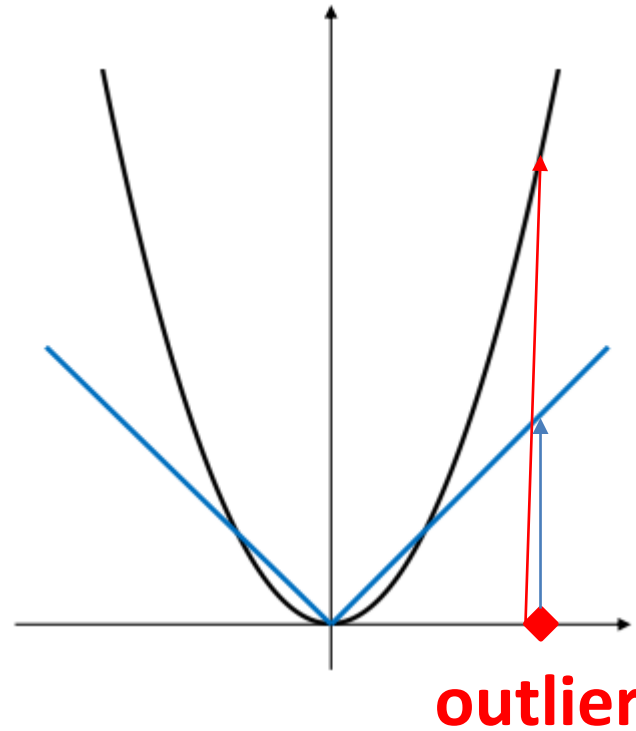
# Robust Regression

- Robust Regression (L1 loss)

$$f(w) = \sum_{i=1}^n (wx_i - y_i)^2$$



$$f(\mathbf{w}) = \sum_{i=1}^n |\mathbf{w}^T \mathbf{x}_i - y_i|$$



# Robust Regression

- Least Trimmed Squares

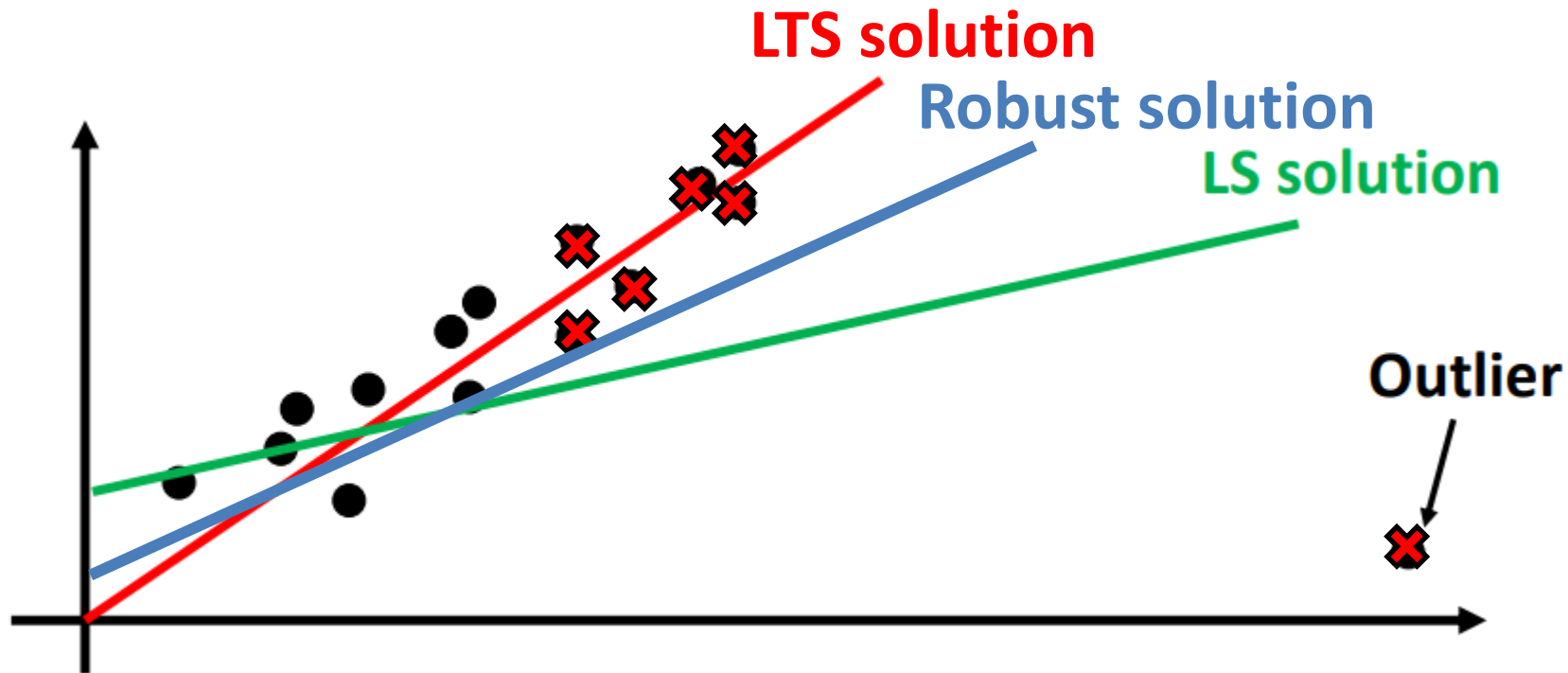
```
ltsReg(x, ...)  
  
# S3 method for formula  
ltsReg(formula, data, subset, weights, na.action,  
        model = TRUE, x.ret = FALSE, y.ret = FALSE,  
        contrasts = NULL, offset, ...)  
  
# S3 method for default  
ltsReg(x, y, intercept = TRUE, alpha = , nsamp = , adjust = ,  
        mcd = TRUE, qr.out = FALSE, yname = NULL,  
        seed = , trace = , use.correction = , wgtFUN = , control = rrcov.control(),  
        ...)
```

alpha

the percentage (roughly) of squared residuals whose sum will be minimized, by default 0.5. In general, `alpha` must be between 0.5 and 1.

# Robust Regression

- Least Square(LS) vs Robust vs Least Trimmed Squares(LTS)



# Robust Regression

---

## ■ Conclusion

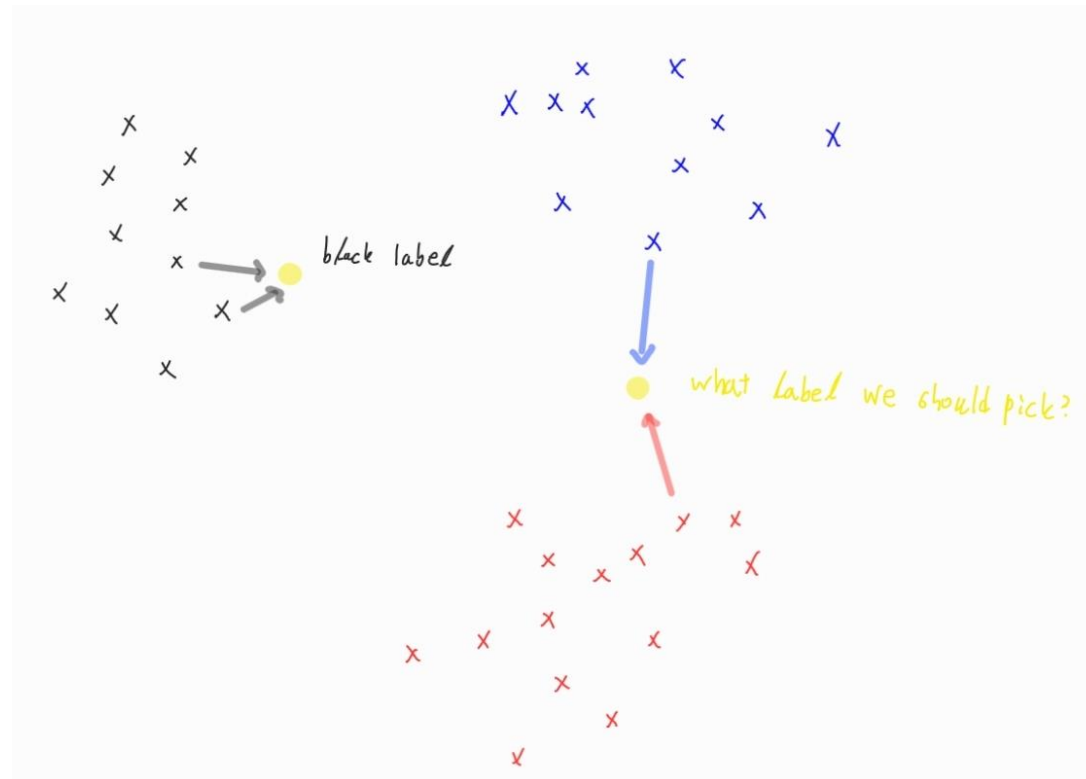
- ‘Outlier 에 의미를 부여한다’ 라는 해석을 위해서는, Outlier가 모델의 Loss(잔차) 계산에 포함되어 있어야 한다고 판단됨
- LTS 는 특정 비율만큼 Loss가 큰 데이터를 순서대로 제외하고, LS solution을 진행하기 때문에 Outlier 가 모델의 Loss 계산에 포함되지 않음
- 따라서, Outlier에 의미를 부여하며 모델의 성능을 높이기 위해선, 성능이 조금 떨어지더라도 LTS 대신 Robust 모델을 사용하는 것이 바람직해 보임

# 1.kNN



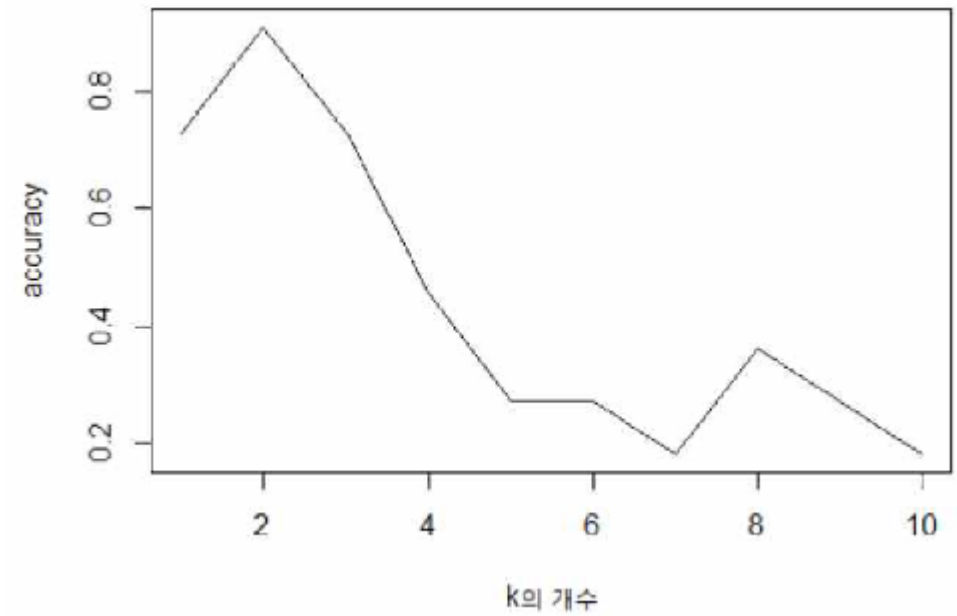
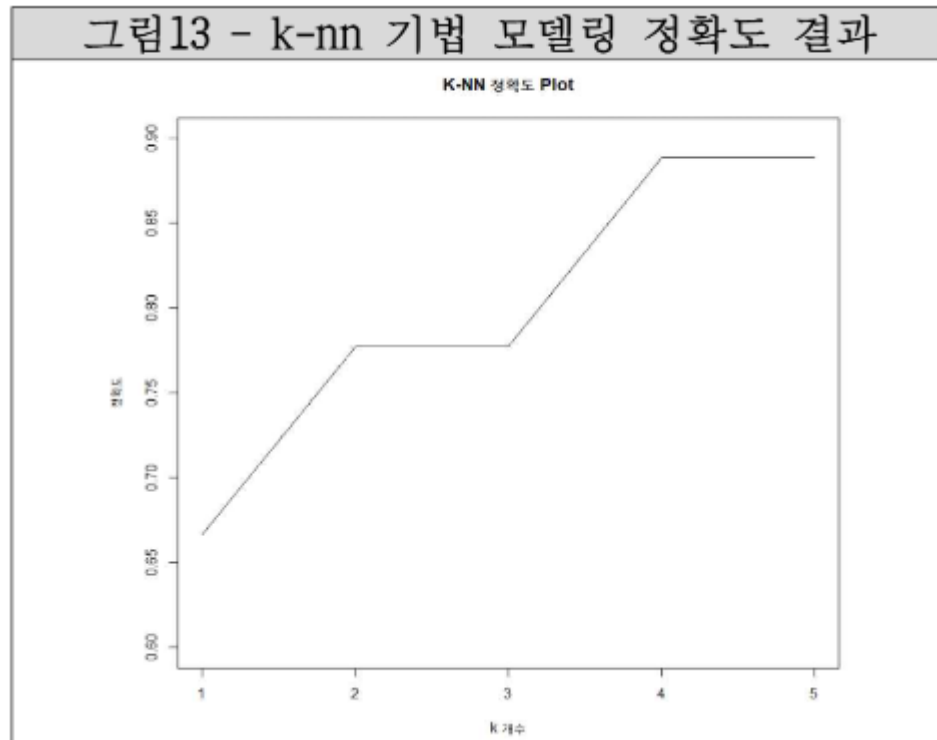
# kNN

- Why we need to pick  $k$  as odd number?
  - If  $k = 2$ ,



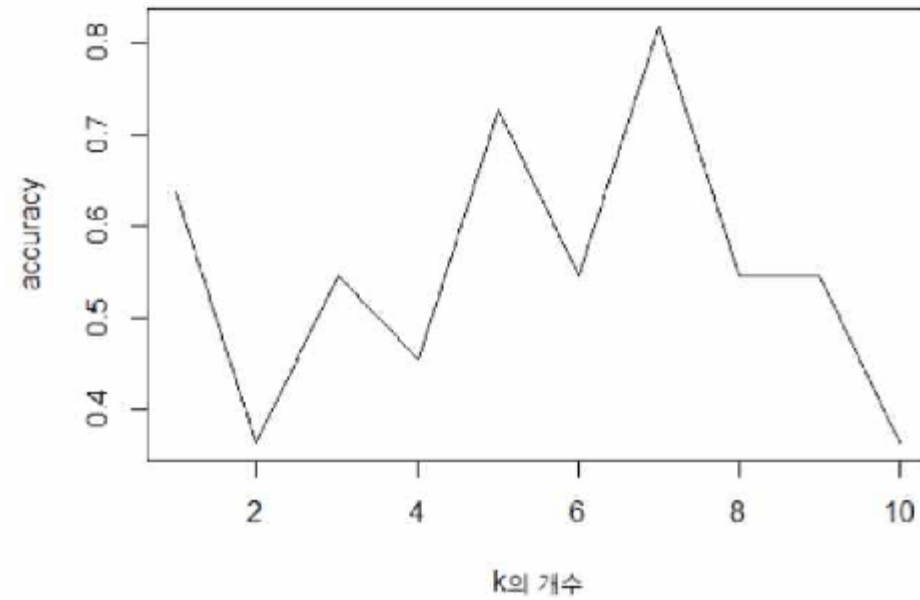
# kNN

- Bad case



# kNN

- Good case



# kNN

---

- Conclusion
  - kNN을 사용할 때,  $k$ 를 짝수로 설정하면 필연적으로 성능 하락이 야기됨
  - Bad case의 경우  $k$ 를 짝수로 선택했을 때, 오히려 성능이 상승하는 기이한 현상을 보임
  - Good case의 경우  $k$ 를 짝수로 선택했을 때 성능이 하락하고, 홀수로 선택했을 때 성능이 상승함
  - 이런 현상이 일어난 원인에 대해서는 추가적인 논의 및 분석이 필요함
    - 예상되는 원인: feature 간 공분산 문제, cross-validation 문제, 데이터 양 문제