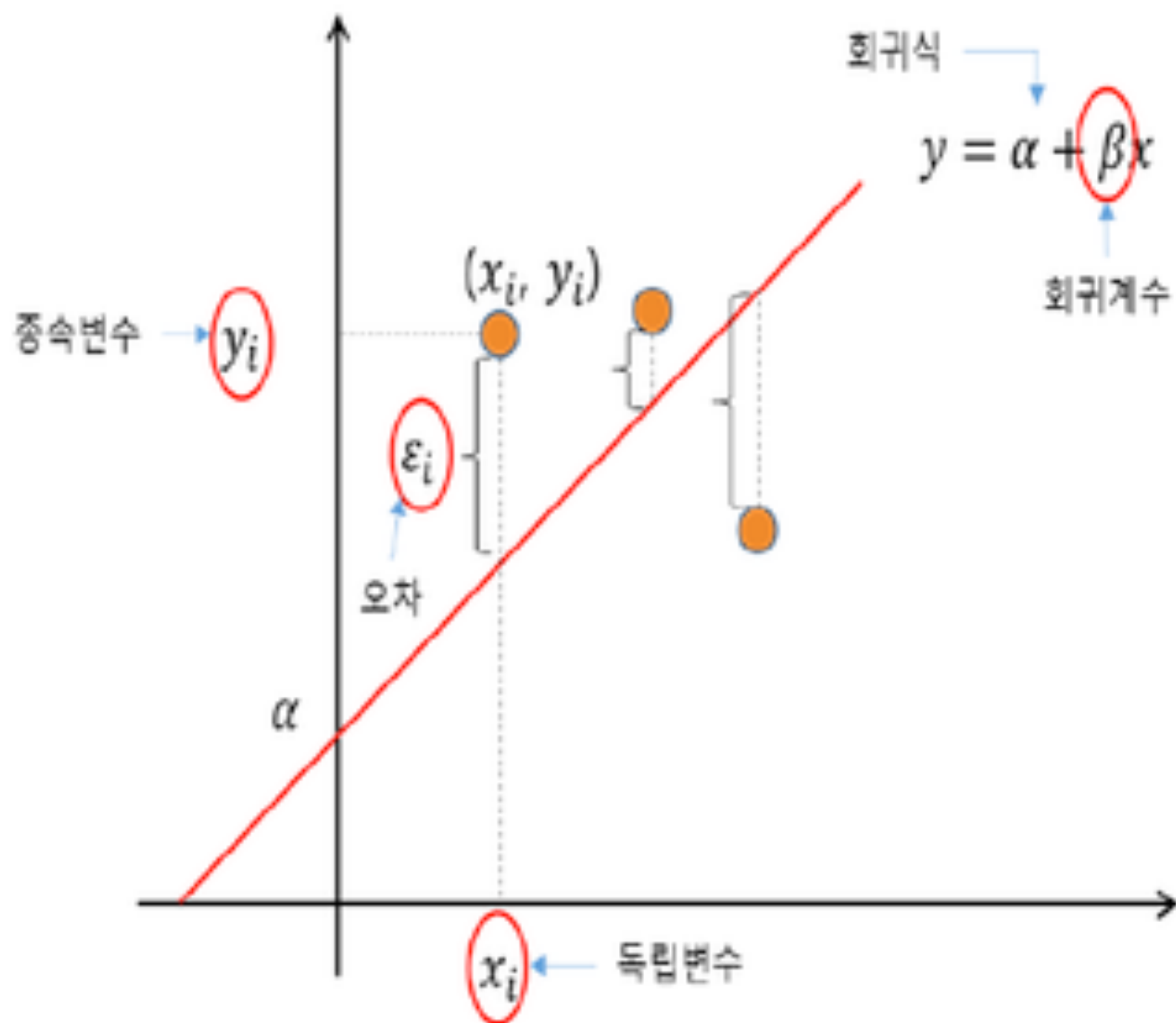


회귀 평가 지표

231008_BDA 데분중

회귀 분석 이해



회귀의 비용 함수(Cost function)

$$RSS(w_0, w_1) = \frac{1}{N} \sum_{i=1}^N (y_i - (w_0 + w_1 * x_i))^2$$

(i는 1부터 학습 데이터의 총 건수 N까지)

회귀에서 이 RSS는 비용(Cost)이며 w 변수(회귀 계수)로 구성되는 RSS를 **비용 함수**라고 합니다. 머신 러닝 회귀 알고리즘은 데이터를 계속 학습하면서 이 비용 함수가 반환하는 값(즉, 오류 값)을 지속해서 감소시키고 최종적으로는 더 이상 감소하지 않는 최소의 오류 값을 구하는 것입니다. 비용 함수를 손실함수(loss function)라고도 합니다.

경사하강법

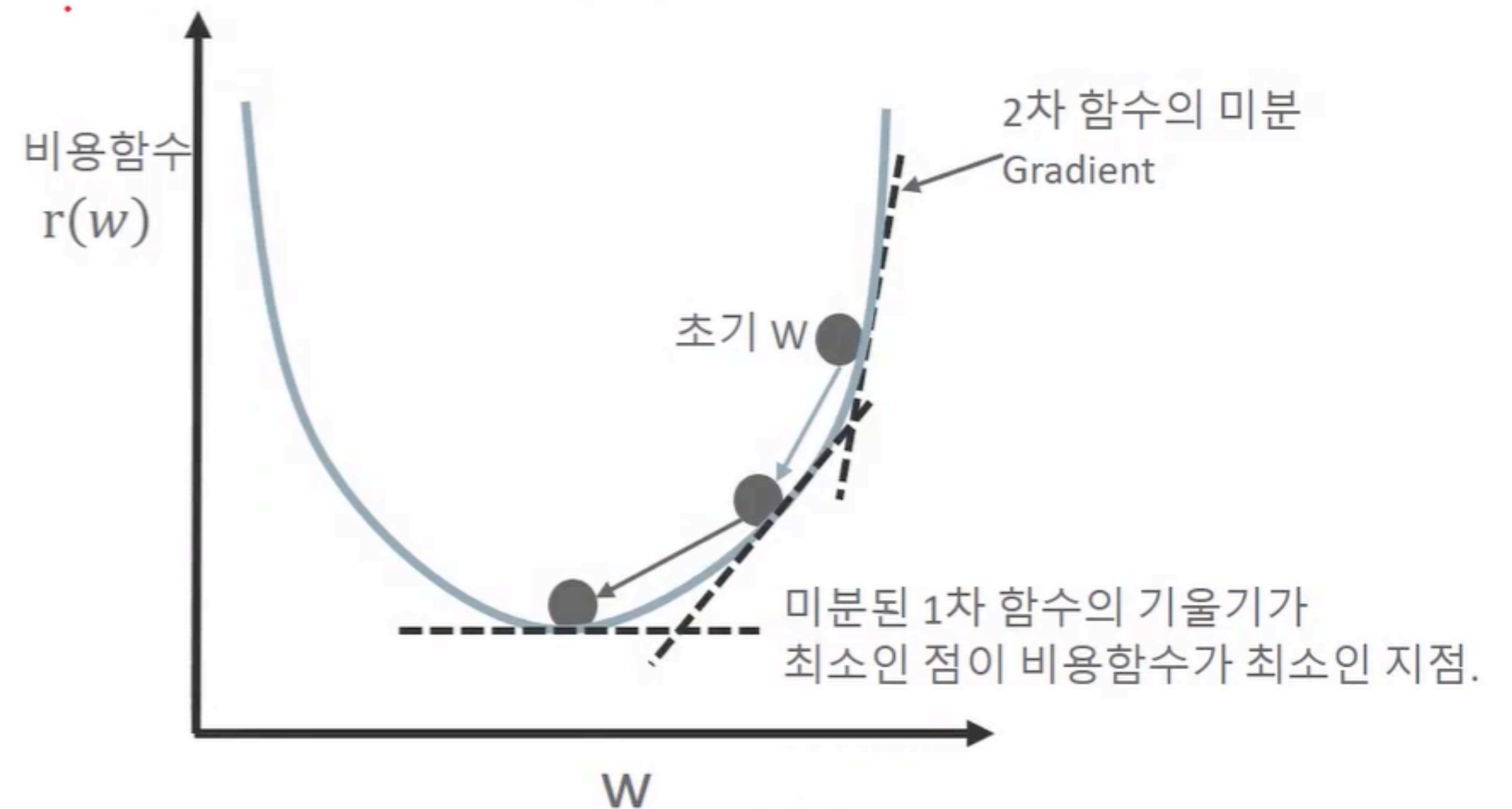
미분을 통해 비용 함수의 최소값을 찾기

어떻게 하면 오류가 작아지는 방향으로 w 값을 보정할 수 있을까?

미분은 증가 또는 감소의 방향성을 나타냅니다.

비용 함수가 다음 그림과 같은 포물선 형태의 2차 함수라면 경사 하강법은 최초 w 에서부터 미분을 적용한 뒤 이 미분 값이 계속 감소하는 방향으로 순차적으로 w 를 업데이트 합니다.

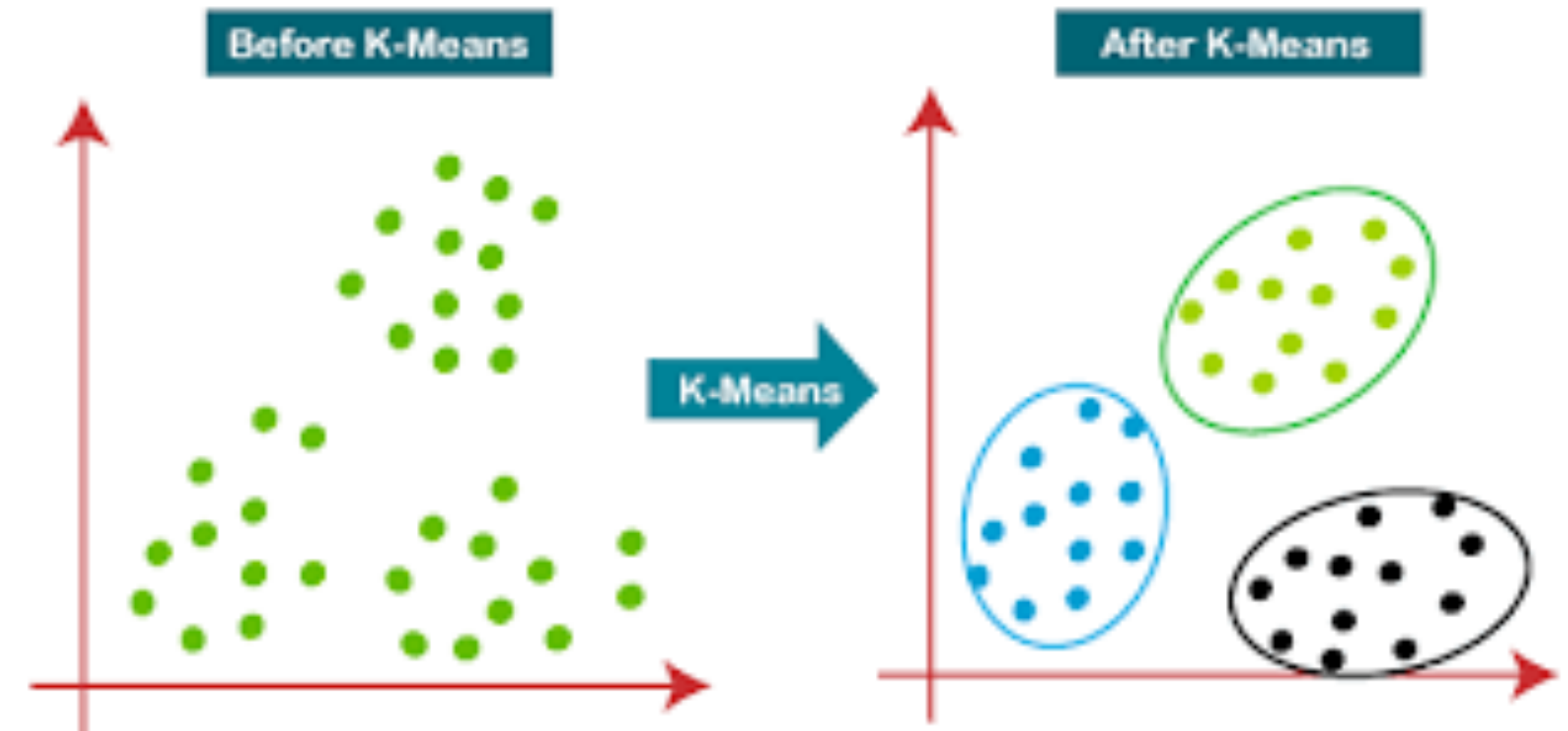
마침내 더 이상 미분된 1차 함수의 기울기가 감소하지 않는 지점을 비용 함수가 최소인 지점으로 간주하고 그때의 w 를 반환합니다



회귀 평가 지표

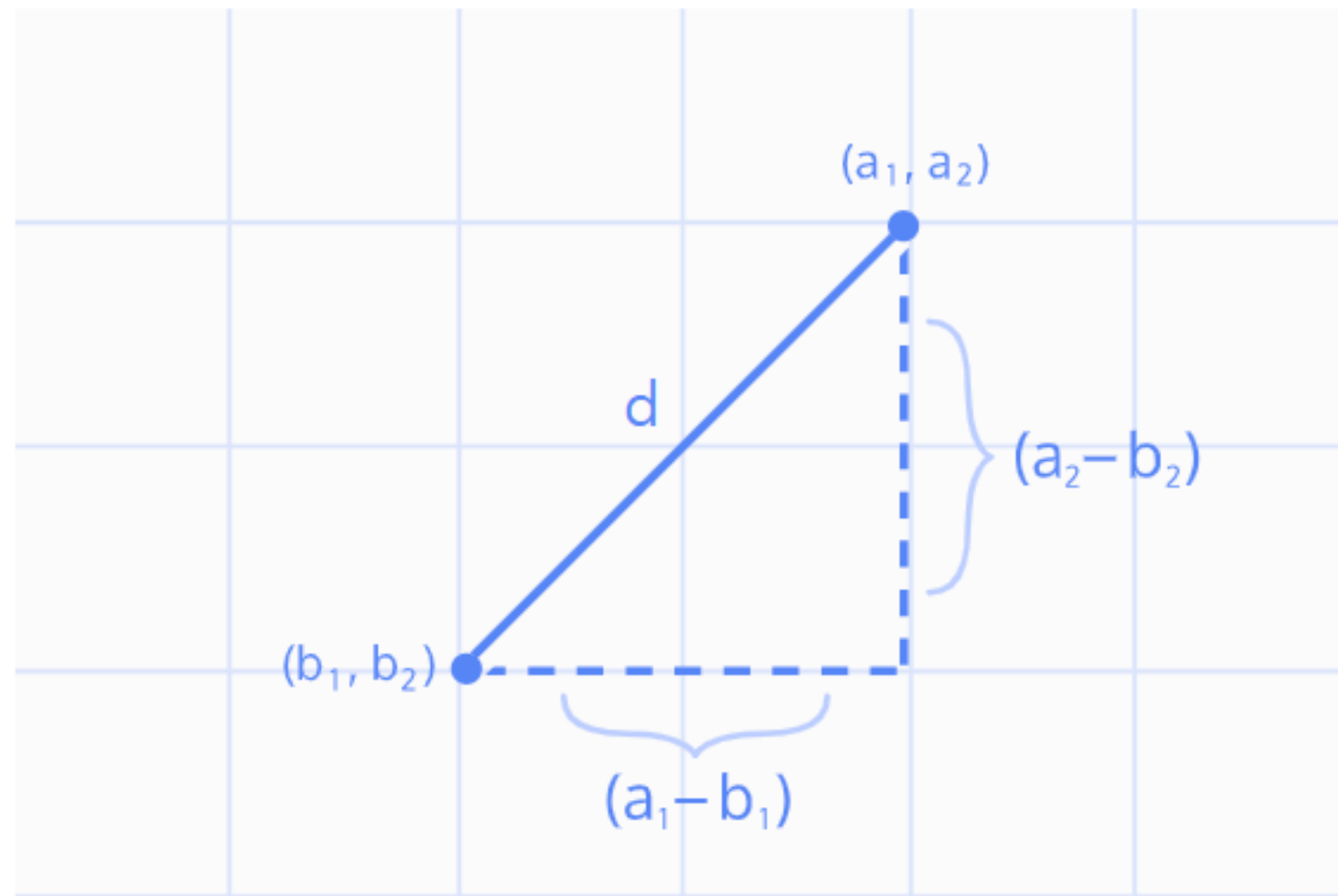
- MSE (Mean Squared Error) = $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- MAE (Mean absolute error) = $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$
- RMSE (Root Mean Squared Error) = \sqrt{MSE}
- R-squared (Coefficient of determination) = $1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$
- 참고
 - SSE(Sum of Squares Error, 관측치와 예측치 차이): $\sum_{i=1}^n (y_i - \hat{y}_i)^2$
 - SSR(Sum of Squares due to Regression, 예측치와 평균 차이): $\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2$
 - SST(Sum of Squares Total, 관측치와 평균 차이): $\sum_{i=1}^n (y_i - \bar{y}_i)^2$, SSE + SSR

KNN- 알고리즘



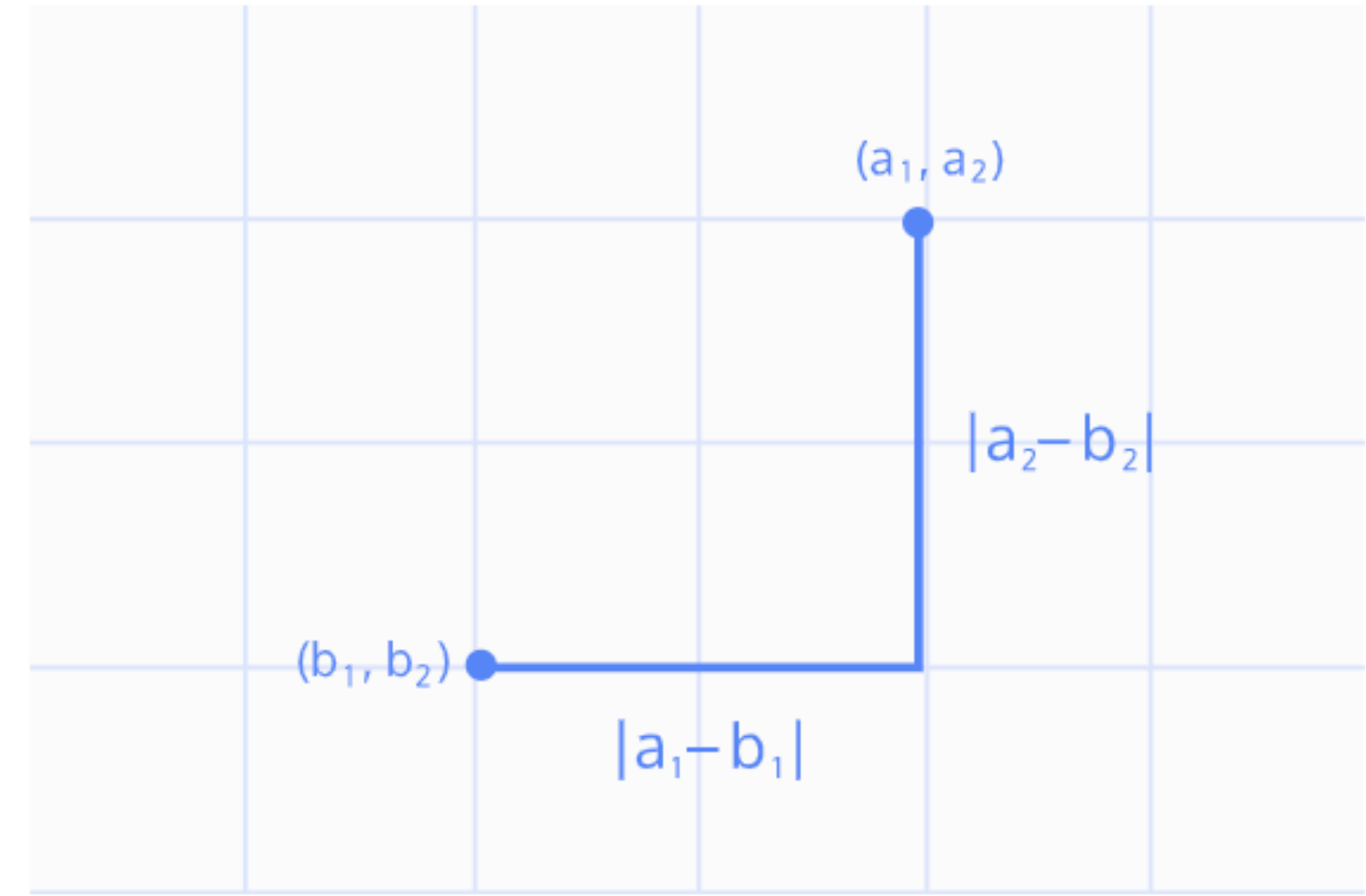
지도학습과 비지도학습

유클리드 거리 (Euclidean distance)



$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$



$$d = |a_1 - b_1| + |a_2 - b_2|$$

$$|a_1 - b_1| + |a_2 - b_2| + \dots + |a_n - b_n|$$