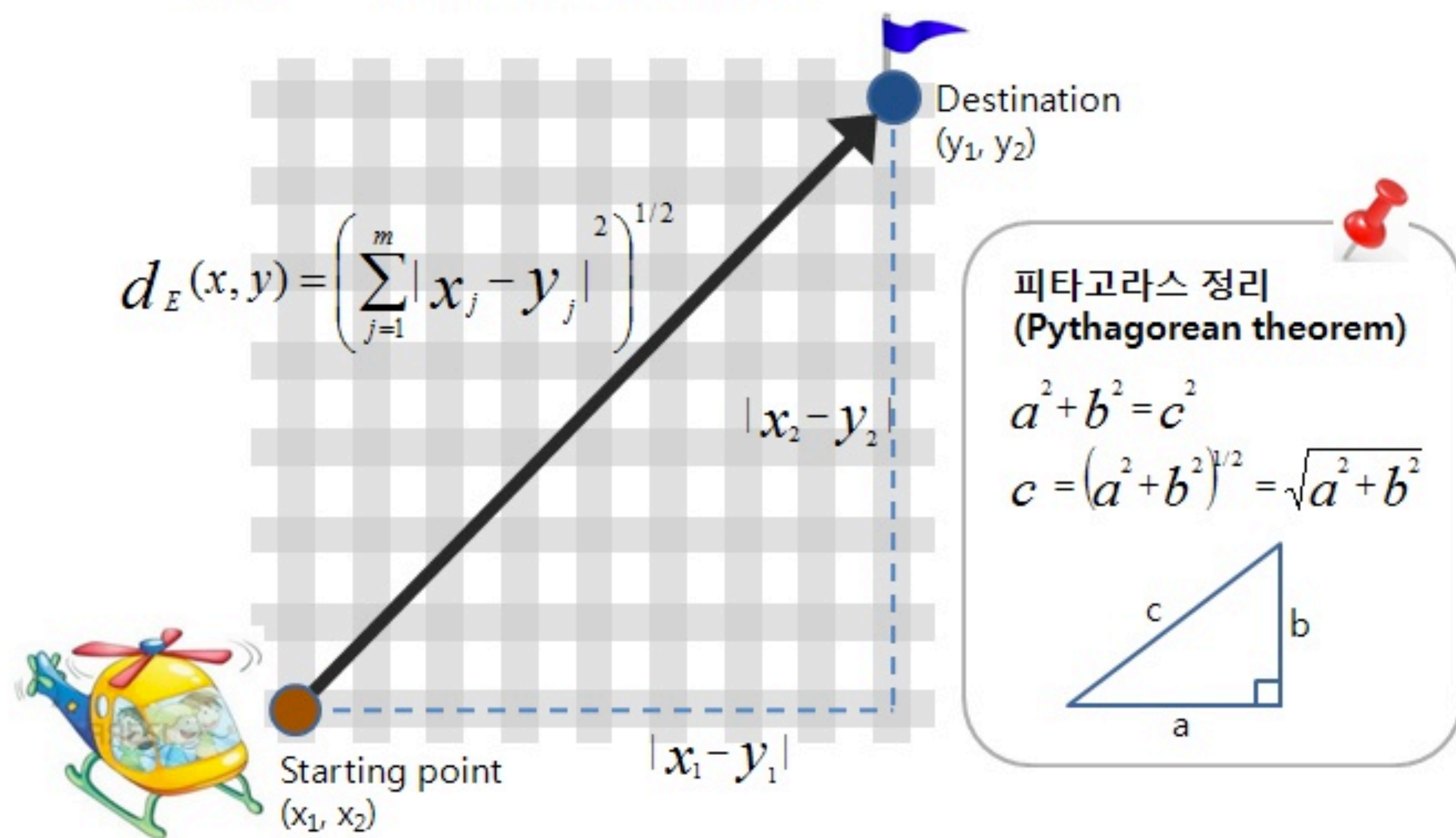


데이터 거리 측정 방법

파이썬 문법 응용반 7기

대표적인 거리 측정 방법

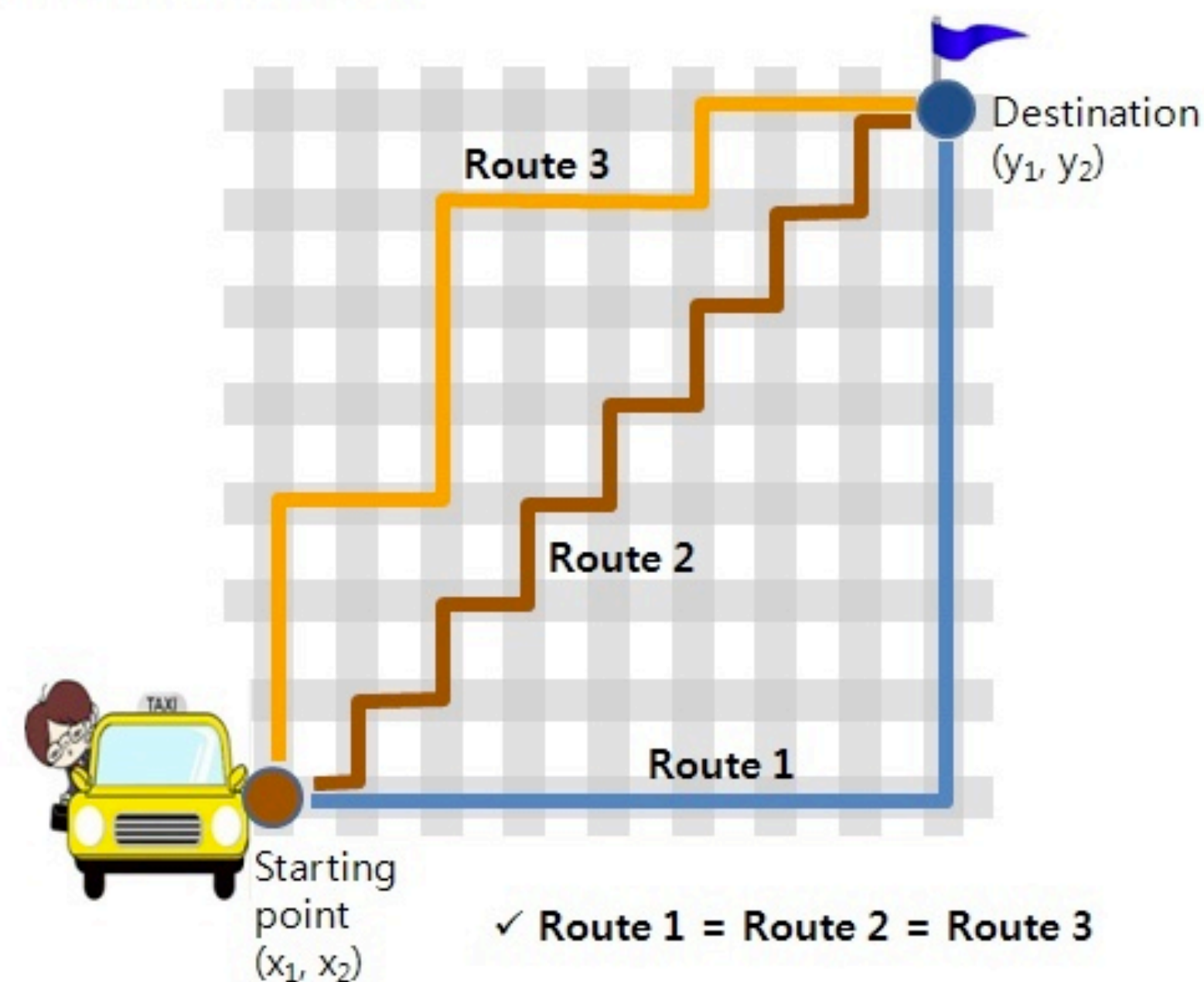
유클리드 거리 (Euclidean Distance)



[R 분석과 프로그래밍] <http://rfriend.tistory.com>

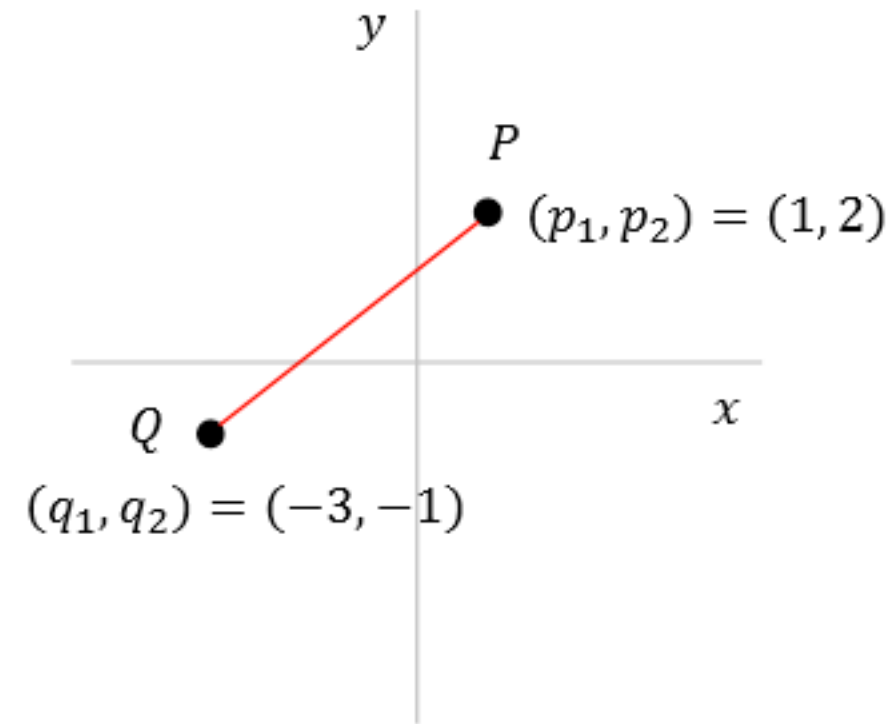
맨하탄 거리 (Manhattan Distance)

$$d_M(x, y) = \sum_{j=1}^m |x_j - y_j|$$



[R 분석과 프로그래밍] <http://rfriend.tistory.com>

대표적인 거리 측정



▪ Manhattan distance

$$d = \sum_{i=1}^n |p_i - q_i|$$

$$d = |1 - (-3)| + |2 - (-1)| \\ = 7$$

▪ Euclidean distance

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$$d = \sqrt{(1 - (-3))^2 + (2 - (-1))^2} \\ = 5$$

▪ Minkowski distance

$$d = \sqrt[m]{\sum_{i=1}^n |p_i - q_i|^m}$$

$$d = \sqrt[3]{|1 - (-3)|^3 + |2 - (-1)|^3} \\ = 4.498$$

민코프스키 거리

$$L_{\infty} = \sqrt[n]{|x_1|^{\infty} + |x_2|^{\infty} + |x_3|^{\infty} + \dots + |x_n|^{\infty}} = \max(|x_i|)$$

$$D_{chebyshev} = \max_i (|x_i - y_i|)$$

체비쇼프 거리

마할라노비스 거리 (Mahalanobis distance)

관측벡터 X 와 Y 의 마할라노비스 거리는

$$d_m(X, Y) = \left[(X - Y)^T S^{-1} (X - Y) \right]^{1/2}$$

$S = \{s_{ik}\}$: 관측벡터의 공분산행렬 (covariance matrix)

[R 분석과 프로그래밍] <http://rfriend.tistory.com>

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$