

데이터 과학

L11: Principal Component Analysis

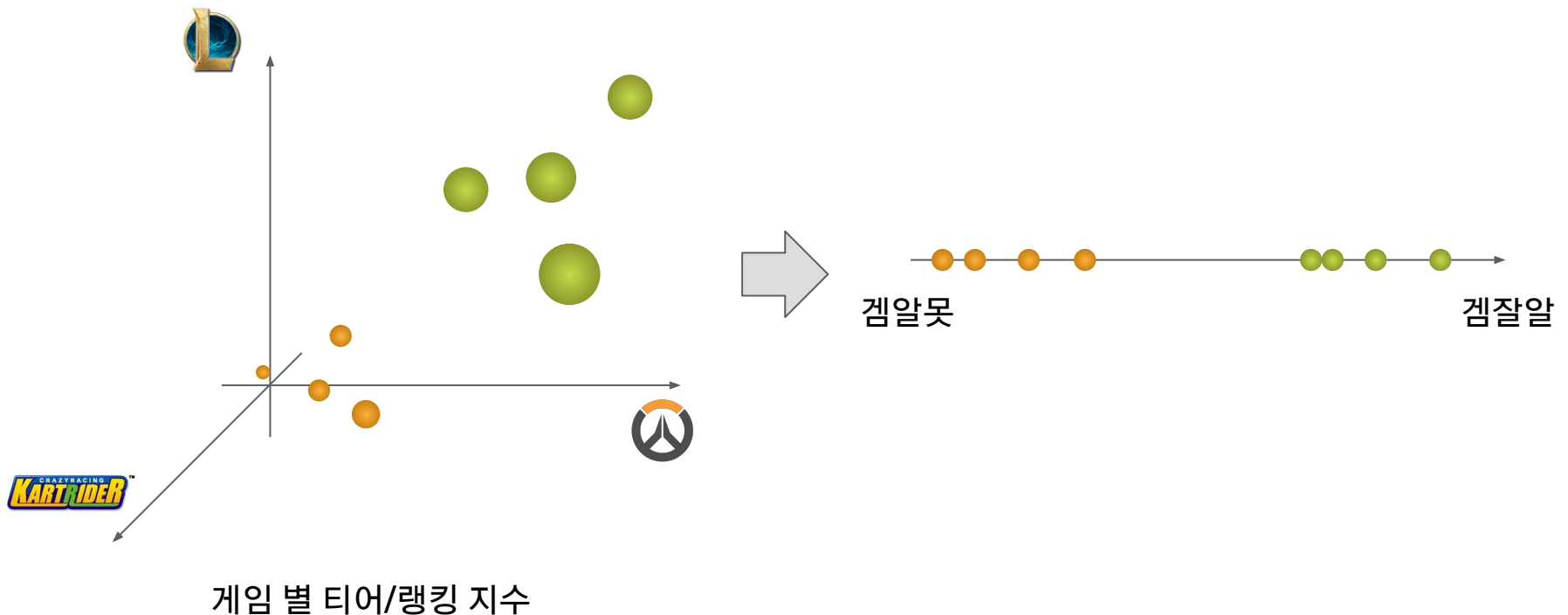
Kookmin University

Principal Component Analysis

- 주 성분 분석

- 데이터의 분포를 결정하는 핵심 성분 찾기

- 예) 원래 데이터: 게임별 티어 → 주 성분: 게임DNA

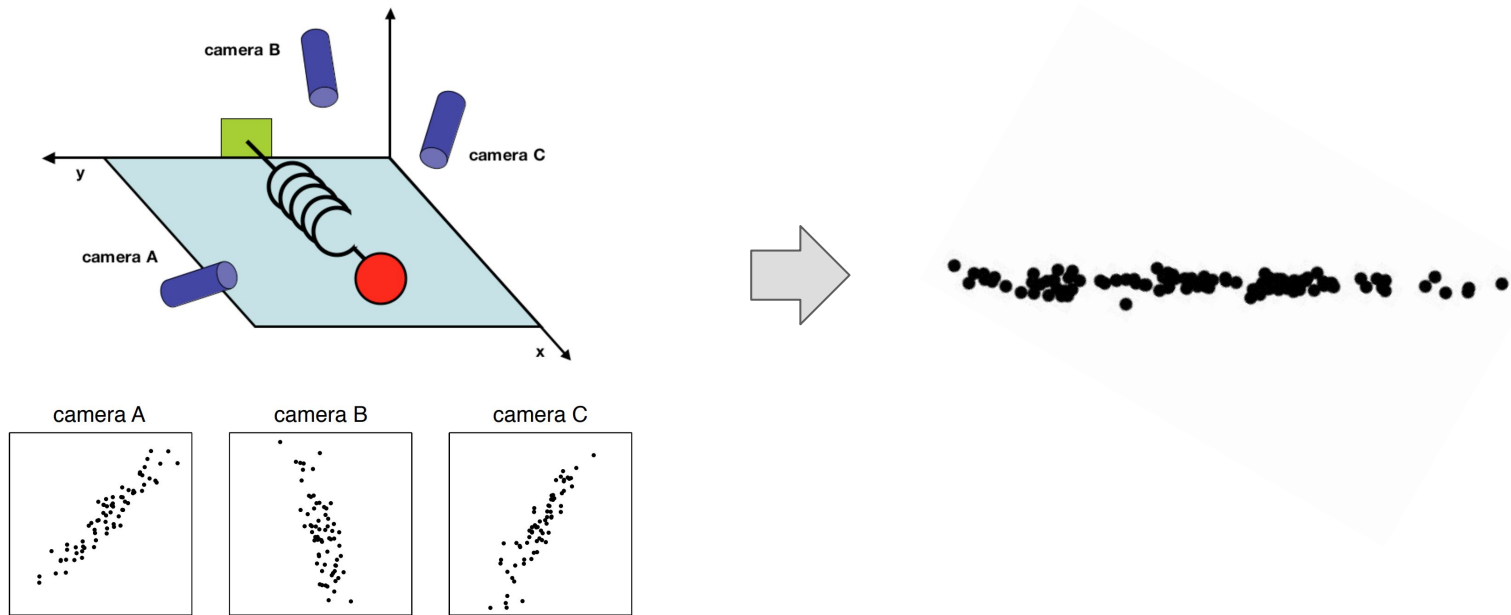


Principal Component Analysis

- 주 성분 분석

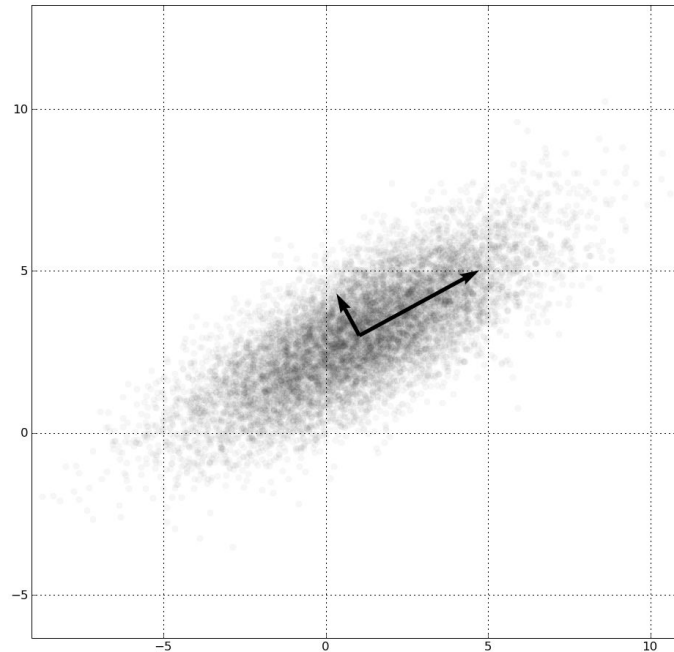
- 데이터의 분포를 결정하는 핵심 성분 찾기

- 예) 원래 데이터: 게임별 티어 → 주 성분: 게임DNA
 - 예) 원래 데이터: 카메라별 공의 위치 → 주 성분: 스프링의 힘



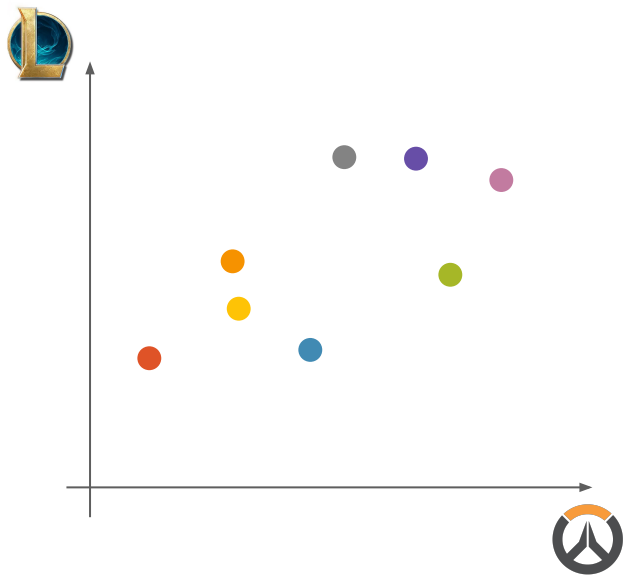
Principal Component Analysis











- 주 성분 분석
 - 분산을 최대화 하면서 서로 직교하는 새로운 축을 찾음



차원 축소

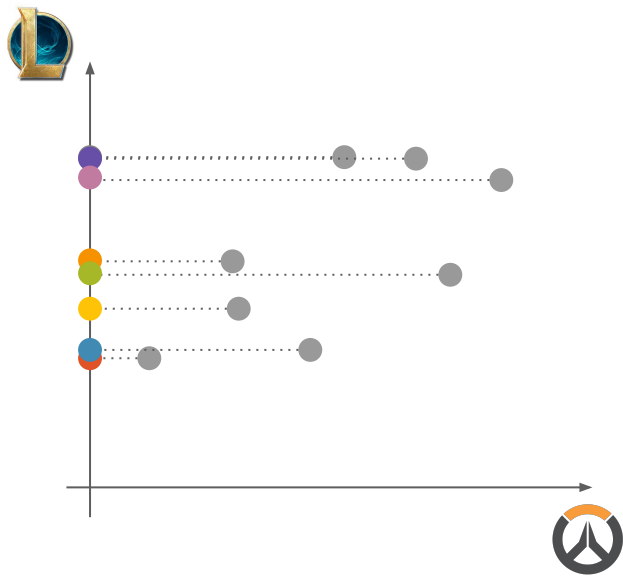
- 차원 축소 방법





								
	A	B	C	D	E	F	G	H
	3.1	3.4	4.6	3.2	7.9	7.8	4.4	7.5
	1.0	4.2	4.0	5.7	6.2	8.1	9.3	9.9

차원 축소

- 차원 축소 방법
 - 방법1. 아무 차원이나 지운다.



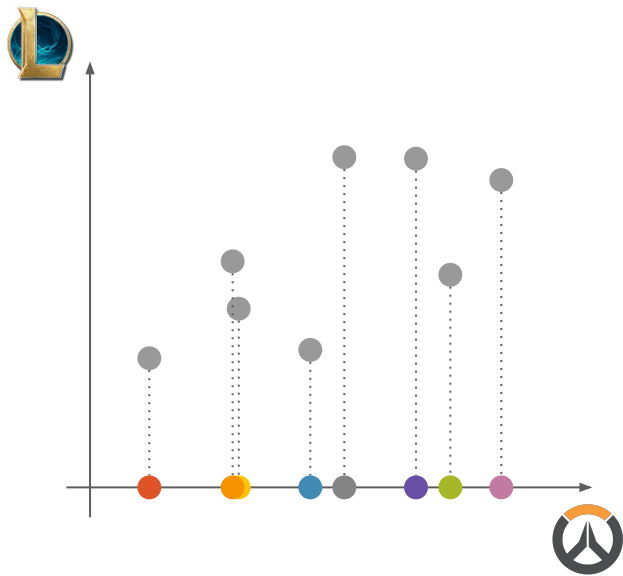
	A	B	C	D	E	F	G	H
	3.1	3.4	4.6	3.2	7.9	7.8	4.4	7.5
	1.0	4.2	4.0	5.7	6.2	8.1	9.3	9.9



차원 축소

- 차원 축소 방법

- 방법1. 아무 차원이나 지운다.

- 어떤 차원을 지우는 것이 더 좋은가?



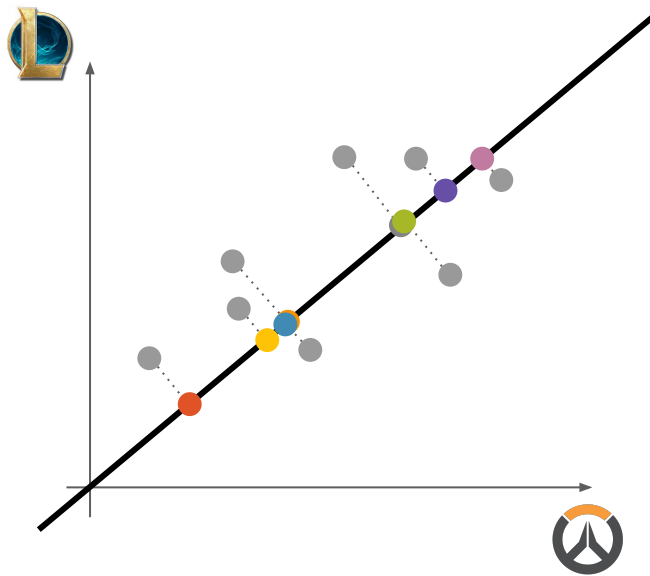
	A	B	C	D	E	F	G	H
	3.1	3.4	4.6	3.2	7.9	7.8	4.4	7.5
	1.0	4.2	4.0	5.7	6.2	8.1	9.3	9.9



차원 축소

- 차원 축소 방법

- 방법2. 새로운 축(선분)을 찾는다. = 주 성분 찾기

- 분산을 최대로..!
 - 어떻게 찾지...?

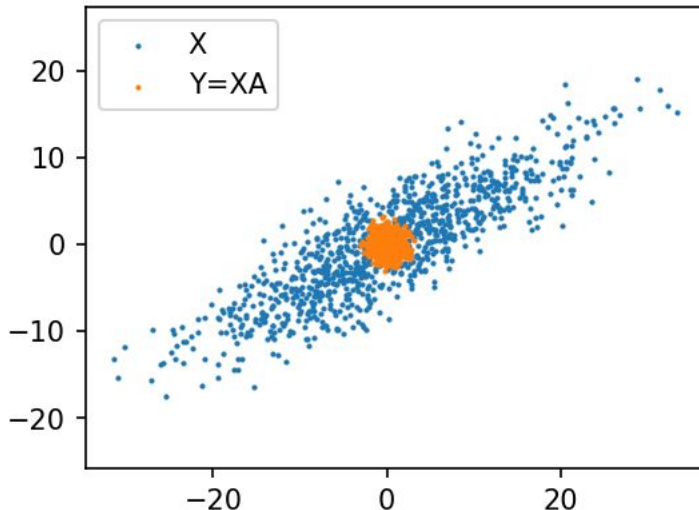


	A	B	C	D	E	F	G	H
	3.1	3.4	4.6	3.2	7.9	7.8	4.4	7.5
	1.0	4.2	4.0	5.7	6.2	8.1	9.3	9.9
?	3.3	5.4	6.1	6.5	10.0	11.2	10.3	12.4

<http://i.imgur.com/Uv2dlsH.gif>

주성분 찾기

- 표준 데이터 X
 - 각 차원의 평균 = 0, 분산 = 1, 차원간 공분산 = 0
 - $d = X$ 의 차원
- $A = d \times d$ 대칭 행렬
- $Y = XA$



$Y =$

	
3.1	1.0
3.4	4.2
4.6	4.0
3.2	5.7
7.9	6.2
7.8	8.1
4.4	9.3
7.5	9.9

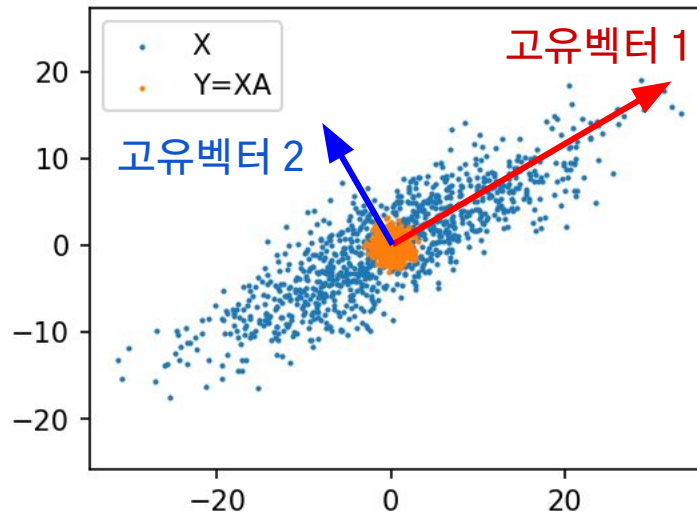
$$A = \begin{bmatrix} 10 & 4 \\ 4 & 5 \end{bmatrix}$$

주성분 찾기

- $(Y = XA)$ 의 주성분은 A 의 고유벡터이다!
 - 행렬 A 에 대해 다음 수식을 만족하는 벡터 v 를 고유벡터라 함: $Av = \lambda v$ (단, λ 는 임의의 상수=고윳값)

질문 1. Y 에 대한 행렬 A 를 어떻게 구하지?

질문 2. 행렬 A 의 고유벡터를 어떻게 구하지?



주성분 찾기

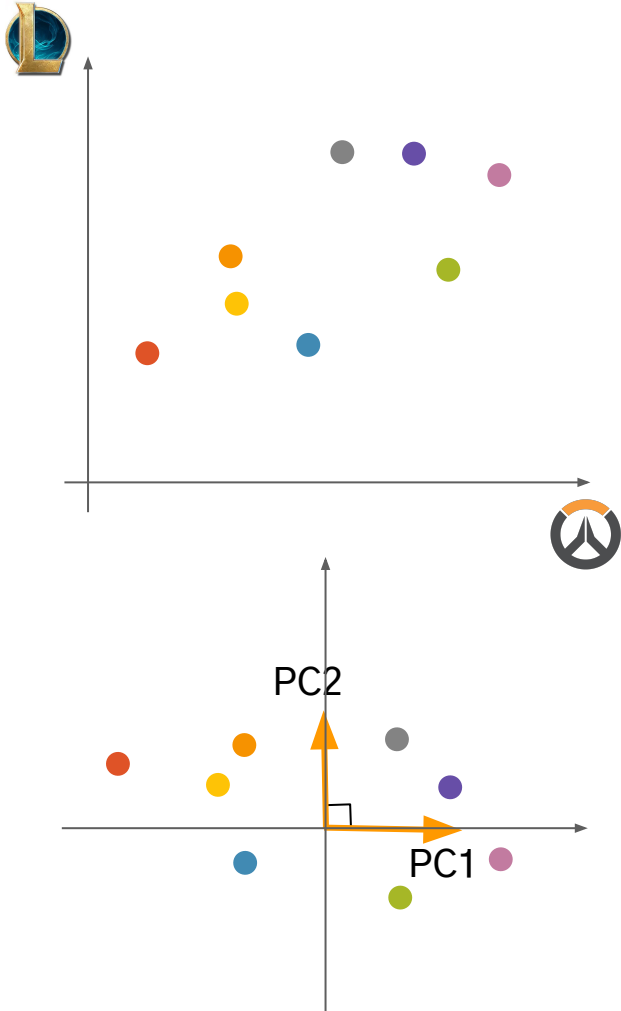
● 질문 1. Y 에 대한 행렬 A 를 어떻게 구하지?











- A 를 구하기 어렵기 때문에, Y 의 공분산 행렬(covariance matrix) Σ 를 활용! $\rightarrow \Sigma = A^2$ (증명?)
 - $\Sigma_{ij} = Y$ 의 i 번째 차원과 j 번째 차원의 공분산
 - $\Sigma = (Y^T Y)/n$ (단, $Y^T = Y$ 의 전치행렬, $n = Y$ 의 행 수)

● 질문 2. 행렬 A 의 고유벡터를 어떻게 구하지?

- A 의 고유벡터 = Σ 의 고유벡터 (증명?)
- Σ 를 고윳값 분해 (eigen decomposition)
 - Power method를 반복
 - 기타 다양한 고윳값 분해 Solver 활용

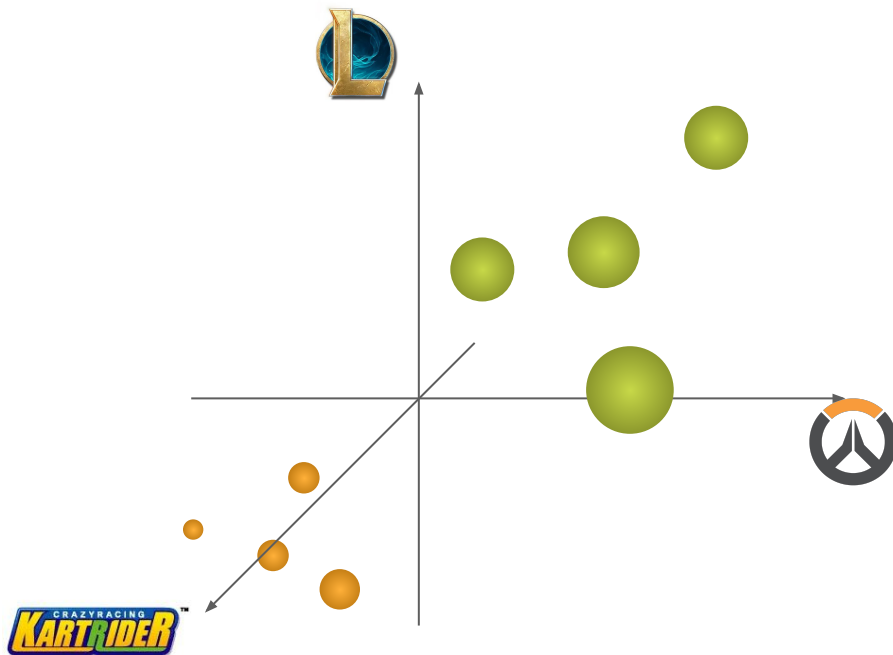
주성분으로 데이터 표현



	 A	 B	 C	 D	 E	 F	 G	 H
	3.1	3.4	4.6	3.2	7.9	7.8	4.4	7.5
	1.0	4.2	4.0	5.7	6.2	8.1	9.3	9.9

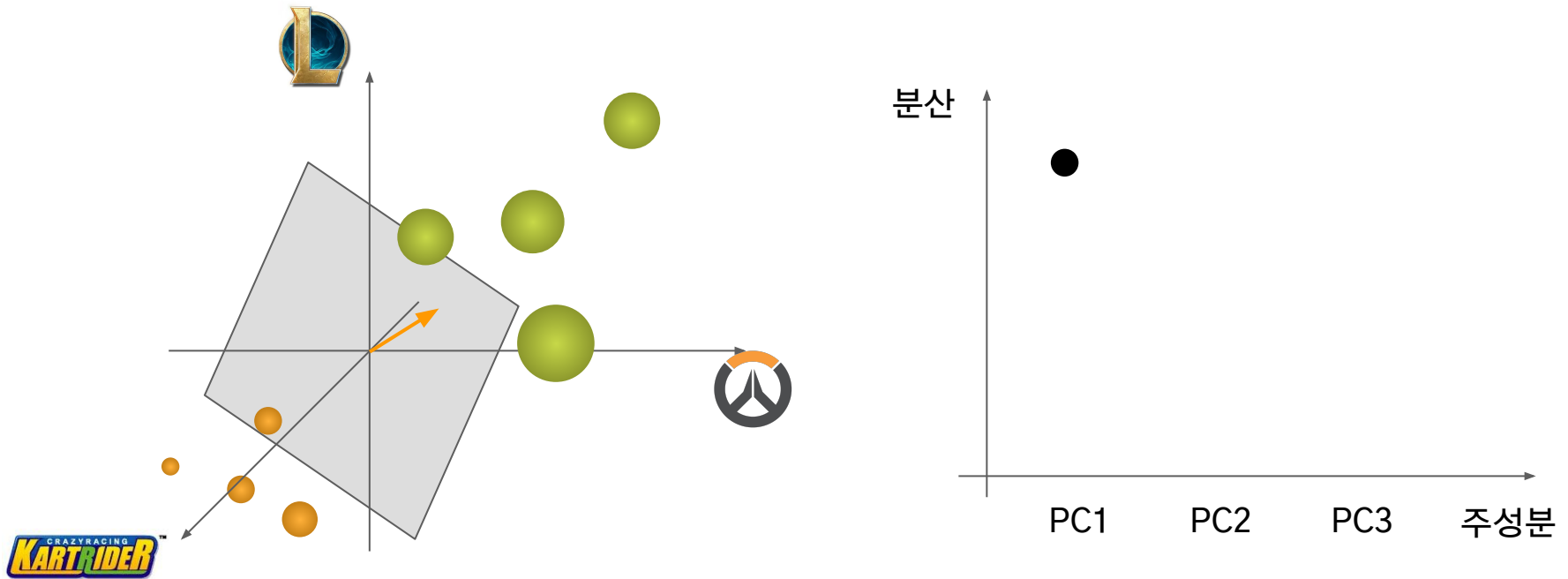
	A	B	C	D	E	F	G	H
PC1	-4.9	-2.8	-2.1	-1.6	1.9	3.1	2.1	4.3
PC2	1.1	0.7	1.5	-0.3	1.7	0.3	-0.9	-0.3

3차원 예시



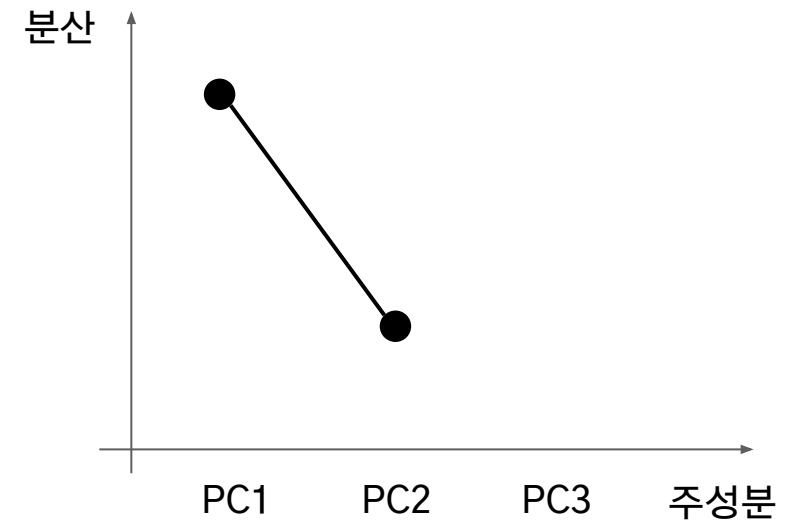
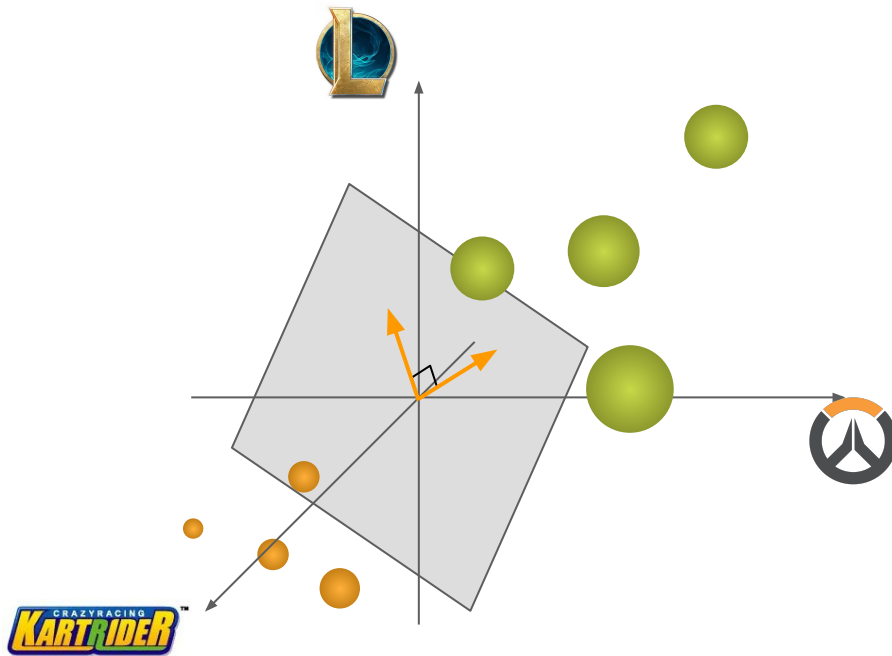
3차원 예시

- PC1 찾기: 사영했을 때 분산이 가장 커지는 벡터



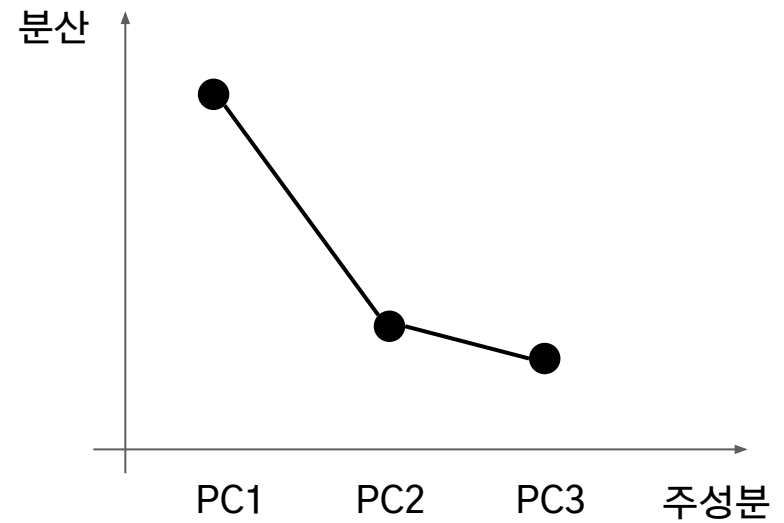
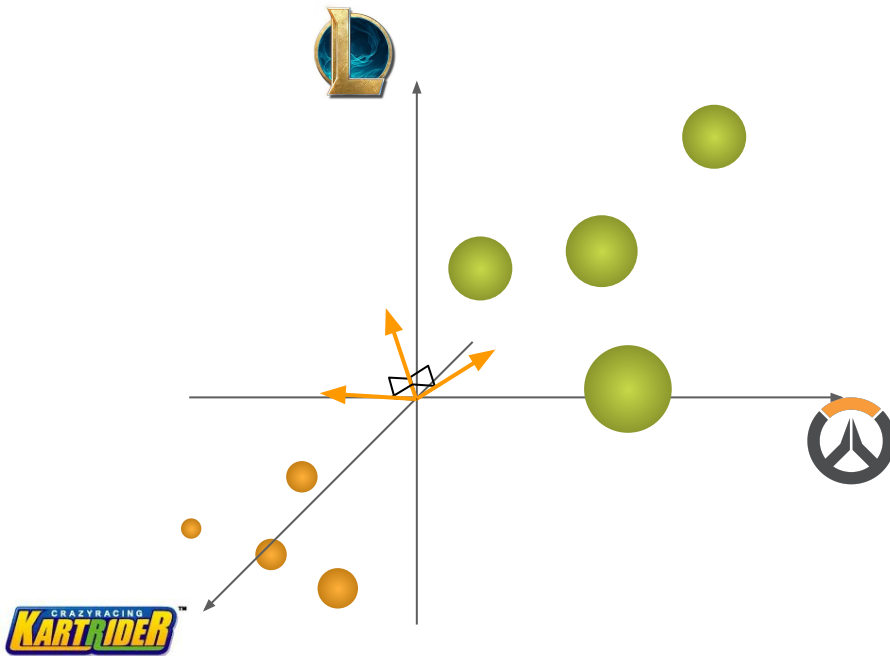
3차원 예시

- PC1의 직교평면에서 PC2 찾기



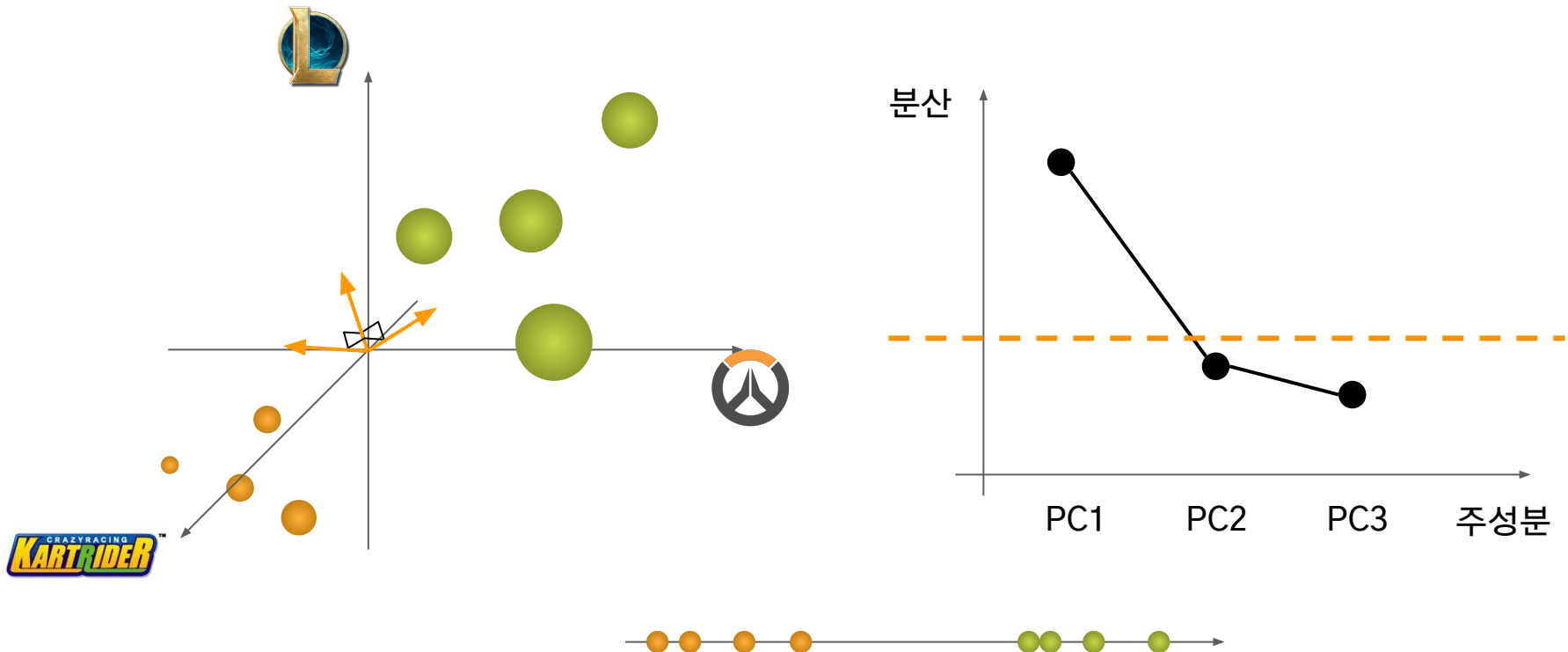
3차원 예시

- PC1과 PC2에 모두 직교하는 벡터 = PC3



3차원 예시

- PC1과 PC2에 모두 직교하는 벡터 = PC3



Questions?