### kmeans

유클리디안 거리를 기반으로 입력 레코드를 K개의 클러스터로 분류합니다.

#### 문법

kmeans [k=INT] [iter=INT] FIELD, ...

필수 매개변수

**FIELD, ...**

계산 대상 필드 목록. 쉼표(,)를 구분자로 사용합니다. 필드 값은 숫자형이어야 하며, 지정된 필드 값이 숫자가 아닌 입력 레코드는 무시됩니다. 최대 10만개의 입력 레코드를 허용하며, **\_cluster** 필드에 1부터 증가하는 번호로 클러스터를 분류하여 출력합니다. 유효 입력 레코드가 10만개를 넘으면 쿼리를 종료합니다.

선택 매개변수

**k=INT**

클러스터 수(기본값: 3)

**iter=INT**

계산 반복 횟수(기본값: 10000)

#### 사용 예

머신 러닝에서 흔히 인용되는 iris 데이터를 이용하여 시험할 수 있습니다. 길이와 너비를 이용하여 분류를 수행한 후, 실제 종(species) 이름과 비교해봅니다(다운로드: <https://github.com/illinois-cse/data-fa14/blob/gh-pages/data/iris.csv>).

csvfile /opt/logpresso/iris.csv| eval sepal\_length = double(sepal\_length), sepal\_width = double(sepal\_width)| kmeans k=4 iter=100000 sepal\_length, sepal\_width