* Thuật toán k-mean?
* K-mean phù hợp với thuật toán có dạng lồi.
* Để đánh giá ksết quả phân cụm ta có thể tính tb khoảng cách từ trung tâm đến cluster các centroid của nó.
* The BFR Algorithm?
* Thuật toán BFR là mở rộng của k-means dành cho trường hợp của dữ liệu lớn.
* Các bước của thuật toán BFR?

**\* Lưu ý**: Đủ gần là khoảng cách điểm dữ liệu đó đến các dữ liệu khác và centroid là đủ gần. Với các điểm dữ liệu không đủ gần tiến hành phân cụm và tạo ra các cluster mới.

Bước 5: Gộp các cluster mới tạo ra ở bước 4 với các cluster có từ ban đầu nếu đủ gần.

Bước 6: Lặp lại bước 2 đến 5 khi các dữ liệu đủ gần.

Load lô dữ liệu đầu tiên với k centroid

Cách 1: Chọn ngẫu nhiên k điểm dữ liệu

Cách 2: Tiến hành phân cụm để có k centroid.

Cách 3: Lần lượt chọn các điểm dữ liệu sao cho điểm dữ liệu càng xa điểm dữ liệu gần càng tốt.

* The “Memory-Load” of Points?

Cách lưu trữ một cluster(summarizing):

+ N: số lượng điểm dữ liệu

+ SUM: tổng các điểm

+ SUMSQ: tổng bình phương các điểm