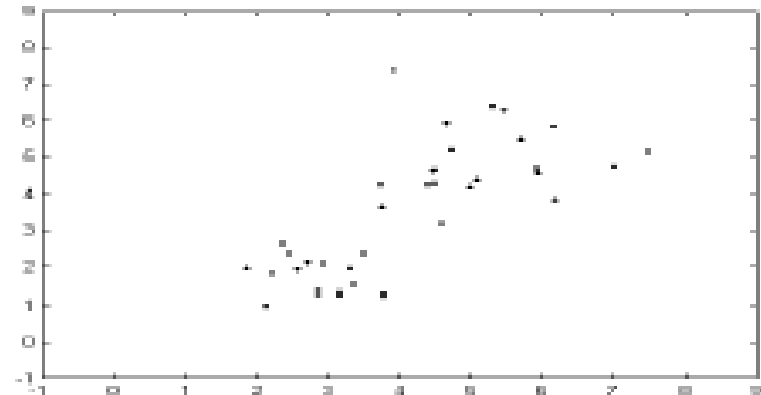
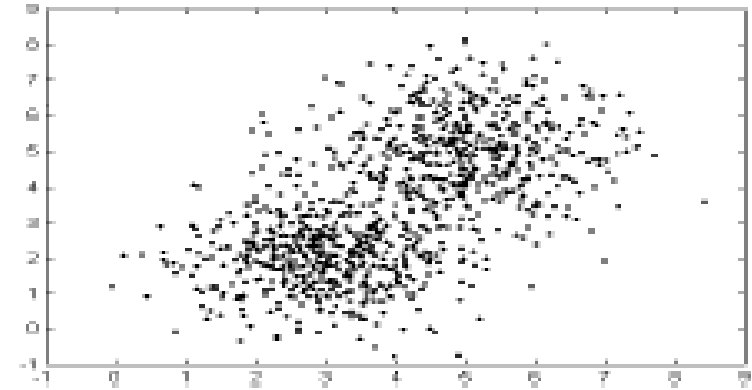


Algoritma Reduksi Data

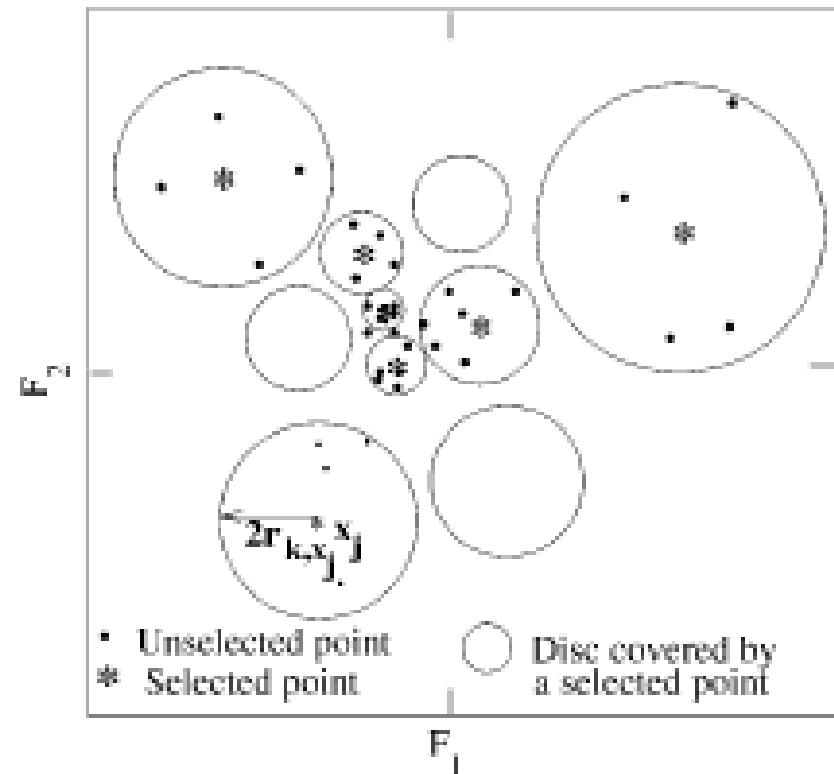
- Pemilihan sub sampel dari data merupakan hal yang biasa dilakukan dalam proses data mining.
- Pemilihan sub sampel dilakukan pada data yang mendekati pada model data sesungguhnya



Density-Based Multiscale Data Condensation

- Prinsip dasar algoritma ini adalah mengurutkan titik-titik berdasarkan estimated densities, memilih titik-titik yang padat, dan menghapus titik lain yang berada dalam jarak tertentu dari titik yang dipilih sebagai bagian dari sampel data.
- Metode non-parametrik dalam memperkirakan probability density function adalah metode k-nearest-neighbor.
- Pada metode k-NN kepadatan titik dihitung berdasarkan area pada suatu lingkaran yang berisi sejumlah k titik yang berketetanggaan.

Density-Based Multiscale Data Condensation (cont.)



Density-Based Multiscale Data Condensation (cont.)

Himpunan $B_N = \{x_1, x_2, \dots, x_N\}$ sebagai data set inputan. Pilih nilai integer positif k

- Untuk tiap titik $x_i \in B_N$ hitung jarak k^{th} nearest neighbor dari x_i pada B_N . Tandai dengan r_{k,x_i} .
- Pilih $x_j \in B_N$ yang mempunyai r_{k,x_j} terkecil dan letakkan pada himpunan E .
- Hapus semua titik dari B_N yang berada dalam lingkaran radius $2 r_{k,x_j}$ yang berpusat di x_j dan titik-titik yang tersisa di himpunan sebagai B_N .