

Data Mining - Latihan 3

Ali Akbar Septiandri
Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Al-Azhar Indonesia

June 2, 2017

Clustering

Diberikan data dengan nilai numerik tunggal sebagai berikut: $\{1, 2, 4, 7, 11, 16, 17\}$

1. Deskripsikan bagaimana algoritma k-Means dapat membagi data tersebut ke dalam k klaster!
2. Asumsikan $k = 2$ dan inisialisasi $\mu_1 = 0$ dan $\mu_2 = 10$. Buatlah daftar objek yang masuk ke dalam tiap klaster setelah iterasi pertama dan kedua. Kapan Anda akan menghentikan iterasi?
3. Asumsikan Anda akan menggunakan algoritma EM untuk mencocokkan Gaussian Mixture Models pada data tersebut. Gunakan nilai awal $\mu_1 = 2$ dan $\mu_2 = 15$ dan $\sigma_1 = \sigma_2 = 1$. Asumsikan bahwa Anda menggunakan *uniform prior*. Hitung probabilitas posterior untuk kedua Gaussians untuk observasi $x = 7$. Laporkan hasilnya hingga dua desimal.
4. Lakukan *agglomerative clustering* untuk data tersebut dengan menggunakan metode *single link*. Gambarkan dendrogram yang dihasilkan!

Dimensionality Reduction

Diberikan data sebagai berikut:

```
x1 = [1 2 3 4]
x2 = [5 6 7 8]
x3 = [9 0 1 2]
x4 = [1 5 9 9]
x5 = [1 6 9 9]
```

1. Empat vektor eigen $e_1 \dots e_4$ dari dekomposisi dengan PCA untuk data di atas adalah sebagai berikut:

```
e1 = [ 0.00  0.26  0.52 -0.81]
e2 = [ 0.07 -0.82  0.56  0.09]
e3 = [ 0.88  0.28  0.25  0.26]
e4 = [-0.46  0.41  0.58  0.52]
mean = [ 3.40  3.80  5.80  6.40]
```

Nilai eigennya adalah $\lambda_1 = 0.04, \lambda_2 = 1.94, \lambda_3 = 25.87, \lambda_4 = 146.15$. Hitunglah koordinat vektor x_1 dalam dua dimensi dengan menggunakan vektor eigen e_1 dan e_2 .

2. Anda ingin menggunakan komponen prinsipal yang menjelaskan $\geq 95\%$ variansi dalam data. Bagaimana Anda dapat melakukannya?