14. Собствени вектори и собствени стойности на ковариационна матрица. Метод на главните компоненти за редуциране на признаковото пространство [14].

(Виж страница 59 ПР–част1)

Собствени стойности и собствени вектори

Задачата се състои в намиране на N числа λ_n и съответните им N-мерни вектори x_n такива, че за дадена квадратна матрица A да имаме:

$$A \cdot x_n = \lambda_n x_n \tag{6}$$

или

$$(A-\lambda_n I)x_n=0$$

Числата λ_n се наричат **собствени стойности** на матрицата A (говорим още за **спектър** на собствените стойности) а n-мерните вектори x се наричат **собствени вектори**. Преди да се запознаем по-подробно с някои методи за намиране на собствените стойности и вектори на една матрица ще припомним някои основни понятия и свойства свързани с матриците:

1. Една матрица се нарича **симетрична** ако е равна на транспонираната си:

$A=A^T$ или $a_{ij}=a_{ji}$;

2. Една матрица се нарича **Ермитова** ако е равна на комплексно спрегнатата на транспонираната си:

$A=A^{*T}=A^{\dagger}$

За реални матрици понятията Ермитова и симетрична съвпадат.

3. Една матрица се нарича **ортогонална** ако транспонираната й е равна на обратната й

$A^T.A=A^{-1}.A=I$

4. Една матрица се нарича **нормална** ако комутира с Ермитово спрегнатата си

$A^{\dagger}.A=A.A^{\dagger}$.

Всяка Ермитова матрица е нормална (обратното не е вярно – не всяка нормална матрица е Ермитова). Собствените вектори на една нормална матрица образуват пълен базис в n-мерното пространство.

5. Ако собствените стойности на една матрица са λ_n, то собствените стойности на Ермитово спрегнатата й са комплекно спрегнатите числа $\overline{\lambda_n}$.

Оттук следва, че собствените стойности на всяка Ермитова матрица са реални.

Равенството (1) представлява една хомогенна система от n линейни уравнения с n неизвестни x_n. Както е известно такава система има нетривиално решение само ако детерминантата й е нула:

$$|A - \lambda I| = 0 \tag{7}$$

или разписано подробно за квадратна четиримерна матрица:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22}-\lambda & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33}-\lambda & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44}-\lambda \end{vmatrix} = 0 \tag{8}$$

От определението за детерминанта следва, че (7) представлява полином от n-та степен спрямо неизвестната величина λ. Този полином се нарича **характеристичен полином**. Един полином отn-та степен има n корена – не непременно различни. Ако матрицата A е реална корените могат да бъдат както реални така и комплексни (два по два комплексно спрегнати). Ако A е Ермитова корените са реални.

И така задачата се сведе до две подзадачи:

А. Намиране на характеричния полином

Б. Намиране на корените на характеристичния полином

В случай на най-общ вид на матрицата A намирането на собствените стойности и собствени вектори е много сложна и изисква специални знания по теория на матриците. Затова ще се спрем (както и в повечето учебници) на случая, когато A е симетрична. Има и още една причина симетричните матрици да представляват такъв интерес – много физични задачи водят до системи линейни уравнения със симетрични матрици.

За да стане ясно защо симетричните матрици представляват особен клас в тази задача ще въведем понятието **подобно преобразование**. Казваме, че две матрици A и Á са **подобни**, (еквивалентни, свързани с подобно преобразование) ако съществува неизродена матрица H, такава че:

$$Á=H^{-1}AH \qquad (9)$$

В сила са следните две твърдения:

Подобните матрици имат еднакъв спектър от собствени стойности
Всяка симетрична матрица може да се преобразува в диагонална посредством подобни преобразования.

Тези две теореми са в основата на почти всички популярни методи за намиране на собствени стойности и вектори на симетрични матрици (всъщност второто твърдение се отнася за всяка нормална матрица). Стратегията на такива методи е следната: посредством серия от подобни преобразования матрицата все повече се приближава до диагонална. Спектърът на диагоналната матрица се намира особено просто: това са самите диагонални елементи. С това задачата е решена.

Метод на главните компоненти за редуциране на признаковото пространство.

Когато данните ни с в много имерно пространство това създава проблеми:

- много физическо място за описание на данните
- по-сложност
- по-голяма вероятност за грешка

Затова би било хубаво ако можем да отделим важните фактори за разпознаването от по-маловажните.

Едно решение е като се опитаме да направим чрез линейни трансформации проекция на многомерното пространство в пространство с по-малко измерения. Освен това искаме това проектиране да стане с най-доброто по-възможност представяне на данните в смисъла на минималната квадратична грешка.

Искаме максимално количество информация с колкото се може по-малко ортогонални линейни компоненти.

Нека представим x

$$x = \sum_{i=1}^{n} y_i \varphi_i = Y\Phi, \text{където } Y = [y1, y2, \dots, yn]^T, a\ \Phi = [\varphi_1, \varphi_2, \dots, \varphi_n].$$

освен това за φ:

$$\varphi_i^T \varphi_j = \begin{cases} 1, & \text{при } i=j \\ 0, & \text{при } i \neq j \end{cases}$$

$$y_i = \varphi_i^T x$$

Интересува ни квадратичната грешка при това представяне. За да я изчислим нека представим x само с m на брой координати, а изберем останалите произволно:

$$\hat{x}(m) = \sum_{i=1}^{m} y_i \varphi_i + \sum_{i=m+1}^{n} b_i \varphi_i$$

$$\Delta x(m) = x - \hat{x}(m) = \sum_{i=m+1}^{n} b_i \varphi_i$$

Квадратичната грешка:

$$\varepsilon^2(m) = E\left(\left|\left|\Delta x(m)\right|\right|^2\right) = E\left\{\sum_{i=m+1}^{n}\sum_{j=m+1}^{n} (y_i - b_i)\left(y_j - b_j\right)\varphi_i^T \varphi_j\right\}$$

$$= \sum_{i=m+1}^{n} E\{(y_i - b_i)^2\}$$

Искаме минимална грешка => производната по bi трябва да е равна 0:

$$2[E\{y_i\} - b_i] = 0$$

$$b_i = E\{y_i\}$$

Заместваики обратно получаваме:

$$\varepsilon^2(m) = \sum_{i=m+1}^{n} E\{(y_i - E\{y_i\})^2\}$$

Заместваме yi:

$$\varepsilon^2(m) = \sum_{i=m+1}^{n} \varphi_i^T E\{(x - E\{x\})^2\}^T \varphi_i = \sum_{i=m+1}^{n} \varphi_i^T \Sigma_x \varphi_i$$

Тук Σ_x е ковариационна матрица за x, за да минимизираме грешката остава да кажем, че трябва φ_i да са собствените вектори на Σ_x така получаваме:

$$\varepsilon^2(m) = \sum_{i=m+1}^{n} \lambda_i$$

Ако m=0 => $\varepsilon^2(m) = \sum_{i=1}^{n} \lambda_i$ – което се нарича следа на ковариационната матрица(T_r) и сумата е константа за някакъв избор на векторите x.

Сега за да редуцираме признаковото пространство по m признака остава да подредим :

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n$$

и така вектора с най-голяма собствена стойност е първия главен компонент, и т.н. Остава ни само да изберем първите m.

П.С. Тук явно ще трябва да се рисуват картинки или да се описва какво са собствени стойности и вектори... НЗ ☹