秦格华-5140219335
03/05/2017

Homework 3

6.1.1 Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if i divides b with no remainder. Thus, item 1 is in all the baskets, item 2 is in all fifty of the even-numbered baskets, and so on. Basket 12 consists of items 1, 2, 3, 4, 6, 12, since these are all the integers that divide 12. Answer the following questions:

(a) If the support threshold is 5, which items are frequent?

Solution.

If the items' threshold is 5, which means they occurs in exactly 5 baskets, they can exactly divide 5 integers within 1 to 100. Since only $x \in [1, 20]$ can be divisor more or equal as 5 times, we only have to check these 20 numbers to see whether satisfied. As a result, we conclude $\forall x \leq 20$ is frequent with threshold as 5.

6.1.5 For the data of Exercise 6.1.1, what is the confidence of the following association rules?

(a) $\{5, 7\} \to 2$;

(b) $\{2, 3, 4\} \to 5$.

Solution.

(a) We know 5 and 7 appear simultaneously in baskets 35 and 70, which only basket 70 contains 2. Thus, the confidence is $1/2$.

(b) We know 2, 3, and 4 appear in 12, 24, 36, 48, 60, 72, 84, and 96, amony which 5 only appears in 60. Thus, the confidence is $1/8$.

11.1.3 For any symmetric 3 × 3 matrix

$$\begin{bmatrix} a-\lambda & b & c \\ b & d-\lambda & e \\ c & e & f-\lambda \end{bmatrix}$$

there is a cubic equation in λ that says the determinant of this matrix is 0. In terms of a through f, find this equation.

Solution.

$$\begin{vmatrix} a-\lambda & b & c \\ b & d-\lambda & e \\ c & e & f-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda)(f-\lambda) + 2bce - b^2(f-\lambda) - c^2(d-\lambda) - e^2(a-\lambda)$$

$$= -\lambda^3 + (a+d+f)\lambda^2 + (b^2+c^2+e^2-ad-af-df)\lambda + adf + 2bce = 0$$

11.2.1 Let M be the matrix of data points

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix}$$

(a) What are M^TM and MM^T?

(b) Compute the eigenpairs for M^TM.

(c) What do you expect to be the eigenvalues of MM^T?

(d) Find the eigenvectors of MM^T, using your eigenvalues from part (c).

Solution.

(a) $M^TM = \begin{bmatrix} 30 & 100 \\ 100 & 354 \end{bmatrix}$ $MM^T = \begin{bmatrix} 2 & 6 & 12 & 20 \\ 6 & 16 & 42 & 72 \\ 12 & 42 & 90 & 156 \\ 20 & 72 & 156 & 272 \end{bmatrix}$

(b) I used library functions in `numpy` to get the eignpairs:

```
>>>import numpy
>>>x =numpy.array([[30., 100.], [100., 354.]])
>>>a,b =numpy.linalg.eig(x)
```

The results are as follows:

```
>>>a
>>>array([ 1.62142978, 382.37857022])
b
>>>array([[-0.9620125 , -0.27300539],
      [0.27300539, -0.9620125 ]])
```

a contains the eignvalues and b contains corresponding eignvectors.

(c) From the lecture we know the following equation holds,

$$M^T Me = \lambda e \Rightarrow MM^T(Me) = M\lambda e = \lambda(Me). \tag{1}$$

We know the eigenvalues of MM^T are these of $M^T M$ adding two 0, which is $[1.62142978, 382.37857022, 0, 0]$

(d) From (c) we can derive that the eigenvetors of MM^T are these of $M^T M$.

I also used Python to get the result:

```
>>>v*vv
>>>matrix([[ 2., 6., 12., 20.],
       [6., 20., 42., 72.],
       [12., 42., 90., 156.],
       [20., 72., 156., 272.]])

>>>p,q =numpy.linalg.eig(v*vv)
>>>p
>>>array([ 3.82378570e+02, 1.62142978e+00, 3.61455268e-15,
       -4.00465957e-16])
q
>>>matrix([[-0.06315773, 0.54109638, 0.80632338, -0.51223888],
       [-0.22470839, 0.65339536, -0.26378205, -0.18715112],
       [-0.484652 , 0.33689693, -0.45461397, 0.76177371],
       [-0.84298854, -0.40839891, 0.27127066, -0.349695 ]])
```

11.3.2 : Use the SVD from Fig. 11.7. Suppose Leslie assigns rating 3 to Alien and rating 4 to Titanic, giving us a representation of Leslie in "movie space"of $[0, 3, 0, 0, 4]$. Find the representation of Leslie in concept space. What does that representation predict about how well Leslie would like the other movies appearing in our example data?

Solution.

We can multiply $[0, 3, 0, 0, 4]$ with V:

$$\begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix}^T \times \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.0092 \\ 1.0092 \\ 1.0092 \\ 2.0164 \\ 2.0164 \end{bmatrix}$$

This result shows that Leslie is more likely to like *Casablanca*.

11.4.2 Find the CUR-decomposition of the matrix of Fig. 11.12 when we pick two "random" rows and columns as follows:

(a) The columns for *The Matrix* and *Alien* and the rows for Jim and John.

Solution.

For the columns, scale the two columns by $\sqrt{rq_1} = \sqrt{rq_2} = \sqrt{2 \times 51/243} = 0.648$, so the matrix C is:

$$C = \begin{bmatrix} 1.54 & 1.54 \\ 4.63 & 4.63 \\ 6.17 & 6.17 \\ 7.72 & 7.72 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Sililarly, we also scale the two rows. For Jim, we have $\sqrt{rp_2} = \sqrt{2 \times 27/243} = 0.471$. For John, we have $\sqrt{rp_3} = \sqrt{2 \times 48/243} = 0.6285$. So the matrix R is:

$$C = \begin{bmatrix} 6.37 & 6.37 & 6.37 & 0 & 0 \\ 6.36 & 6.36 & 6.36 & 0 & 0 \end{bmatrix}$$

For matrix W,

$$W = X\Sigma Y^T = \begin{bmatrix} -0.6 & -0.8 \\ -0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\Sigma^+ = \begin{bmatrix} \frac{1}{5\sqrt{2}} & 0 \\ 0 & 0 \end{bmatrix}$$

$$U = Y(\Sigma^+)^2 X^T = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0.02 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.6 & -0.8 \\ -0.8 & -0.6 \end{bmatrix} = \begin{bmatrix} 0.0085 & 0.0085 \\ -0.0113 & 0.0113 \end{bmatrix}$$