CS389: Foundations of Data Science Homework II

Zihao Ye

December 26, 2016

EXERCISE 2.20

Suppose r is the radius of the upper surface of cylinder, the volume of the cylinder will be:

$$\sqrt{1-r^2} \cdot r^d V(d)$$

Since $V(d)$ has nothing to do with r, we just need to optimize $\sqrt{1-r^2} \cdot r^{d-1}$, considering

$$\frac{\partial}{\partial r}\left(\sqrt{1-r^2} \cdot r^{d-1}\right) = \frac{r^{d-2}\left((d-1)(r^2-1)+r^2\right)}{\sqrt{1-r^2}}$$

when $r = \sqrt{\frac{d-1}{d}}$, i.e. $h = \sqrt{\frac{1}{d}}$ the volume will reach its peak.

EXERCISE 2.26

The intersection of a narrow slice at the equator and a narrow annulus at the surface of the ball is not empty(In 2D and 3D situations, it looks like a doughnut). In high dimensions, we can prove that most of the volume of a annulus at the surface of the ball concentrate on a narrow slice at the equator.

EXERCISE 2.27

The experiment result shows that about $330-360$ points are in each band. Only 75 points are in all five bounds.

To analyse the threshold more precisely, I draw a graph of the relation between c(coefficient of the threshold) and the probability that all points are in all five bounds.

This figure shows that let the band width be $2 \times \frac{4}{\sqrt{50}}$, then with high probability all points are in all five bands.

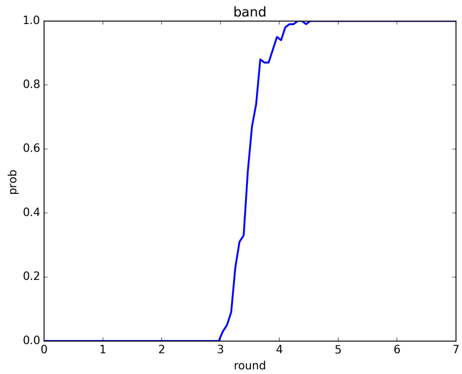


Figure 0.1: Prob-c

EXERCISE 2.51

1. Most of the volume of a ball in high dimensions is contained in the annulus.
2. Most of the volume of a ball in high dimensions is contained in the equators.
3. If we draw two points at random from the unit ball, with high probability their vectors will be nearly orthogonal to each other.
4. In high dimensions, a unit cube will not be entirely covered by a unit ball.
5. Under some conditions, it's possible for us to project high-dimension points to low-dimension ones while preserving the distances between each pair by a coefficient $(1 \pm \varepsilon)\sqrt{k}$.

EXERCISE 3.5

(A)

$$M = \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ 3 & 0 \end{bmatrix}$$

According to the fact

$$\mathbf{v}_1 = \arg\max_{|\mathbf{v}|=1} |A\mathbf{v}|$$

we derive

$$V = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

and(by calculating $|A\mathbf{v}|^2$, $A\cdot V = U\cdot D$)

$$D = \begin{bmatrix} \sqrt{11} & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$$

$$U = \begin{bmatrix} \sqrt{\frac{2}{11}} & 0 & -\frac{3}{\sqrt{11}} \\ \frac{3}{\sqrt{22}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{11}} \\ \frac{3}{\sqrt{22}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{11}} \end{bmatrix}$$

(B)

$$M = \begin{bmatrix} 0 & 2 \\ 2 & 0 \\ 1 & 3 \\ 3 & 1 \end{bmatrix}$$

Use the same technique, we derive

$$V = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

and

$$D = \begin{bmatrix} 2\sqrt{5} & 0 \\ 0 & 2\sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$U = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{1}{2} & -\frac{1}{\sqrt{14}} & -\frac{9}{2\sqrt{35}} \\ \frac{1}{\sqrt{10}} & -\frac{1}{2} & -\frac{3}{\sqrt{14}} & \frac{1}{2\sqrt{35}} \\ \sqrt{\frac{2}{5}} & \frac{1}{2} & 0 & \frac{\sqrt{\frac{7}{5}}}{2} \\ \sqrt{\frac{2}{5}} & -\frac{1}{2} & \sqrt{\frac{2}{7}} & -\frac{3}{2\sqrt{35}} \end{bmatrix}$$

Exercise 3.8

(A)

$|A\mathbf{v}_1|$ will reach its peak when $\mathbf{v}_1 = \begin{bmatrix} \sqrt{\frac{k}{n}} & \sqrt{\frac{k}{n}} & \cdots & 0 \end{bmatrix}$, whose first $\sqrt{\frac{k}{n}}$ entries are not zero.

$\mathbf{v}_2$ must be orthogonal with $A\mathbf{v}_1$, if some of its entries related to $\mathbf{v}_1$ is not 0, then their sum must be 1, which has nothing to do with $|A\mathbf{v}_2|$. Thus all its non-zero entries will have no intersection with those related to $\mathbf{v}_1$, then we will derive $\mathbf{v}_1 = \begin{bmatrix} 0 & 0 & \cdots & \sqrt{\frac{k}{n}} & \sqrt{\frac{k}{n}} & \cdots & 0 \end{bmatrix}$, whose second $\sqrt{\frac{k}{n}}$ entries are not zero.

By using this routine again and again, we derive $\mathbf{v}_i$ has exactly $\sqrt{\frac{k}{n}}$ non-zero entries corresponding to the i-th block.

(B)

If $a_1 = a_2 = \cdots, a_k$, the order of these singular vectors isn't important, thus we could arrange those singular vectors in any order.

EXERCISE 3.27

(A)

Here are some images generated by Python.

(a) Use top 5% singular values

(b) Use top 10% singular values

(c) Use top 25% singular values

(d) Use top 50% singular values

Figure 0.2: Images generated by SVD using different number of singular values.

(B)

The following tables shows how *Forbenius Norm* changes when ratio of used singular values grows.

Table 0.1: My caption

Ratio of used singular values	0.05	0.10	0.25	0.50
Percentage of Forbenius norm	0.99683	0.99862	0.99966	0.99995

EXERCISE 3.30

(A)

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T = \mathbf{x}_i\mathbf{x}_i^T + \mathbf{x}_j\mathbf{x}_j^T - 2\mathbf{x}_i\mathbf{x}_j^T$$

$$\mathbf{x}_i\mathbf{x}_j^T = \frac{1}{2}\left(\mathbf{x}_i\mathbf{x}_i^T + \mathbf{x}_j\mathbf{x}_j^T - d_{ij}^2\right)$$

From the facts that

$$\left(\sum_{k=1}^{n}\mathbf{x}_k\right)\left(\sum_{k=1}^{n}\mathbf{x}_k\right)^T = 0$$

$$\left(\sum_{k=1}^{n}\mathbf{x}_k\right)\mathbf{x}_i^T = 0$$

$$\left(\sum_{k=1}^{n}\mathbf{x}_k\right)\mathbf{x}_j^T = 0$$

We derive:

$$\sum_{k=1}^{n}\mathbf{x}_i\mathbf{x}_i^T = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}^2$$

$$\mathbf{x}_i\mathbf{x}_i^T = \frac{1}{n}\left(\sum_{k=1}^{n}d_{ki}^2 - \sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^T\right)$$

$$\mathbf{x}_j\mathbf{x}_j^T = \frac{1}{n}\left(\sum_{k=1}^{n}d_{kj}^2 - \sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^T\right)$$

Combining these facts, we have:

$$\mathbf{x}_i\mathbf{x}_j^T = \frac{1}{2}\left[-d_{ij}^2 + \frac{1}{n}\sum_{k=1}^{n}d_{ki}^2 + \frac{1}{n}\sum_{k=1}^{n}d_{kj}^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}^2\right]$$

(B)

The left singular vectors $\mathbf{u}_j$ of X are eigenvectors of XX^T with eigenvalue σ^2.

Algorithm Description:

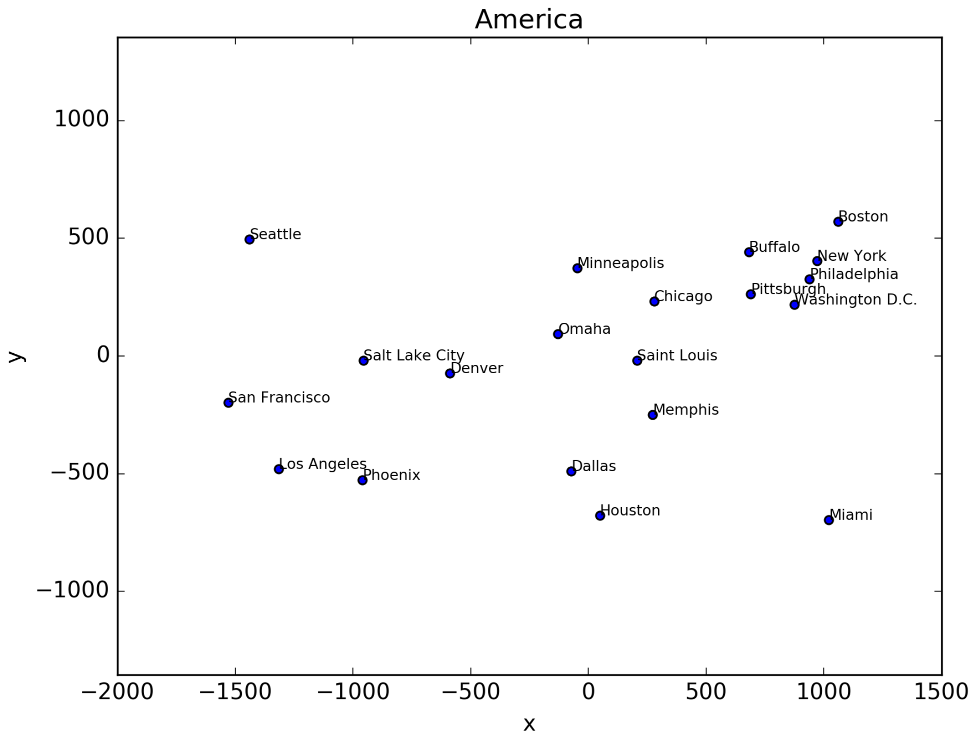
1. Compute the eigenvalue and eigenvectors of XX^T.
2. Choose the eigenvectors correspond to positive eigenvalues(If there are more than d positive eigenvalues, choose the largest d of them), let them be U' and $\Sigma'\Sigma'^T$, then $U' \cdot \Sigma'$ is a feasible solution.

Proof: Suppose $X = U\Sigma V^T$, then $U\Sigma = XV$, which is equivalent to X by translation, rotation or reflection.

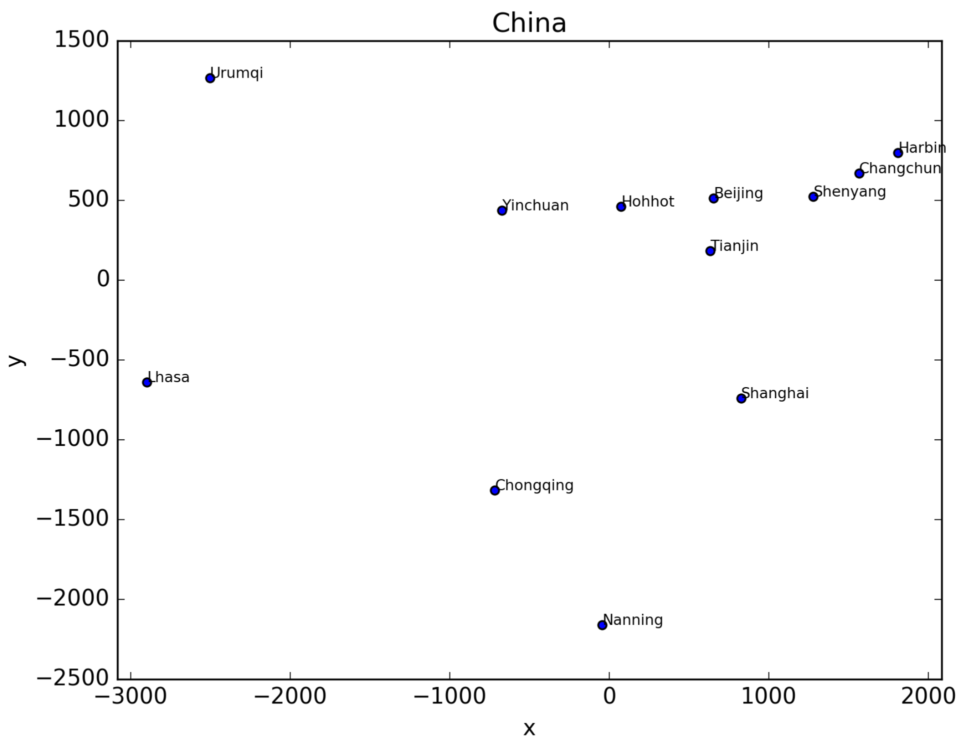
EXERCISE 3.31

(A)

Here are two images generated by Python using classical multidimensional scaling algorithm.



(a) American cities



(b) Chinese cities

Figure 0.3: Placement in cities of US and China generated by classical multidimensional scaling.

(B)

If we regard the airline distance as the shortest path in euclidean space, we could apply the algorithm stated above to construct a 3-dimensional world model.

Otherwise we need to know the radius of the earth, which is a little bit hard to acquire from the data.

Exercise 4.39

(A)

$$E[X] = \sqrt{n}p$$

When $p = o\left(\frac{1}{\sqrt{n}}\right)$, $n \to \infty \Longrightarrow E[x] \to 0$. When $p > \frac{1}{\sqrt{n}}$, $P = 1-(1-p)^{\sqrt{n}} > 1-e^{-p\sqrt{n}}$, with high probility, $N(n,p)$ will contain a perfect square. Thus $\frac{1}{\sqrt{n}}$ is a threshold for this property.

(B)

Use the same technique as showed above.

$$E[X] = \sqrt[3]{n}p$$

$$P = 1-(1-p)^{\sqrt[3]{n}} > 1-e^{-p\sqrt[3]{n}}$$

We derive that $\frac{1}{\sqrt[3]{n}}$ is a threshold for this property.

(C)

$$E[X] = pn/2$$

$$P = 1-(1-p)^{n/2} > 1-e^{-pn/2}$$

$\frac{2}{n}$ is a threshold for this property.

(D)

$$E[X] \approx \frac{n^2}{2}p^3$$

$$E[X^2] < \frac{n^2}{2}p^3 + 2\left(\frac{2n^2}{2}p^4 + \frac{3n^3}{2}p^5 + \frac{n^4}{8}p^6\right)$$

When $p = o\left(\sqrt[3]{\frac{1}{2n^2}}\right)$, $n \to \infty \Longrightarrow E[x] \to 0$. When $p > \sqrt[3]{\frac{1}{2n^2}}$, $E[X^2] = E[X]^2(1+o(1))$. By the second moment argument, $\sqrt[3]{\frac{1}{2n^2}}$ is a threshold for this statement.