University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2019)

Solutions 4

Solutions

Problem 1: Consider the matrix

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 2 \\ 1 & -2 \\ -1 & -2 \end{pmatrix}.$$

1. Run the *power method* starting from $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ for $k = 3$ steps. What does this give as estimates for v_1 and σ_1?
2. What are the actual values of v_i's, σ_i's and u_i's? You might find it helpful to first compute the eigenvalues and eigenvectors of $B = A^\intercal A$.
3. Suppose matrix A is a database of restaurant ratings: each row corresponds to a person, each column to a restaurant, and the entries A_{ij} represent how much person i likes restaurant j. What might v_1 represent? What about u_i? What about the gap $\sigma_1 - \sigma_2$?

Solution:

1. Recall that the power method computes a sequence of vectors $\{x_n\}$ such that $x_i = Bx_{i-1}$ for all $1 \le i \le k$, where the matrix $B = A^\intercal A$. In our case we have that

$$B = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 2 & 2 & -2 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 2 \\ 1 & -2 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}$$

   After $k = 3$ runs of the power method, we obtain a vector

$$x_3 = B^3 x = \begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}^3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 64 & 0 \\ 0 & 4096 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 64 \\ 4096 \end{pmatrix}$$

   The estimate for v_1 is given by

$$\tilde{v_1} = \frac{x_3}{\|x_3\|} \simeq \begin{pmatrix} 0.0152 \\ 0.9998 \end{pmatrix}.$$

   Also, the estimate for σ_1 is given by

$$\tilde{\sigma_1} = \|A\tilde{v_1}\| \simeq 3.9996.$$

2. Since the matrix B is already in diagonal form, its eigenvaues are simply the entries on the diagonal. Thus we have that $\lambda_1 = 16$ and $\lambda_2 = 4$. Recall that the eigenvalues of B are the squares of the singular values of the matrix A, therefore $\sigma_1 = 4$ and $\sigma_2 = 2$. Moreover, we know that the right-singular vectors v_i are the eigenvectors of B corresponding to λ_i. One has that $v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. For the left-singular vectors u_i we compute

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{4} \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

and

$$u_1 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

3. Recall that

$$v_1 \triangleq \arg\max_{\|v\|=1} \|Av\| = \arg\max_{\|v\|=1} \sum_{i=1}^{d} (A_i v) .$$

Each A_i represents the ratings that customer i gives to each restaurant. Therefore v_1 is indicating the most preferred restaurant according to the customers. Similarly, u_1 indicates the most prefered customer from the perspective of the most popular restaurant. The gap $\sigma_1 - \sigma_2$ indicates the difference between the top two most preferred restaurants.

Problem 2: Let $v \in \mathbb{R}^n$ such that $\|v\| = 1$. Sample uniformly $x \in \{-1, 1\}^n$, and define $S = \langle x, v \rangle$. Prove that

$$\mathbf{E}\left[S^4\right] = 3 \sum_{i=1}^{n} v_i^2 - 2 \sum_{i=1}^{n} v_i^4 \leq 3.$$

That is, prove the inequality from the Proof of Lemma 2 in Lecture 6.

***Solution*:** We have that

$$\begin{aligned}
\mathbf{E}\left[\, S^4 \,\right] &= \mathbf{E}\left[\left(\sum_{i=1}^{n} x_i v_i \right)^2 \right] \\
&= \mathbf{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{\ell=1}^{n} x_i x_j x_k x_\ell v_i v_j v_k v_\ell \right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{\ell=1}^{n} \mathbf{E}\left[\, x_i x_j x_k x_\ell \,\right] v_i v_j v_k v_\ell \\
&= \sum_{i=1}^{n} \mathbf{E}\left[\, x_i^4 \,\right] v_i^4 + \frac{1}{2}\binom{4}{2} \sum_{i \neq j} \mathbf{E}\left[\, x_i^2 x_j^2 \,\right] v_i^2 v_j^2 \\
&= \sum_{i=1}^{n} v_i^4 + 3 \sum_{i \neq j} v_i^2 v_j^2 \\
&= 3 \left(\sum_{i=1}^{n} v_i^2 \right) \left(\sum_{j=1}^{n} v_j^2 \right) - 2 \sum_{i=1}^{n} v_i^4 \\
&= 3 \|v\|^4 - 2 \sum_{i=1}^{n} v_i^4 \\
&\leq 3.
\end{aligned}$$

In the third line we used the linearity of the expectation. The equality in the fourth line comes from the fact that under expectation, all products of x_i's vanish when at least one factor has odd power. Finally the last inequality comes from the fact that we chose v to be a unit vector.

Problem 3: Let $A \in \mathbb{R}^{n \times n}$ be a symmetric and PSD matrix. Show that the power method can be applied to approximately compute the smallest eigenvalue of A.

Solution: Suppose A has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, counting multiplicities. First, we can run the power method to find a good approximation of the largest eigenvalue of A, say d is the approximated largest eigenvalue of λ_1. Using this, we can upper bound λ_1 by a constant, say $2d$. Consider the matrix $B = 2D - A$, where D is a diagonal matrix with each diagonal entry being equal to d. Notice that this ensures that matrix B is a PSD matrix. We claim that for every eigenvalue λ_i of A with corresponding eigenvector v_i, $2d - \lambda_i$ is an eigenvalue of B. Indeed we have that

$$Bv_i = (2D - A)v_i = 2Dv_i - Av_i = 2dv_i - \lambda_i v_i = (2d - \lambda_i)v_i.$$

Also note that the smallest eigenvalue of A, i.e. λ_n, corresponds to the largest eigenvalue of B, which is $2d - \lambda_n$. Hence we can run the power method for B to get an estimate for $2d - \lambda_n$ and subtract it from $2d$ to get an estimate of λ_n.

Problem 4: Let u be a fixed vector. Show that maximising $x^\intercal uu^\intercal(1-x)$ subject to $x_i \in \{0,1\}$ is equivalent to partitioning the coordinates of u into two subsets where the sum of the elements in both subsets are as equal as possible.

Solution: Suppose that the vectors x and u are n-dimensional. Let $f(x) = x^\intercal uu^\intercal(1-x)$. We have that

$$\begin{aligned} f(x) &= \left(\sum_{i=1}^n x_i u_i\right)\left(\sum_{j=1}^n u_j(1-x_j)\right) \\ &= \sum_{i,j=1}^n x_i(1-x_j)u_i u_j \\ &= \sum_{i:x_i=1}\sum_{j:x_j=0} x_i(1-x_j)u_i u_j \\ &= \left(\sum_{i:x_i=1} u_i\right)\left(\sum_{j:x_j=0} u_j\right). \end{aligned}$$

Let $a = \left(\sum_{i:x_i=1} u_i\right)$ and $b = \left(\sum_{j:x_j=0} u_j\right)$. Note that $a + b = \sum_{i=1}^n u_i = c$ for some constant c since the vector u is fixed. Therefore, the problem of maximising $f(x)$ subject to x, is equivalent to maximising the product ab, subject to the constraint $a + b = c$. We have seen from last week's tutorial that ab is maximised for $a = b$. In our case a and b take discrete values over the random sampling of x, hence $f(x)$ is maximised when $|a - b|$ is minimised. In other words, when we can partition the entries of u into two sets such that the sum of entries in the two sets is as equal as possible.

Problem 5 *(Optional)*: Let $x_1, x_2, \ldots, x_n$ be n points in a d-dimensional space and let X be an $n \times d$ matrix whose rows are the n points. Suppose we know only the matrix D of pairwise distances between points and not the coordinates of the points themselves. The set of points $x_1, x_2, \ldots, x_n$ giving rise to the matrix D is not unique since any translation, rotation or reflection of the coordinate system preserves the distances. Fix the origin of the coordinate system so that the centroid of the set of points is at the origin. That is, $\sum_{i=1}^n x_i = 0$.

1. Show that the elements of $XX^\intercal$ are given by

$$x_i^\intercal x_j = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{k=1}^n d_{ik}^2 - \frac{1}{n}\sum_{k=1}^n d_{kj}^2 + \frac{1}{n^2}\sum_{k=1}^n\sum_{\ell=1}^n d_{k\ell}^2\right).$$

2. Describe an algorithm for determining the matrix X whose rows are the x_i.

Solution: We will write the points $x_i^\intercal = (x_i^1, x_i^2, \dots, x_i^d)$ for each $1 \le i \le n$. We will refer to x_i^r as the r'th entry of the i'th point x_i. Since the mean of the points is the origin, it holds that $\sum_{i=1}^n x_i^r = 0$ for every $1 \le r \le d$. Moreover,

$$d_{ij}^2 = \|x_i - x_j\|^2 = \sum_{r=1}^d (x_i^r - x_j^r)^2$$

.

1. Observe that

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^n d_{ik}^2 &= \frac{1}{n}\sum_{k=1}^n \sum_{r=1}^d (x_i^r - x_k^r)^2 \\
&= \frac{1}{n}\sum_{k=1}^n \sum_{r=1}^d \left((x_i^r)^2 + (x_k^r)^2 - 2x_i^r x_k^r\right) \\
&= \left(\frac{1}{n}\sum_{r=1}^d n(x_i^r)^2\right) + \frac{1}{n}\left(\sum_{k=1}^n \sum_{r=1}^d (x_k^r)^2\right) - \frac{2}{n}\sum_{r=1}^d x_i^r \left(\sum_{k=1}^n x_k^r\right) \\
&= \sum_{r=1}^d (x_i^r)^2 + \frac{1}{n}\sum_{k=1}^n \sum_{r=1}^d (x_k^r)^2.
\end{aligned}$$

Similarly we have that

$$\frac{1}{n}\sum_{k=1}^n d_{kj}^2 = \sum_{r=1}^d (x_j^r)^2 + \frac{1}{n}\sum_{k=1}^n \sum_{r=1}^d (x_k^r)^2.$$

Finally we see that

$$\begin{aligned}
\frac{1}{n^2}\sum_{k=1}^n \sum_{\ell=1}^n d_{k\ell}^2 &= \frac{1}{n^2}\sum_{k=1}^n \sum_{\ell=1}^n \sum_{r=1}^d (x_k^r - x_\ell^r)^2 \\
&= \frac{1}{n^2}\sum_{k=1}^n \sum_{\ell=1}^n \sum_{r=1}^d (x_k^r)^2 + (x_\ell^r)^2 - 2x_k^r x_\ell^r \\
&= \frac{1}{n}\sum_{r=1}^d \sum_{k=1} (x_k^r)^2 + \frac{1}{n}\sum_{r=1}^d \sum_{\ell=1} (x_\ell^r)^2.
\end{aligned}$$

Putting everything together we see that

$$\begin{aligned}
& d_{ij}^2 - \frac{1}{n}\sum_{k=1}^n d_{ik}^2 - \frac{1}{n}\sum_{k=1}^n d_{kj}^2 + \frac{1}{n^2}\sum_{k=1}^n \sum_{\ell=1}^n d_{k\ell}^2 \\
&= \sum_{r=1}^d (x_i^r - x_j^r)^2 \\
&- \sum_{r=1}^d (x_i^r)^2 - \frac{1}{n}\sum_{k=1}^n \sum_{r=1}^d (x_k^r)^2 \\
&- \sum_{r=1}^d (x_j^r)^2 - \frac{1}{n}\sum_{k=1}^n \sum_{r=1}^d (x_k^r)^2 \\
&+ \frac{1}{n}\sum_{r=1}^d \sum_{k=1} (x_k^r)^2 + \frac{1}{n}\sum_{r=1}^d \sum_{\ell=1} (x_\ell^r)^2 \\
&= -2\sum_{r=1}^d x_i^r x_j^r \\
&= -2x_i^\intercal x_j.
\end{aligned}$$

Dividing by -2 gives our desired result.

2. Let $A = XX^\intercal$. We can compute the singular value decomposition for A and write it as

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^\intercal. \tag{1}$$

Taking the transpose in (1) and using the fact that A is symmetric, we see that

$$A = \sum_{i=1}^{n} \sigma_i v_i u_i^\intercal.$$

Therefore A can be written as $A = U\Sigma U^\intercal$, where Σ is the diagonal matrix with σ_i in the i'th row and matrix U has columns precisely the singular vectors of A. Take $X = U\Sigma^{1/2}$ to conclude the result.