Singular Value Decomposition

Algoritmos Para Data Science - Eduardo Laber

Daniel Menezes, Guilherme Varela, Matheus Telles

23 de outubro de 2017

Question 2 part 1

Provar

$$\mathbf{x}_i\mathbf{x}_j^T = -\frac{1}{2}\left[d_{ij}^2 - \frac{1}{n}\sum_i^n d_{ij}^2 - \frac{1}{n}\sum_j^n d_{ij}^2 + \frac{1}{n^2}\sum_i^n\sum_j^n d_{ij}^2\right] \qquad (1)$$

$$d_{ij}^2 = \left[\sum_k^d (x_{ik} - x_{jk})^2\right] \Leftrightarrow \mathbf{x}_i\cdot\mathbf{x}_i + \mathbf{x}_j\cdot\mathbf{x}_j - 2\mathbf{x}_i\cdot\mathbf{x}_j \qquad (2)$$

$$\begin{aligned}\frac{1}{n}\sum_i^n d_{ij}^2 &= \frac{1}{n}\sum_i^n(\mathbf{x}_i\cdot\mathbf{x}_i + \mathbf{x}_j\cdot\mathbf{x}_j - 2\mathbf{x}_i\cdot\mathbf{x}_j)\\ &= \frac{1}{n}\left(\sum_i^n \mathbf{x}_i\cdot\mathbf{x}_i + n\mathbf{x}_j\cdot\mathbf{x}_j - 2\mathbf{x}_j\sum_i^n \mathbf{x}_i\right)\\ &= \mathbf{x}_j\cdot\mathbf{x}_j + \frac{1}{n}\left(\sum_i^n \mathbf{x}_i\cdot\mathbf{x}_i\right)\\ &= \mathbf{x}_j\cdot\mathbf{x}_j + MSQ \qquad (3)\end{aligned}$$

Question 2 part 1

$$\frac{1}{n}\sum_{j}^{n} d_{ij}^2 = \mathbf{x}_i \cdot \mathbf{x}_i + MSQ \tag{4}$$

$$\begin{aligned}\frac{1}{n^2}\sum_{i}^{n}\sum_{j}^{n} d_{ij}^2 &= \frac{1}{n^2}\sum_{i}^{n}\left(n\mathbf{x}_i \cdot \mathbf{x}_i + SSQ\right) \\ &= \frac{1}{n^2}\left(nSSQ + nSSQ\right) \\ &= 2MSQ \end{aligned} \tag{5}$$

Substituindo 3, 4, 5 em 2 completamos a prova

Question 2 part 2

$$\begin{aligned} X &= U\Sigma V^T \\ XX^T &= (U\Sigma V^T)(U\Sigma V^T)^T = (U\Sigma V^T)(V\Sigma^T U) \\ &= (U\Sigma\Sigma^T U) \\ X &= U\Sigma \end{aligned} \tag{6}$$

Pegando os dois primeiros vetores colunas de tamanho n de 6 obtemos a melhor representação da matrix D em duas dimensões.