EECS 551 Discussion 8

Task 4 - Multidimensional Scaling

Eric Cheek

October 28, 2021

Overview

Today's discussion will cover:

- Chapter 6 Highlights
- Task 4 - Multidimensional Scaling

Chapter 6 Highlights: Low Rank Matrix Approximation via Frobenius Norm

Early on in Chapter 6, we saw that we can solve the following optimizaiton problem

$$\hat{\mathbf{A}}_K = \underset{\mathbf{B} \in \mathcal{L}_K^{MxN}}{\operatorname{argmin}} \|\mathbf{B} - \mathbf{A}\|_F$$

$$\mathcal{L}_K^{MxN} := \{\mathbf{B} \in \mathbb{F}^{MxN} : (\mathbf{B}) \leq K\}$$

Has a solution

$$\hat{\mathbf{A}}_K = \sum_{k=1}^{K} \sigma_k \mathbf{u_k} \mathbf{v_k'}$$

Chapter 6 Highlights: Low Rank Matrix Approximation via Frobenius Norm

Any questions on the proof for the general case?

Diagonal case: proof sketch

First consider a $M \times N$ (rectangular) diagonal matrix $\boldsymbol{\Sigma}$ having rank r, with descending diagonal values, where we want to approximate it by a matrix $\boldsymbol{C}$ of rank at most $K \leq r$, *i.e.*, we want to solve:

$$\hat{\boldsymbol{C}} \triangleq \underset{\boldsymbol{C} \in \mathcal{L}_K^{M \times N}}{\arg\min} \left\|\left\| \boldsymbol{C} - \boldsymbol{\Sigma} \right\|\right\|_{\mathrm{F}}^2 = \underset{\boldsymbol{C} \in \mathcal{L}_K^{M \times N}}{\arg\min} \left\|\left\| \begin{bmatrix} c_{11} & \dots & c_{1N} \\ & \vdots & \\ c_{M1} & \dots & c_{MN} \end{bmatrix} - \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_K & & & \\ & & & \sigma_{K+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_r \\ & & & & & & \mathbf{0} \end{bmatrix} \right\|\right\|_{\mathrm{F}}^2$$

$$= \underset{\boldsymbol{C} \in \mathcal{L}_K^{M \times N}}{\arg\min} \sum_{k=1}^{K} (c_{kk} - \sigma_k)^2 + \sum_{k=K+1}^{r} (c_{kk} - \sigma_k)^2 + \sum_{k=r+1}^{\min(M,N)} (c_{kk} - 0)^2 + \sum_{m \neq n} (c_{mn} - 0)^2.$$

Figure: Optimization problem posed from general case for low rank approximation

Chapter 6 Highlights: Low Rank Matrix Approximation via Frobenius Norm

Any questions on the proof for the general case?

Proof for general case

Now we assume that (6-3) is correct (it is, though not proven here), and use it to prove the general case. Denote an SVD of $\boldsymbol{A}$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$. Rewrite any $\boldsymbol{B} \in \mathbb{F}^{M \times N}$ in terms of the $\boldsymbol{U}$ and $\boldsymbol{V}$ bases as follows:

$$\boldsymbol{B} = \underbrace{(\boldsymbol{U}\boldsymbol{U}')}_{\boldsymbol{I}} \boldsymbol{B} \underbrace{(\boldsymbol{V}\boldsymbol{V}')}_{\boldsymbol{I}} = \boldsymbol{U} \underbrace{(\boldsymbol{U}'\boldsymbol{B}\boldsymbol{V})}_{\hookrightarrow \triangleq \boldsymbol{C} \text{ (not diagonal in general)}} \boldsymbol{V}' = \boldsymbol{U}\boldsymbol{C}\boldsymbol{V}'. \tag{6-4}$$

Because $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary, $\mathsf{rank}(\boldsymbol{B}) = \mathsf{rank}(\boldsymbol{C})$, so (6-1) is equivalent to

$$\hat{\boldsymbol{A}}_K = \boldsymbol{U}\hat{\boldsymbol{C}}\boldsymbol{V}', \quad \hat{\boldsymbol{C}} \triangleq \operatorname*{arg\,min}_{\boldsymbol{C} \in \mathcal{L}_K^{M \times N}} \left\| \underbrace{\boldsymbol{U}\boldsymbol{C}\boldsymbol{V}'}_{\boldsymbol{B}} - \underbrace{\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'}_{\boldsymbol{A}} \right\|_{\mathrm{F}} = \operatorname*{arg\,min}_{\boldsymbol{C} \in \mathcal{L}_K^{M \times N}} \|\boldsymbol{U}(\boldsymbol{C} - \boldsymbol{\Sigma})\boldsymbol{V}'\|_{\mathrm{F}}$$

$$= \operatorname*{arg\,min}_{\boldsymbol{C} \in \mathcal{L}_K^{M \times N}} \|\boldsymbol{C} - \boldsymbol{\Sigma}\|_{\mathrm{F}} = \sum_{k=1}^{K} \sigma_k \boldsymbol{e}_k \tilde{\boldsymbol{e}}_k',$$

Figure: Optimization problem posed from general case for low rank approximation

Chapter 6 Highlights: Eckhart - Young - Mirsky Theorem

We also learned for any unitarily invariant norm, we can state:

Theorem (**Eckart-Young-Mirsky**) (See [4] and [10] for a proof.)
For any **unitarily invariant** matrix norm $\|\cdot\|_{\mathrm{UI}}$, the **low-rank approximation problem** has the same solution using the first (largest) K singular components of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k'$:

$$\hat{\boldsymbol{A}}_K \triangleq \underset{\boldsymbol{B} \in \mathcal{L}_K^{M \times N}}{\arg\min} \underbrace{\|\boldsymbol{B} - \boldsymbol{A}\|_{\mathrm{UI}}}_{\hookrightarrow \text{ unitarily invariant norm}} = \sum_{k=1}^{K} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k'.$$

Figure: Solution for any optimizaiton problem involving a unitarily invariant norm

Example!

Consider the vectors $\mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 & -2 \end{bmatrix}$ the matrix $\mathbf{A} = \mathbf{x}\mathbf{x}' + \mathbf{y}\mathbf{y}'$.
What is the solution to the problem

$$\hat{\mathbf{A}}_K = \underset{\mathbf{B} \in \mathcal{L}_K^{2x2}}{\operatorname{argmin}} \|\mathbf{B} - \mathbf{A}\|_F$$

When $K = 1$?
What is the error for $k > 1$?

Example!

Singular Value Thresholding

Consider the optimization problem:

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\|_{UI}^2 + \beta R(\mathbf{X})$$

Based on the symmetric gauge principles,

$$\hat{\mathbf{X}} = \mathbf{U_r}\hat{\mathbf{\Sigma}}_\mathbf{r}\mathbf{V}'_\mathbf{r},\ \hat{\mathbf{\Sigma}}_\mathbf{r} = Diag\{\hat{w}_k\}, \hat{w}_k = h_k(\sigma_k;\beta)$$

$R(\cdot) = rank(\cdot),\ h_k(\sigma_k;\beta) = h_{hard}(\sigma_k;\beta) = \sigma_k H(\sigma_k - \sqrt{2\beta})$
$R(\cdot) = \|\cdot\|_*, h_k(\sigma_k;\beta) = h_soft(\sigma_k;\beta) = [\sigma_k - \beta]_+$

Example!

Consider the following problem (Exam 3, Fall 2020)

3. Let $\boldsymbol{Y}$ be a 6×7 matrix with singular values $0, 2, 3, 8, 9, 13$ and define

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X} \in \mathbb{R}^{6\times 7}}{\arg\min}\, \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}^2 + 8\|\boldsymbol{X}\|_*$$

$$\hat{\boldsymbol{Z}} = \underset{\boldsymbol{Z} \in \mathbb{R}^{6\times 7}}{\arg\min}\, \frac{1}{2}\|\boldsymbol{Z} - \boldsymbol{Y}\|_{\mathrm{F}}^2 + 8\,\mathsf{rank}(\boldsymbol{Z})$$

Determine both $\hat{\mathbf{X}}$ and $\hat{\mathbf{Z}}$

Example!

Example!

Multi-Dimensional Scaling

Class, I need your help...

I misplaced all of my sensors again. All I have is this matrix $\mathbf{D} \in \mathbb{R}^{dxN}$, $d_{ij} = \|\mathbf{c}_i - \mathbf{c}_j\|_2$

Luckily, with the help of the 551 lecture notes, I might be able to locate my sensors!

Multi-Dimensional Scaling

First, let's define another matrix $\mathbf{S}$ such that $s_{ij} = d_{ij}^2$
Following the derivation in the notes, we can write

$$\mathbf{S} = \mathbf{r}1_{\mathbf{J}}' + 1_{\mathbf{J}}\mathbf{r}' - 2\mathbf{C}'\mathbf{C},\ \mathbf{r}_i = \|\mathbf{c}_i\|_2^2$$

We can next "de-mean" the data by multiplying by $\mathbf{P}^{\perp} = \mathbf{I} - \frac{1}{j}1_{\mathbf{J}}1_{\mathbf{J}}'$ on the left and right.

$$\mathbf{P}^{\perp}\mathbf{S}\mathbf{P}^{\perp} = -2\mathbf{C}'\mathbf{C}$$

Multi-Dimensional Scaling

We then can define a matrix $\mathbf{G}$ solely in terms of $\mathbf{C}$!

$$\mathbf{G} = \mathbf{C}'\mathbf{C} = -\frac{1}{2}\mathbf{P}^{\perp}\mathbf{S}\mathbf{P}^{\perp}$$

From here, we can see that $\mathbf{G}$ is PSD. Using the SVD of $\mathbf{G}$, we can find an expression for $\mathbf{C}$.

$$\mathbf{G} = \mathbf{V_d}\mathbf{\Sigma_d}\mathbf{V_d'},\ \mathbf{C} = \mathbf{\Sigma_d}^{1/2}\mathbf{V_d}$$