4301 - HW4 Solution

November 29, 2020

Problem 1: Missing Entries

Suppose that you are given a symmetric matrix $A \in \mathbb{R}^{n\times n}$ that is missing some entries, e.g., $A_{i,j} =?$ for some indices $i, j \in \{1, \dots, n\}$. To determine which entries are missing, we will use an index matrix $Q \in \{0,1\}^{n\times n}$ such that $Q_{i,j} = 1$ if $A_{i,j} =?$ and $Q_{i,j} = 0$ otherwise.

1. Explain how to formulate the problem of finding the closest symmetric positive semidefinite matrix to A under the Frobenius norm (over the non-missing entries) as a convex optimization problem.

   Answer: : **(Watch Lectures 10/14 and 10/19 and study the 9th set of slides.)**

   Projecting a matrix into the convex set of positive semidefinite matrices has been discussed in the class, and you can find its related materials in the 9th set of slides, page 26- 37. Here, the only difference is those missing entries. So, with a slight modification, remove those particular entries from the Frobenius norm's summation and build the following problem.

   $$\min_{B\in\mathbb{R}^{n\times n}_{\text{sym}}} \frac{1}{2} \sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} (A_{i,j} - B_{i,j})^2$$

   such that

   $$B \succeq 0.$$

2. What is the dual of your optimization problem?

   Answer:

   Again, the main reference here is Pages 26-37 of the 9th set of slides discussion. But the point is, adding matrix B and scaler $\sum_{i,j,Q_{i,j}=0}(A_{i,j} - B_{i,j})$ is not allowed. So, the conventional Lagrangian method that we are used to can not be followed. There is a need to hire an alternative notion for constructing the Largangian function. So, first introduce the Lagrangian coefficient matrix Λ, and then subtract $\langle\Lambda, B\rangle$ from the objective function and define the Lagrangian function as

   $$L(\Lambda, B) = \frac{1}{2} \sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} (A_{i,j} - B_{i,j})^2 - \sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} \Lambda_{i,j} B_{i,j}.$$

   As always, in order to achieve the dual function, we need to take the derivative of the Lagrangian function with respect to the primal variable, and put it equal to zero. So for $a, b \in \{1, \dots, n\}$

   $$\frac{\partial L(\Lambda, B)}{\partial B_{a,b}} = -(A_{a,b} - B_{a,b}) - \Lambda_{a,b} = 0 \rightarrow B_{a,b} = A_{a,b} + \Lambda_{a,b}$$

Therefore,

$$\begin{aligned}
g(\Lambda) &= \inf_B \left[L(\Lambda, B)\right] \\
&= \inf_B \left[\frac{1}{2}\sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} (A_{i,j}-B_{i,j})^2 - \sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} \Lambda_{i,j}B_{i,j}\right] \\
&= -\frac{1}{2}\sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} (\Lambda_{i,j})^2 - \sum_{\substack{i,j\in\{1,\dots,n\}\\ Q_{i,j}=0}} \Lambda_{i,j}A_{i,j}
\end{aligned}$$

such that

$$\Lambda \succeq 0.$$

The main difference in comparison with the case that has been discussed in the Slides is that we removed the missing entries from the Frobenius norm and $\langle \Lambda, B\rangle$.

3. Colab

Problem 2: Matrix Factorizations

1. Consider the following convex function, known as the generalized KL divergence, for two nonnegative matrices $A, B \in \mathbb{R}^{m\times n}$.

$$\mathrm{KL}(A\|B) = \sum_{i=1}^{m}\sum_{j=1}^{n}(A_{i,j}\log(\frac{A_{i,j}}{B_{i,j}}) - A_{i,j} + B_{i,j})$$

Suppose, now that $A \in \mathbb{R}^{m\times n}$ is a nonnegative matrix that we would like to approximate as a product of two nonnegative matrices $C \in \mathbb{R}^{m\times K}$, $U \in \mathbb{R}^{K,n}$. Explain how to formulate the problem of finding the closest pair of nonnegative matrices to A under the generalized KL-divergence as a biconvex optimization problem.

Answer: : **(Watch Lecture 11/11 and study the 12th set of slides.)**

KL-divergence can be used as our objective function because it reaches its minimum at $B = A$, which is our desired palace to end up in (Note that KL-divergence is always nonnegative and $\mathrm{KL}(A\|A) = 0$). So, given $B = CU$, define

$$\begin{aligned}
f(C,U) &= \sum_{i=1}^{m}\sum_{j=1}^{n}(A_{i,j}\log(\frac{A_{i,j}}{(CU)_{i,j}}) - A_{i,j} + (CU)_{i,j}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(A_{i,j}\log(\frac{A_{i,j}}{\sum_{k=1}^{K} C_{i,k}U_{k,j}}) - A_{i,j} + \sum_{k=1}^{K} C_{i,k}U_{k,j}).
\end{aligned}$$

Using the function $f(.,.)$, the problem becomes

$$\min_{C\in\mathbb{R}^{m,K}, U\in\mathbb{R}^{K,m}} f(C,U)$$

such that

$$C, U \geq 0.$$

For detailed explanation of how the block coordinate descent works on the biconvex function $f(C,U)$, first check the Lecture 11/11 and slide 12 and then check part 2 of this question on Colab.

2. Colab

3. Is your block coordinate descent procedure guaranteed to converge to a critical point?

   Answer:

   Although we took the derivative of $f(C, U)$ in part 2, this function is not differentiable everywhere. As a result, the block coordinate descent may not converge to a critical point. One example has been provided in Colab.