

CSCI 6220/4030: Homework 6

Assigned Thursday November 29 2018. Due at beginning of class Thursday December 10 2018.

Remember to typeset your submission, and label it with your name. Please start early so you have ample time to see me during office hours. Provide mathematically convincing arguments for the following problems. Ask me if you are unclear whether your arguments are acceptable.

1. Consider the *collaborative filtering* problem: given a set of m items and n users, and observations of a subset of the scores $\{a_{ij}\}$ that the j th user assigns to the i th item, predict the remaining unknown scores.

A popular method for solving this problem is to assume that the $m \times n$ score matrix \mathbf{A} is low-rank. Recall that we say \mathbf{A} has rank r if it can be written in the form $\mathbf{A} = \mathbf{X}\mathbf{Y}$ for some $\mathbf{X} \in \mathbb{R}^{m \times r}$ and $\mathbf{Y} \in \mathbb{R}^{r \times n}$, and such a factorization doesn't exist for any number smaller than r . If \mathbf{A} has rank r , then it only has $O(r(n+m))$ degrees of freedom — the number of entries it takes to specify \mathbf{X} and \mathbf{Y} — rather than the potentially much larger nm degrees of freedom it takes to specify all the entries of \mathbf{A} . Thus, if we assume that \mathbf{A} has a rank r that is much smaller than $\min\{m, n\}$, then we can expect that we can completely determine \mathbf{A} once we have seen a small subset of its entries.

One can show that, if \mathbf{A} indeed has rank r , and it is not too “spiky”, then with high probability, once we have seen $O(rn \log(n))$ observations selected i.i.d. uniformly at random from \mathbf{A} , we can recover all of \mathbf{A} exactly by solving a (surprisingly tractable!) optimization problem¹.

In practical applications (say to predicting item ratings on Amazon), \mathbf{A} is *not* exactly low-rank, but this method still works very well empirically; intuitively, this is because \mathbf{A} is well-approximated by a low-rank matrix.

Verify the intuition that underlies the good performance of these algorithms: show that, given any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, there is a matrix $\mathbf{B} \in \mathbb{R}^{m \times n}$ that has rank $O(\varepsilon^{-2} \log(m+n))$ and is entrywise close to \mathbf{A} . Specifically, show that \mathbf{B} satisfies

$$\|\mathbf{A} - \mathbf{B}\|_{\infty} \leq \varepsilon \max_{i=1, \dots, n} \|\mathbf{a}_i\|_2.$$

Here \mathbf{a}_i is the i th column of \mathbf{A} and $\|\mathbf{A}\|_{\infty} = \max_{i,j} |a_{ij}|$ is the largest absolute entry.

Hint: use the fact that $\mathbf{A} = \mathbf{I}_m \times \mathbf{A}$.

2. Gas molecules move about randomly in a box that is divided into two halves symmetrically by a partition; there is a hole in the partition. Suppose there are n molecules in the box. Molecular motion can be modeled by keeping the molecules in their positions with probability $1/2$, and with probability $1/2$ choosing a number between 1 and n at random and moving the corresponding molecule to the other side of the partition.
 - (a) Argue that the number of molecules on one side of the partition evolves as a Markov chain. What are the states, and the transition probabilities?
 - (b) Show that the chain is ergodic, and find its stationary distribution. Hint: you can assume that detailed balance is satisfied. If you use this hint, the identities

$$\prod_{i=0}^{k-1} \binom{n-i}{i+1} = \binom{n}{k} \quad \text{and} \quad \sum_{i=0}^n \binom{n}{i} = 2^n$$

may be helpful.

¹This observation and the randomized algorithms that arise from it have driven a lot of applied and theoretical work in applied mathematics, statistics, and machine learning over the past decade.

3. Consider a random vector $\mathbf{V} = (x, y)$ that is distributed according to the pdf

$$p(x, y) = c \exp(-(x^2 y^2 + x^2 + y^2 - 8x - 8y)/2),$$

where $c \approx 1/20216.335877$ is the constant required to make this a pdf.

- (a) Show that $x|y$ and $y|x$ are normal random variables, and determine their mean and variances.
- (b) Implement a Gibbs sampler to draw samples from this distribution. Use this sampler to draw 10^4 samples, and plot the last 1000 along with several contour lines of the density p .
- (c) Attach your code² for both of the MCMC samplers above to your homework submission.

²Make your code coherent and easy to follow. I don't care what language you use