

# Matrix Completion

(Apologies in advance if  $\hat{\cdot}$  (LaTeX hat) is rendered poorly in the HTML. For some browsers, the hats may have to be shifted a touch to the left.)

## A Simple Problem

Consider a  $3 \times 2$  matrix  $A$ , where we are unable to observe one of the values:

$$A = \begin{pmatrix} 1 & ? \\ 4 & 8 \\ 3 & 6 \end{pmatrix}.$$

What would be a good guess for the missing value in  $A$ ? This depends on how we define “good”. One objective would be that we want  $A$  to be as “simple” as possible. Again, though, we are left with the question, what do we mean by “simple”? In matrix algebra, one notion of simplicity is rank: “the maximal number of linearly independent columns,” as per Wikipedia. For  $A$ , for the complete rows, we have the following formula:

$$a_{i2} = 2a_{i1}.$$

That being the case, if we were to complete the matrix as,

$$\hat{A} = \begin{pmatrix} 1 & 2 \\ 4 & 8 \\ 3 & 6 \end{pmatrix},$$

we would have a matrix for which the rank of  $\hat{A}$  is 1:

```
hatA <- matrix(c(1,2,4,8,3,6), ncol=2, byrow=T)
require(Matrix)
```

```
## Loading required package: Matrix
```

```
rankMatrix(hatA)[1]
```

```
## [1] 1
```

Suppose that we insert any other real number value  $k$ , to generate  $\tilde{A}(k)$ . This would result in increasing the rank of  $\hat{A}$  to 2:

```
testVals <- sample(c(-10:1, 3:10), 5)
rankTestRes <- rep(NA, length(testVals))
for(i in 1:length(testVals)){
  hatA.alt <- matrix(c(1,testVals[i],4,8,3,6), ncol=2, byrow=T)
  rankTestRes[i] <- rankMatrix(hatA.alt)[1]
}
print(rankTestRes)
```

```
## [1] 2 2 2 2 2
```

As such,  $\hat{A}$  is the unique solution to the problem of completing the matrix  $A$  with a value that minimizes rank.

## Generalizing

This is an example of a matrix completion problem where we defined an objective function as follows:

$$\begin{aligned} & \text{minimize} && \text{rank}(\hat{A}) \\ & \text{s.t.} && a_{ij} = \hat{a}_{ij} \text{ for } (i, j) \in \Omega, \\ & && \text{where } \Omega \text{ is the set of } (i, j) \text{ with values observed.} \end{aligned}$$

As such, the rank is a matrix-level loss function, and our solution must abide by a constraint that all complete rows are replicated. This generalizes to situations where the matrix  $A$  has more columns and rows, of course.

A complication is with larger matrices, computing rank can be difficult. As such, matrix completion methods are generally based on more tractable loss function. One function is based on the “nuclear norm,” which is the sum of singular values of the matrix,  $A$ . There is a close relationship between the nuclear norm and the rank: a matrix of rank  $r$  has  $r$  non-zero singular values. If a matrix has lots of large singular values, it means that it contains many dimensions of orthogonal variation. To get the intuition, think about the relationship between singular values and eigenvalues. Recall that the eigenvalues of a covariance matrix measure the amount of variance explained by the corresponding principal components. For matrix  $A$ , the singular values correspond to the square root of the eigenvalues of  $A'A$ , where this latter matrix product yields eigenvalues that are scalar multiples of the eigenvalues of the covariance matrix of  $A$ .

Now, we can also state a matrix completion loss function as follows:

$$\min_{\hat{A}} \|A - \hat{A}\|_a + \lambda \|\hat{A}\|_b,$$

where the norms  $a$  and  $b$  in this expression immediately above would correspond to our simple example as follows:  $\|\cdot\|_a$  would correspond to a norm that is minimized when  $\hat{A}$  replicates  $A$  for all  $(i, j) \in \Omega$ , such as the Euclidean norm, and then  $\|\cdot\|_b$  would correspond to a norm that reflects (at least for the simple problem) the rank of the matrix, such as the nuclear norm.

A useful guide to matrix completion methods is available here: <http://statweb.stanford.edu/~candes/papers/MatrixCompletion.pdf>

## Application to Causal Inference

The application of these methods to causal inference, as presented by Athey et al. (2018; link below), is pretty straightforward, then. For causal inference, you could want to compare outcomes under treatment, say, a matrix of  $\mathbf{Y}_1$  values, to outcomes under control, a matrix of  $\mathbf{Y}_0$  values. But of course we have the “fundamental problem of causal inference,” and so for units in treatment, their control outcomes in  $\mathbf{Y}_0$  are not observed, and vice versa for units in control. The solution? Do matrix completion for the  $\mathbf{Y}_0$ , in which case you can estimate the ATT simply by taking the difference in means between the observed values in  $\mathbf{Y}_1$  and the imputed values in  $\mathbf{Y}_0$ .

This is, in essence, the approach proposed by Athey et al.: <https://arxiv.org/abs/1710.10251>

Questions or corrections welcome in the comments.