Linear Algebra and Learning from Data

a note from the author

This message and this new textbook are about an established subject—**linear algebra**—leading to the much newer subject of **deep learning**. May I express separate thoughts about those two subjects, and connect them. Two other subjects are essential to success—**statistics** and **optimization**—and the book shows how and where they play a crucial part.

Linear Algebra is completely accepted as basic to the undergraduate curriculum. But I don't see that its surge in importance is fully recognized. *Multivariable algebra is far more widely used than multivariable calculus*. Our students are really missing out if our teaching is limited to matrix manipulation. It is *factorization* of the matrices that we need in applications—into orthogonal and diagonal and triangular matrices.

Deep Learning is a particularly successful application to understanding data. It constructs a learning function $\boldsymbol{F}(\boldsymbol{v}) = \boldsymbol{w}$. The data vectors are $\boldsymbol{v}$, and their meaning is $\boldsymbol{w}$. $\boldsymbol{F}$ is constructed from a training set of known pairs $\boldsymbol{v}$ and $\boldsymbol{w}$. The word **deep** indicates that $\boldsymbol{F}$ is a composition $F_L(\ldots(F_1(\boldsymbol{v})))$ of L simple steps (the "depth" is L). Each step involves a matrix A_i, a vector $\boldsymbol{b}_i$, and a fixed nonlinear activation function: often $F_i = \max(A_i\boldsymbol{v}_{i-1} + \boldsymbol{b}_i, 0)$. The matrices A_i and the vectors $\boldsymbol{b}_i$ are optimized to reproduce $\boldsymbol{F}(\boldsymbol{v}) = \boldsymbol{w}$ on the known training data, leading to good accuracy on the unseen data.

This is the course I now teach: Math 18.065. And students come to it, knowing that the two subjects are important for their future. They learn quite a lot about linear algebra, and they see how optimization finds those matrices A_i in the learning function. Research labs and companies have data to analyze and understand, and this deep learning approach has become widespread. Students learn key ideas from statistics, to measure the success of the learning function $\boldsymbol{F}$.

The course needs an instructor who wants to help. It begins with linear algebra—matrix factorizations $A = QR$ from Gram-Schmidt orthogonalization and $S = Q\Lambda Q^{\mathrm{T}}$ from eigenvalues and $A = U\Sigma V^{\mathrm{T}}$ from singular values. This is the heart of the subject and you could not teach any mathematics that is more useful.

To help instructors and students, the 2018 lectures were recorded for MIT's OpenCourseWare. They will be on **ocw.mit.edu** in mid-April. It's now 2019 and we have the textbook and more experience with the course. I would be happy to send you the new 2019 videos when SIAM sends you a sample copy of the book.

Gilbert Strang

Department of Mathematics, MIT