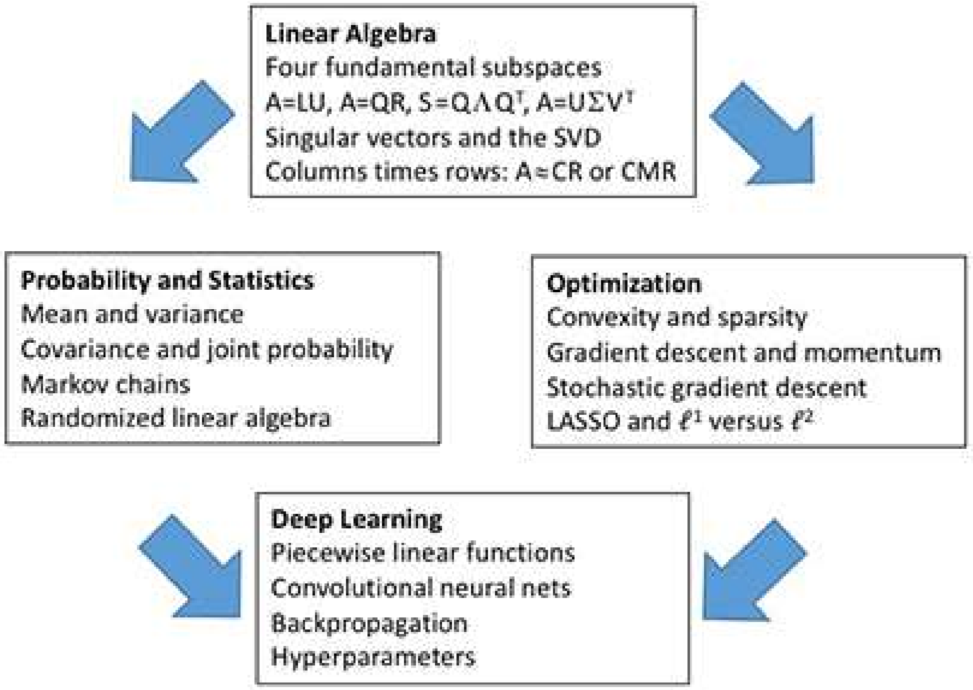
Motivation

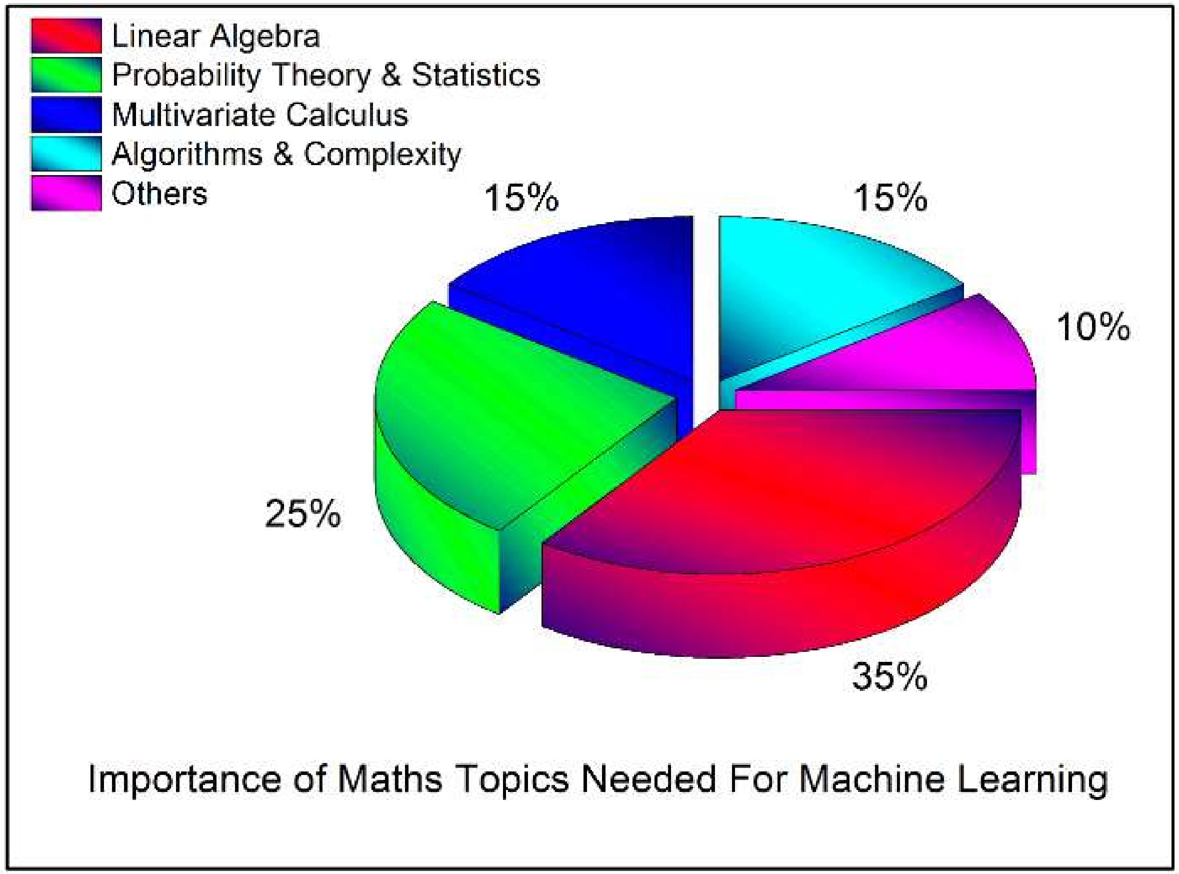
- New master's program in data science?
- Introductory course in mathematical fundamentals for a quantitatively oriented but heterogeneously group of students
  - What data science exactly is
  - What kind of mathematics plays a role
  - What is most important when there is not enough time

Course Description

– Linear algebra concepts are key for understanding and creating machine learning algorithms, especially as applied to deep learning and neural networks.

– This course reviews linear algebra with applications to probability and statistics and optimization–and above all a full explanation of deep learning.





Linear Algebra
Probability Theory & Statistics
Multivariate Calculus
Algorithms & Complexity
Others
15%
15%
10%
25%
35%
Importance of Maths Topics Needed For Machine Learning

Two Essential Topics

- Linear Algebra
  - Symmetric, orthogonal
  - Factorization
- Deep Learning
  - Create a(learning) function
  - Input: image (driverless car: pedestrian, pole, handwriting: zip-code)
  - Speech: Siri

$$\begin{bmatrix} \text{input} \\ (\text{data}) \end{bmatrix} \rightarrow \underbrace{[\text{function}]}_{\substack{\text{matrix multiplication} \\ \text{all linear? fail} \Rightarrow \text{nonlinear}}} \rightarrow [\text{output}]$$

Two Supporting Subjects

- Optimization
  - Find entities in those matrices → learning function
  - Minimizing error → multivariable calculus, giant calculation
- Statistics
  - Keep the mean and variance at the right spot
  - Numbers: good range
    - Bad news for learning function
    - Grow out of sight exponentially
    - Drop to zero

Linear Algebra

- Basic Course
  - Elimination to solve Ax=b
  - Matrix operations, inverses, determinants
  - Vector spaces and subspaces
  - Independence, dimension, rank of a matrix
  - Eigenvalues and eigenvectors
- Stronger Course
  - Ax=b in all cases: square system, too many equations and unknowns
  - Factor A in LU, QR, $U\Sigma V^T$, CMR: columns times rows
  - Four fundamental subspaces: dimensions, orthogonality, good bases
  - Diagonalizing A by eigenvectors and left and right singular vectors
  - Applications: graph, convolution, iteration, covariance, projection, filter, network, image, matrices of data

Deep Learning and Neural Nets

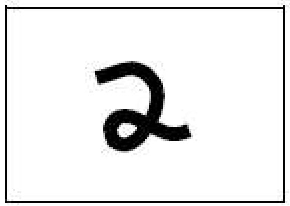
- Mathematical pillars of machine Learning
  - Linear algebra, Probability/Statistics, Optimization
- Goal: to construct a (learning) function that classifies the training data correctly, so it can generalize to unseen test data
- For the problem of identifying handwritten digits
  - Input: image (a matrix of pixels)
  - Output: ten numbers from 0 to 9
  - Function: create by assigning weights to different pixels in the image (architecture of an underlying neural nets)
  - Big problem of optimization to choose weights
  - MNIST set: 70,000 handwritten digits (60,000 for training data, 10,000 for test data)

Linear and Nonlinear Learning Functions

- Inputs: samples v
- Outputs: computed classifications w=F(v)
- Simplest learning function: w=Av
- Affine functions: F(v)=Av+b (b: bias vector), too simple
- Linearity: very limiting requirement
  - halfway between I and III, I and XIX
  - Handwritten digits: two zeros→8, one and zero→9 or 6, images don't add
- Nonlinearity: squaring?
  - sigmoidal functions with S-shaped graphs: A(S(Bv))
  - Simple ramp function: ReLU(x)=max(0,x)

Structure of F(v)

- Functions that yield deep learning
  - Composition of affine functions Lv=Av+b with nonlinear functions R
  - F(v)=L(R(L(R(…(Lv)))))
  - A, b: weights in the learning function
  - Output from first hidden layer: v_1=ReLU(A_1v+b_1)
  - More layers in F → typically more accuracy in F(v)
- Choose weights A_k and b_k to minimize the total loss over all training samples
  - Least squares: $||F(v)-(\text{true output})||^2$

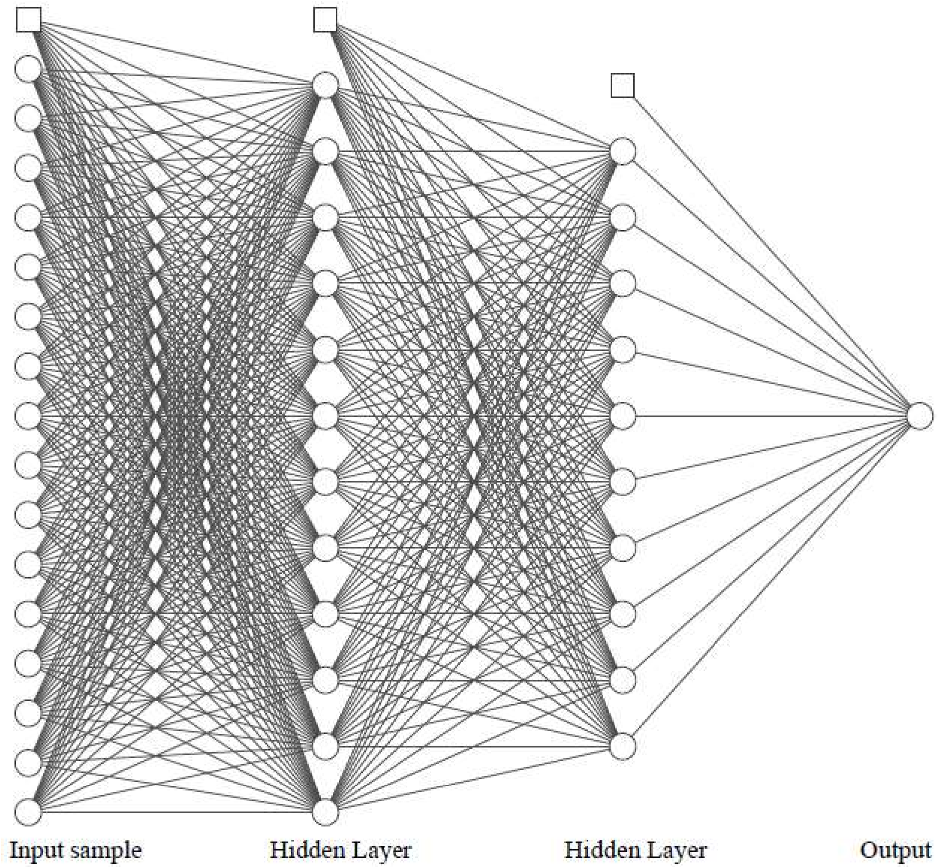
One input $v =$  **One output** $w = 2$

Neural Nets (1)

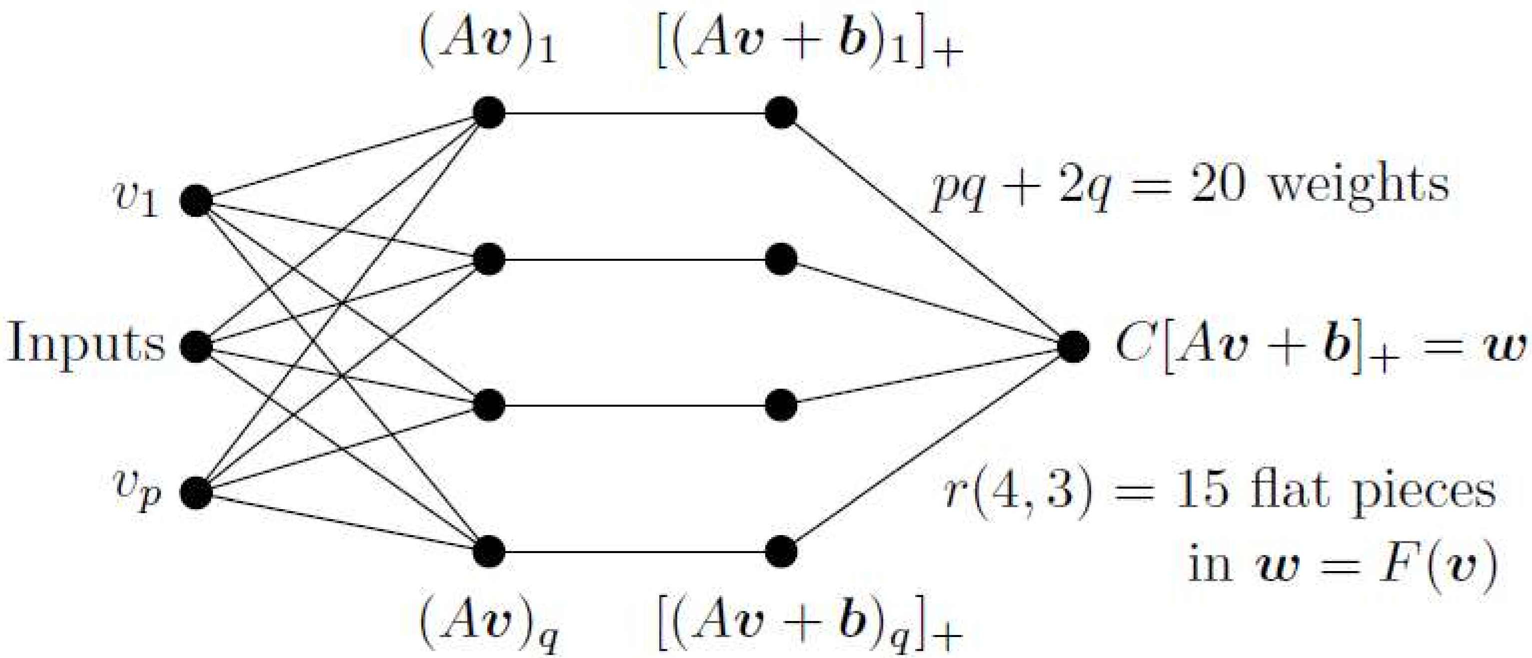
- Input layer: training samples $v=v_0$
- Output: their classification $w=F(v)$
- Hidden layers add depth to the network
  - Allow composite function F
- Feed-forwarded fully connected network
  - For images, convolutional neural net(CNN) is often appropriate and weights are shared, constant diag(A)
- Deep learning works amazing well, when the architecture is right

Neural Nets (2)

- Each diagonal: weight to be learned by optimization
  - Edges from the squares contain bias vector b
  - The other weights are in A



Functions of Deep Learning



– If A is q by p, the input space R^p is sliced by q hyperplanes into r pieces → measure the "expressivity" of the overall function F(v)

$$r(q,p) = \binom{q}{0} + \binom{q}{1} + \cdots + \binom{q}{p}$$

DL: Linear Algebra and Calculus

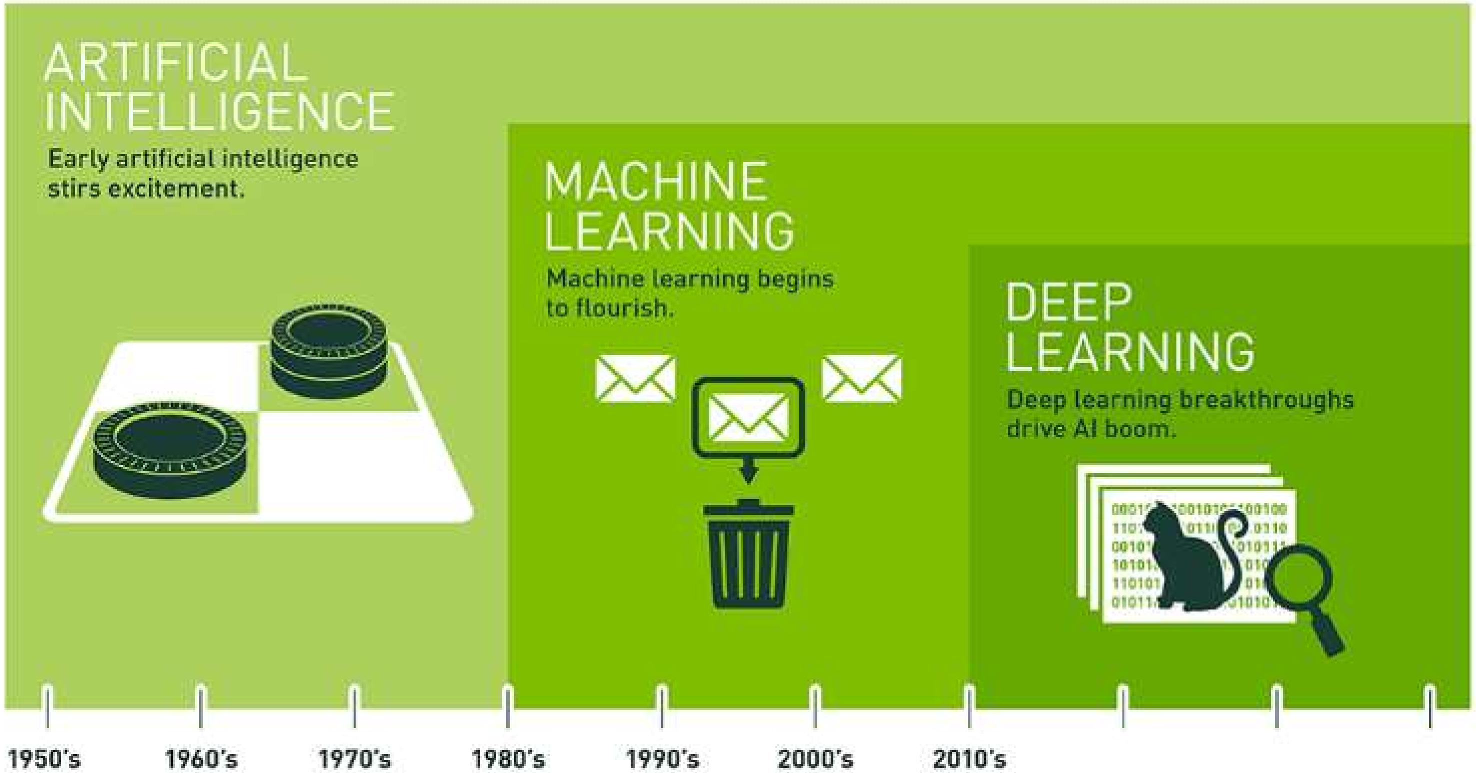
- Linear algebra
  - Compute “weights” that pick out the important features of the training data, and those weights go into matrices
  - Form of learning function
- Calculus
  - Show the direction to move in order to improve the current wrights
  - Reduce the loss function L(x) by moving in the direction of fastest decrease
  - Reduce the error L(x) by $\mathbf{x}_{k+1} = \mathbf{x}_k - s_k \nabla L$

$$\text{minus sign : downhill}$$

$$s_k \text{ : step size (learning rate)}$$

What's the Difference Between AI, ML and DL?

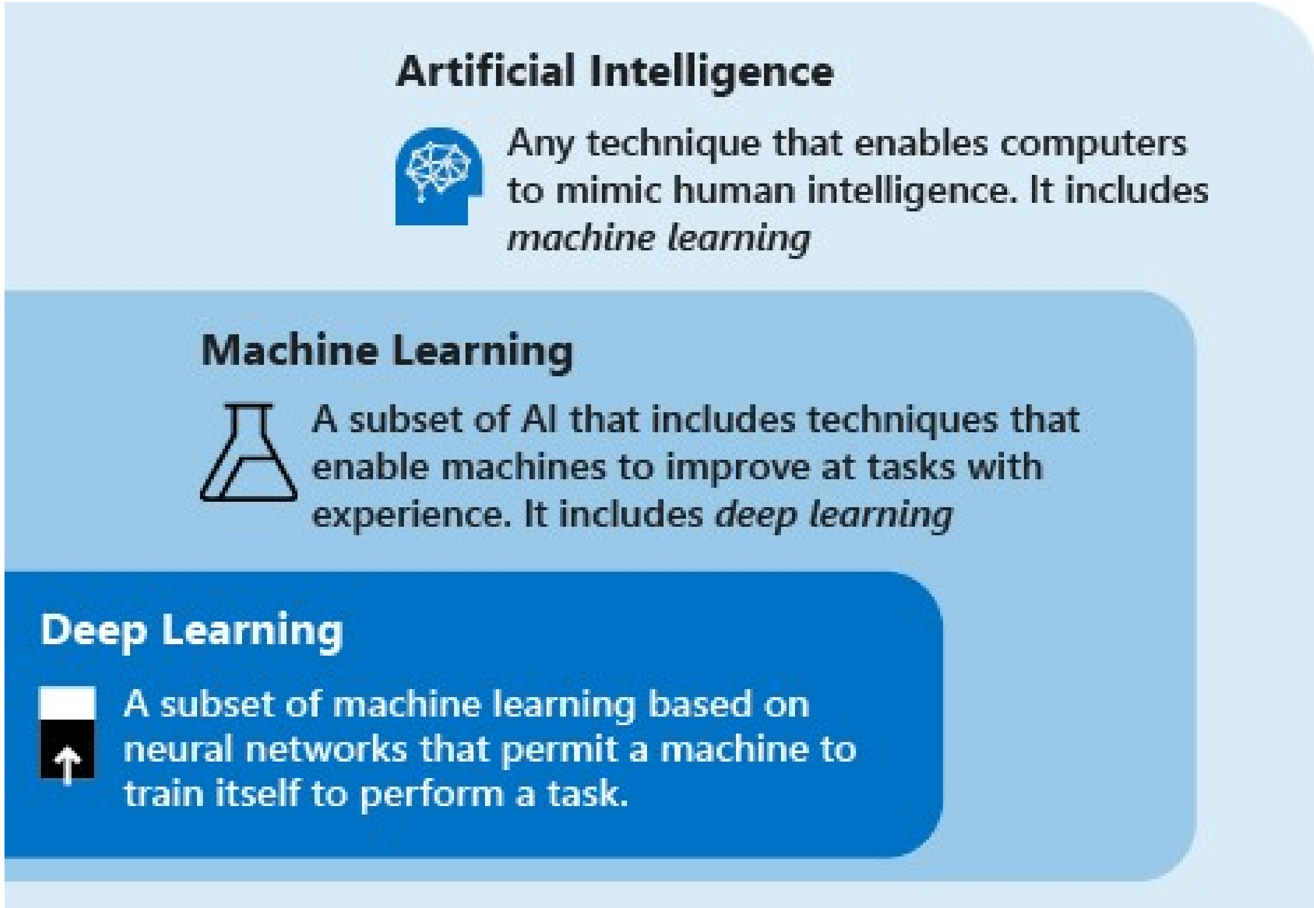
– (blogs.nvidia.com) July 29, 2016 by MICHAEL COPELAND



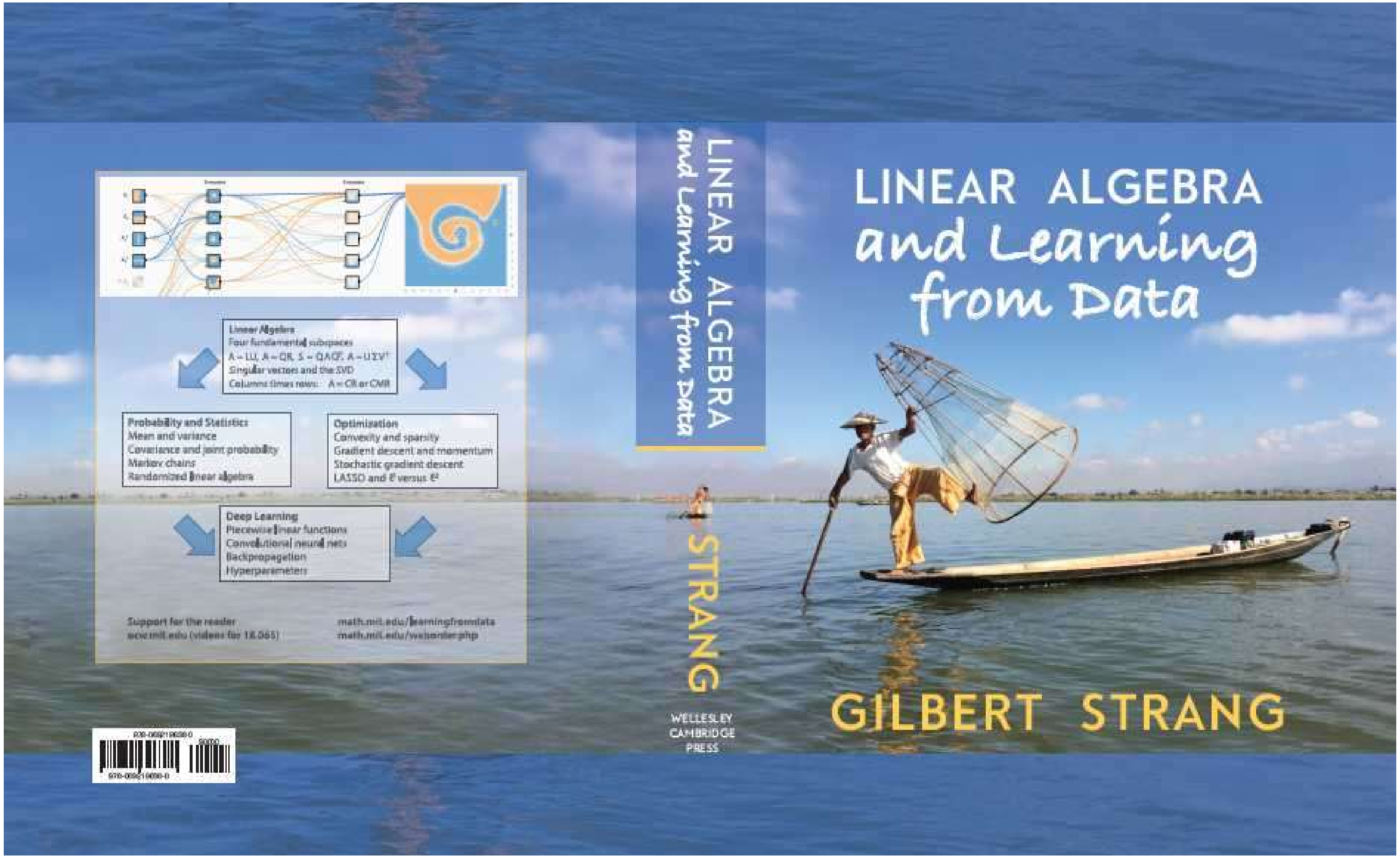
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Deep Learning vs. Machine Learning

- (docs.microsft.com) 03/05/2020



Textbook



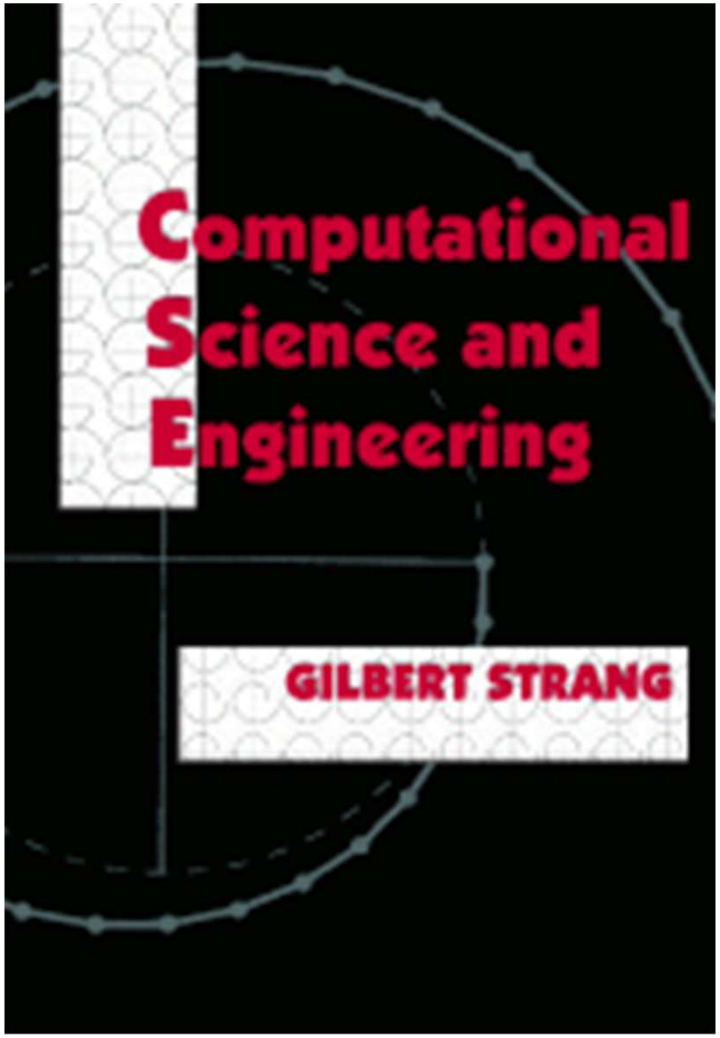
Linear Algebra
Four fundamental subspaces
Singular vectors and the SVD
Probability and Statistics
Mean and variance
Covariance and joint probability
Markov chains
Randomized linear algebra
Optimization
Convexity and sparsity
Gradient descent and momentum
Stochastic gradient descent
Deep Learning
Piecewise linear functions
Convolutional neural nets
Backpropagation
Hyperparameters
Support for the reader
math.mit.edu/learningfromdata
LINEAR ALGEBRA and Learning from Data
STRANG
WELLESLEY CAMBRIDGE PRESS
LINEAR ALGEBRA
and Learning
from Data
GILBERT STRANG

Applied Mathematics for Deep Learning
딥러닝수학

- MIT Math 18.065, Matrix Methods in Data Analysis, Signal Processing, and Machine Learning

Ch	Contents	
1	Highlights of Linear Algebra	○
2	Computations with Large Matrices	○
3	Low Rank and Compressed Sensing	
4	Special Matrices	
5	Probability and Statistics	○
6	Optimization	○
7	Learning from Data	○

AY: 2008~2017



http://math.mit.edu/~gs/cse/

Applied Mathematics for Computational Design and Analysis
전산설계 및 해석을 위한 응용수학

- MIT Math 18.065 Computational Science and Engineering I

Ch	Contents	
1	Applied Linear Algebra	○
2	A Framework for Applied Mathematics	○
3	Boundary Value Problems	○
4	Fourier Series and Integrals	
5	Analytic Functions	
6	Initial Value Problems	
7	Solving Large Systems	
8	Optimization and Minimum Principles	○

Advanced Numerical Methods in Engineering
수치해석특론

Ch	Contents	
1	Applied Linear Algebra	
2	A Framework for Applied Mathematics	
3	Boundary Value Problems	
4	Fourier Series and Integrals	○
5	Analytic Functions	
6	Initial Value Problems	○
7	Solving Large Systems	○
8	Optimization and Minimum Principles	
	Variational Method	○