Oleksandr Romanko, Ph.D.
Associate Director, Financial Risk Quantitative Research, SS&C Algorithmics
Adjunct Professor, University of Toronto

MIE1624H – Introduction to Data Science and Analytics

Lecture 4 – Linear Algebra and Matrix Computations

University of Toronto
October 3, 2023

Lecture outline

Matrix computations

- Matrix operations
- Computing determinants and eigenvalues

Linear algebra

- Solving systems of linear equations
- Solving non-linear equations (Bisection method, Newton’s method)
- Solving systems of non-linear equations
- Solving unconstrained non-linear optimization problems

Derivatives

- Gradients and Hessians
- Taylor series expansion

Functions and convexity

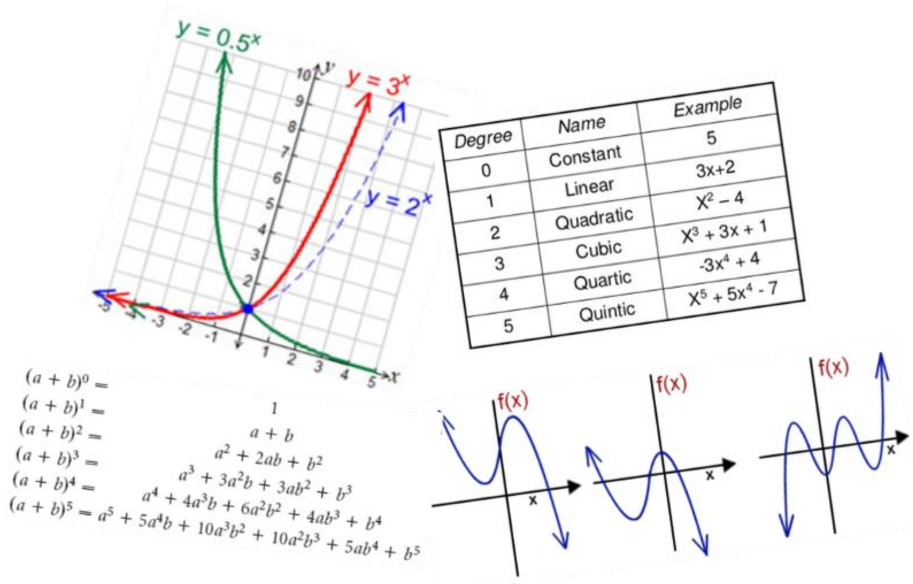
- Convex and concave functions
- Checking convexity
- Properties of convex functions

📖 Math for Data Science

Functions, variables, equations, graphs

What: basic stuff like the equation of a line to binomial theorem and its properties

- Logarithm, exponential, polynomial functions, rational numbers
- Basic geometry and theorems, trigonometric identities
- Real and complex numbers and basic properties
- Series, sums, and inequalities
- Graphing and plotting, Cartesian and polar co-ordinate systems, conic sections



Usage examples: how a search runs faster on a million item database after you sorted it, you will come across the concept of binary search; to understand the dynamics of it, logarithms and recurrence equations need to be understood; if you want to analyze a time series you may come across concepts like periodic functions and exponential decay.

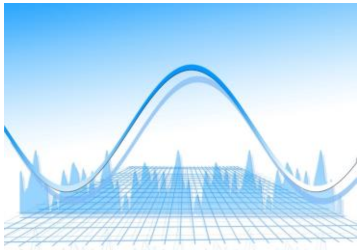
Online resources:

- ❑ Data Science Math Skills – Coursera
- ❑ Introduction to Algebra – edX
- ❑ Khan Academy Algebra

Source: Tirthajyoti Sarkar “Essential Math for Data Science - ‘Why’ and ‘How’”

Statistics

What: solid grasp over essential concepts of statistics and probability, many practitioners in the field call classical (non neural network) machine learning nothing but statistical learning.



- Data summaries and descriptive statistics, central tendency, variance, covariance, correlation
- Basic probability: basic idea, expectation, probability calculus, Bayes theorem, conditional probability
- Probability distribution functions – uniform, normal, binomial, chi-square, student's t-distribution, Central limit theorem (CLT)
- Sampling, measurement, error, random number generation
- Hypothesis testing, A/B testing, confidence intervals, p-values
- ANOVA, t-test, chi-square test
- Linear regression, regularization

Online resources:

- Statistics with R specialization – Coursera
- Statistics and Probability in Data Science using Python – edX
- Business Statistics and Analysis Specialization – Coursera

Usage examples: in interviews, as a prospective data scientist, if you can master all of the concepts mentioned above, you will impress the other side of the table really fast; and you will use some concept or other pretty much every day of your job as data scientist.

Source: Tirthajyoti Sarkar "Essential Math for Data Science - 'Why' and 'How'"

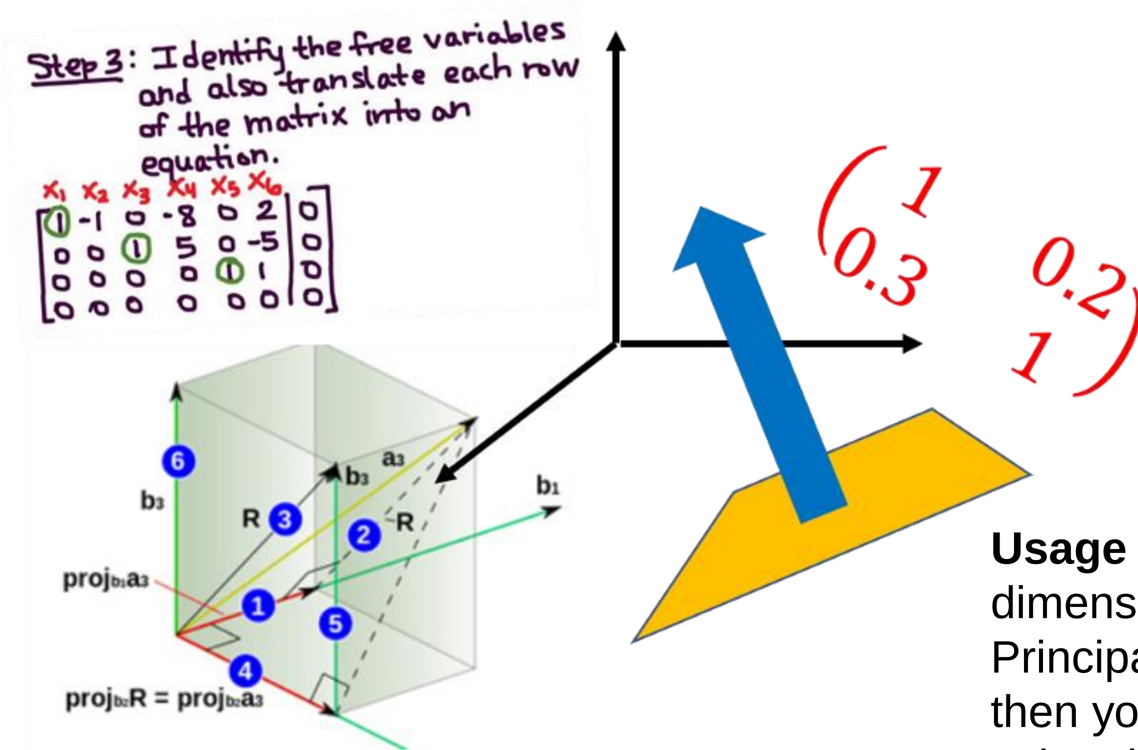
Linear algebra

What: friend suggestion on Facebook, song recommendation in Spotify, transferring your selfie to a portrait drawing Salvador Dali style using Deep Transfer learning – matrices and matrix algebra in all of them; this is an essential branch of mathematics to study for understanding how most machine learning algorithms work on a stream of data to create insight.

- Basic properties of matrices and vectors – scalar multiplication, linear transformation, transpose, conjugate, rank, determinant
- Matrix computations – inner and outer products, matrix multiplication rule and various algorithms, matrix inverse
- Special matrices – square matrix, identity matrix, triangular matrix, idea about sparse and dense matrices, unit vectors, symmetric matrix, Hermitian, skew-Hermitian and unitary matrices
- Matrix factorization concept/LU decomposition, Gaussian/Gauss-Jordan elimination, solving systems of linear equations ($Ax=b$)
- Vector space, basis, span, orthogonality, orthonormality, linear least squares
- Eigenvalues, eigenvectors, and diagonalization, singular value decomposition (SVD)
- Solving systems of nonlinear equations, bisection and Newton algorithms

Source: Tirthajyoti Sarkar “Essential Math for Data Science - ‘Why’ and ‘How’”

Linear algebra (continued)



Usage examples: if you have used a dimensionality reduction technique Principal Component Analysis (PCA), then you have likely used the singular value decomposition to achieve a compact dimension representation of your dataset with fewer parameters, all neural network algorithms use linear algebra techniques to represent and process the network structures and learning operations.

Online resources:

- [Linear Algebra: Foundation to Frontier – edX]
- [Mathematics for Machine Learning: Linear Algebra – Coursera]

Source: Tirthajyoti Sarkar “Essential Math for Data Science - ‘Why’ and ‘How’”

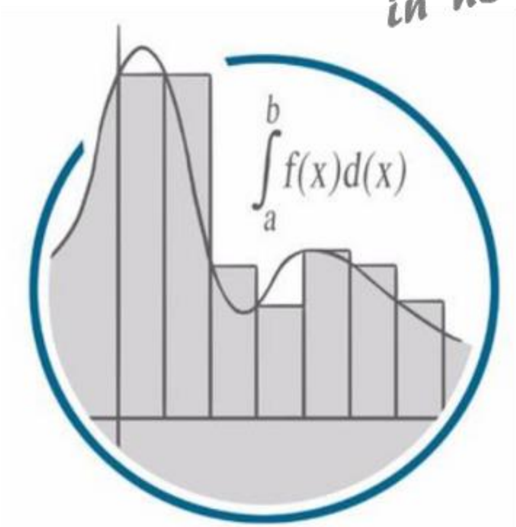
Calculus

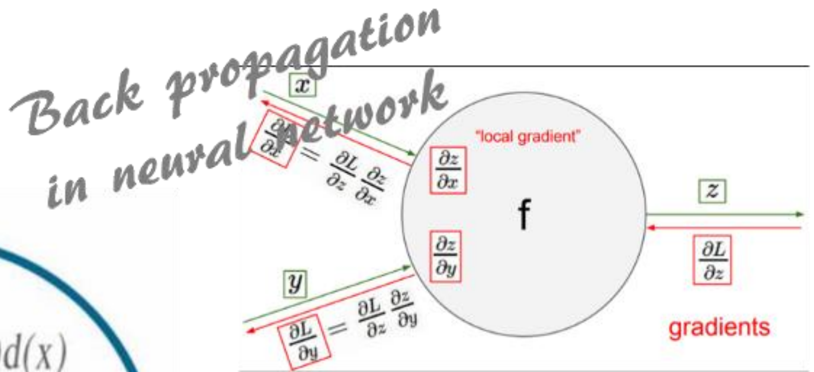
What: concepts and applications of calculus pop-up in numerous places in the field of data science or machine learning; it is behind the simple looking analytical solution of ordinary least square problem in linear regression, or it is embedded in every back-propagation your neural network makes to learn a new pattern.

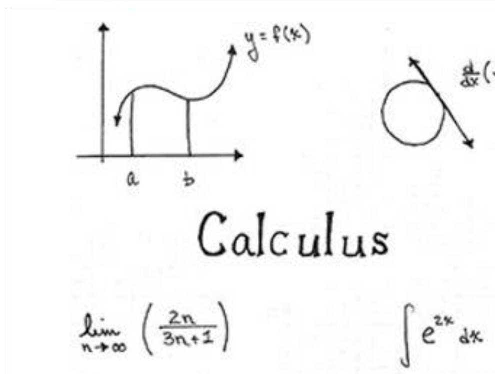
- Functions of single variable, limit, continuity and differentiability
- Mean value theorems, indeterminate forms and L'Hospital rule
- Maxima and minima
- Product and chain rule
- Taylor's series, infinite series summation/integration concepts
- Fundamental and mean value-theorems of integral calculus, evaluation of definite and improper integrals
- Beta and Gamma functions
- Functions of multiple variables, limit, continuity, partial derivatives, gradient vector, Hessian matrix
- Basics of ordinary and partial differential equations (not too advanced)

Source: Tirthajyoti Sarkar "Essential Math for Data Science - 'Why' and 'How'"

Calculus (contunued)







Online resources:

- Pre-University Calculus – edX
- Khan Academy Calculus all content
- Mathematics for Machine Learning: Multivariable Calculus – Coursera

Usage examples: ever wondered how exactly a logistic regression algorithm is implemented, there is a high chance it is using a method called 'gradient descent' to find the minimum loss function, and to understand how it is working, you need to use concepts from calculus –gradient, derivatives, limits, and chain rule.

Source: Tirthajyoti Sarkar "Essential Math for Data Science - 'Why' and 'How'"

Discrete mathematics

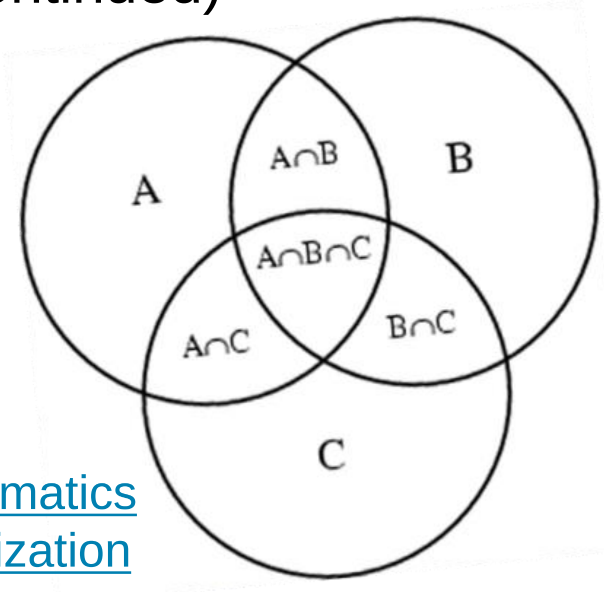
What: all modern data science is done with the help of computational systems and discrete math is at the heart of such systems; a refresher in discrete math will imbue the learner with concepts critical to daily use of algorithms and data structures in analytics project.

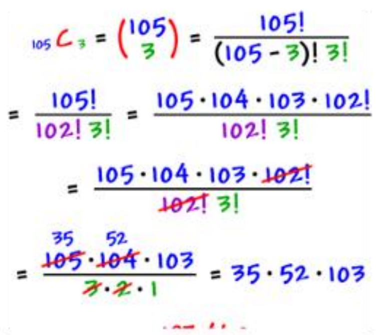
- Sets, subsets, power sets
- Counting functions, combinatorics, countability
- Basic Proof Techniques – induction, proof by contradiction
- Basics of inductive, deductive, and propositional logic
- Basic data structures – stacks, queues, graphs, arrays, hash tables, trees
- Graph properties – connected components, degree, maximum flow/minimum cut concepts, graph coloring
- Recurrence relations and equations
- Growth of functions and O(n) concept

Usage examples: in any social network analysis you need to know properties of graph and fast algorithm to search and traverse the network; to choose an algorithm you need to understand the time and space complexity, i.e., how the running time and space requirements grow with input data size, by using O(n) (Big-Oh) notation.

Source: Tirthajyoti Sarkar "Essential Math for Data Science - 'Why' and 'How'"

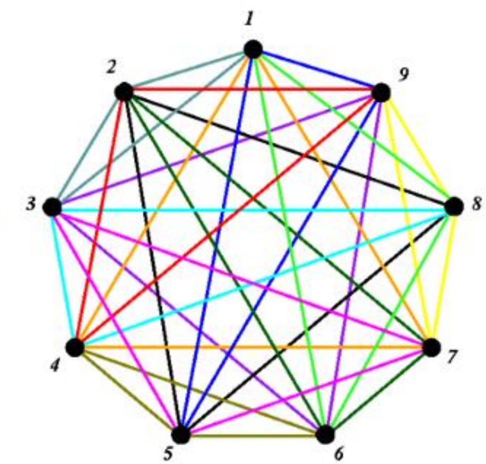
Discrete mathematics (continued)





Online resources:

- [Introduction to Discrete Mathematics for Computer Science Specialization – Coursera]
- [Introduction to Mathematical Thinking – Coursera]
- [Master Discrete Mathematics: Sets, Math Logic, and More – Udemy]



Usage examples: in any social network analysis you need to know properties of graph and fast algorithm to search and traverse the network; to choose an algorithm you need to understand the time and space complexity, i.e., how the running time and space requirements grow with input data size, by using O(n) (Big-Oh) notation.

Source: Tirthajyoti Sarkar “Essential Math for Data Science - ‘Why’ and ‘How’”

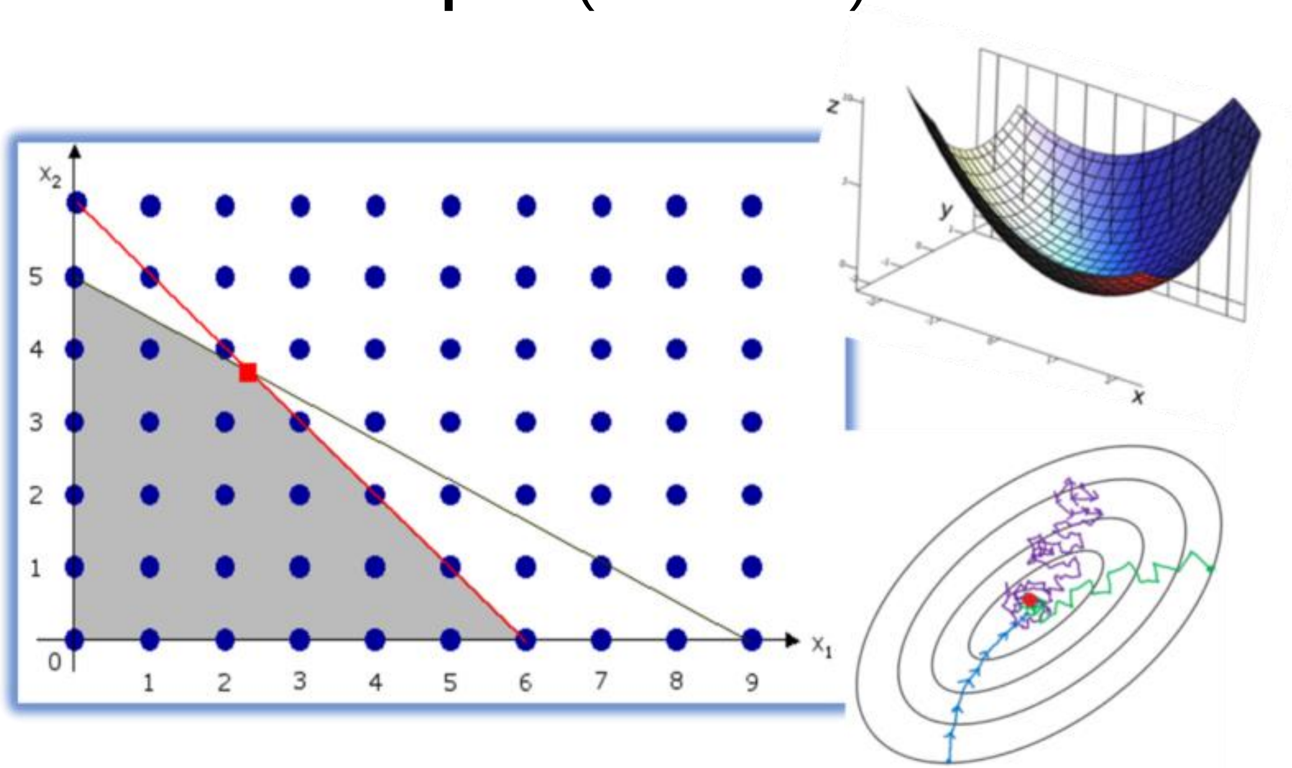
Optimization, operation research topics

What: these topics are little different from the traditional discourse in applied mathematics as they are mostly relevant and most widely used in specialized fields of study – theoretical computer science, control theory, or operation research, however, a basic understanding of these powerful techniques can be immensely fruitful in the practice of machine learning; virtually every (supervised) machine learning algorithm/technique aims to minimize some kind of estimation error subject to various constraints and that is an optimization problem.

- Basics of optimization – how to formulate the problem, unconstrained vs. constrained optimization, nonlinear vs. linear/quadratic optimization
- Maxima, minima, convex functions, local and global optimum
- Linear, quadratic and second-order conic optimization (programming), simplex algorithm, interior-point method (IPM)
- Nonlinear optimization – gradient descent algorithm, Newton and quasi-Newton algorithm, derivative-free optimization
- Integer optimization, mixed-integer optimization
- Constraint programming, Knapsack problem
- Randomized optimization techniques – hill climbing, simulated annealing, Genetic algorithms

Source: Tirthajyoti Sarkar “Essential Math for Data Science - ‘Why’ and ‘How’”

Optimization, operation research topics (continued)



Online resources:

- Optimization Methods in Business Analytics – edX
- Discrete Optimization – Coursera
- Deterministic Optimization – edX

Usage examples: simple linear regression problems using least-square loss function often have an exact analytical solution, but logistic regression problems don't; to understand the reason, you need to know the concept of convexity in optimization; this line of thinking will also explain why we have to remain satisfied with 'approximate' solutions in most machine learning problems.

Source: Tirthajyoti Sarkar "Essential Math for Data Science - 'Why' and 'How'"

🕮 Review of Mathematics for Data Science

Vectors and matrices

- Vectors and matrices are the terms used to describe arrays of data
- Matrices can be arrays of any dimensions, for example, $N \times M$
- Vectors are matrices that have only one row or column, and are typically written as column arrays of dimensions $N \times 1$
- When an array consists of a single number, that is of dimensions 1×1, it is referred to as a scalar
- The element in the i^{th} row and the j^{th} column of the matrix array $\boldsymbol{A}$, for example, is denoted a_{ij}
- Important matrix arrays:
  - nullmatrix, $\boldsymbol{0}$, whose elements are all zeros
  - identity matrix, usually denoted $\boldsymbol{I}$**,** which contains 1's in its left-to-right diagonal, and zeros everywhere else

Matrix computations

- Matrix equality – two matrices are equal only of their dimensions are the same and they have the same elements

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

- Matrix transpose – transpose of an $N \times M$ matrix is an $M \times N$ matrix whose elements are the same as the elements of the original matrix, but are "swapped" around the left-to-right diagonal. The transpose of a matrix $\boldsymbol{A}$ is denoted $\boldsymbol{A}^T$.

$$\boldsymbol{b}^T = \begin{pmatrix} 46 \\ -15 \end{pmatrix}^T = (46 \ \ -15) \qquad \boldsymbol{A}^T = \begin{pmatrix} 3 & 8 \\ 10 & -7 \end{pmatrix}^T = \begin{pmatrix} 3 & 10 \\ 8 & -7 \end{pmatrix}$$

- Matrix addition – when two matrices are added, we simply add the corresponding elements

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix}$$

- Multiplication by a scalar – when a matrix is multiplied by a scalar (a number), the resulting matrix is a matrix whose elements are all multiplied by that number

$$5 \cdot \boldsymbol{A} = 5 \cdot \begin{pmatrix} 3 & 8 \\ 10 & -7 \end{pmatrix} = \begin{pmatrix} 15 & 40 \\ 50 & -35 \end{pmatrix}$$

Matrix computations

- Matrix multiplication – suppose that we want to multiply two matrices, $\boldsymbol{P}$ of dimensions $N \times M$ and $\boldsymbol{Q}$ of dimensions $M \times T$. We have

$$\boldsymbol{P} \cdot \boldsymbol{Q} = \begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NM} \end{pmatrix} \cdot \begin{pmatrix} q_{11} & \cdots & q_{1T} \\ \vdots & \ddots & \vdots \\ q_{M1} & \cdots & q_{MT} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{M} p_{1i} q_{i1} & \cdots & \sum_{i=1}^{M} p_{1i} q_{iT} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{M} p_{Ni} q_{i1} & \cdots & \sum_{i=1}^{M} p_{Ni} q_{iT} \end{pmatrix}$$

Example

$$\boldsymbol{A} \cdot \boldsymbol{x} = \begin{pmatrix} 3 & 8 \\ 10 & -7 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 & 8 \cdot 5 \\ 10 \cdot 2 & -7 \cdot 5 \end{pmatrix} = \begin{pmatrix} 46 \\ -15 \end{pmatrix}$$

- Matrix inverse – inverse of a matrix $\boldsymbol{A}$ (denoted by $\boldsymbol{A}^{-1}$) is simply the matrix that, when multiplied by the original matrix, produces an identity matrix

$$\boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{I}$$

Systems of linear equations

- System of linear equations

$$
\begin{aligned}
3 \cdot x_1 + 8 \cdot x_2 &= 46 \\
10 \cdot x_1 - 7 \cdot x_2 &= -15
\end{aligned}
$$

- To solve this system of equations we express one of the variables through the other from one of the equations and plug into the other equation:

$$x_2 = \frac{46 - 3 \cdot x_1}{8}$$

$$10 \cdot x_1 - 7 \cdot \frac{46 - 3 \cdot x_1}{8} = -1.5$$

- Therefore $x_1 = 2,\ x_2 = 5$
- Matrix notation:

$$\boldsymbol{A} = \begin{pmatrix} 3 & 8 \\ 10 & -7 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 46 \\ -15 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

- System of linear equations in matrix form:

$$\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b}$$

Gaussian elimination

Gaussian elimination

System of equations	Row operations	Augmented matrix
$\begin{aligned} 2x + y - z &= 8 \\ -3x - y + 2z &= -11 \\ -2x + y + 2z &= -3 \end{aligned}$		$\left[\begin{array}{ccc\|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right]$
$\begin{aligned} 2x + y - z &= 8 \\ \frac{1}{2}y + \frac{1}{2}z &= 1 \\ 2y + z &= 5 \end{aligned}$	$\begin{aligned} L_2 + \frac{3}{2}L_1 &\to L_2 \\ L_3 + L_1 &\to L_3 \end{aligned}$	$\left[\begin{array}{ccc\|c} 2 & 1 & -1 & 8 \\ 0 & 1/2 & 1/2 & 1 \\ 0 & 2 & 1 & 5 \end{array}\right]$
$\begin{aligned} 2x + y - z &= 8 \\ \frac{1}{2}y + \frac{1}{2}z &= 1 \\ -z &= 1 \end{aligned}$	$L_3 + -4L_2 \to L_3$	$\left[\begin{array}{ccc\|c} 2 & 1 & -1 & 8 \\ 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right]$
The matrix is now in echelon form (also called triangular form)		
$\begin{aligned} 2x + y &= 7 \\ \frac{1}{2}y &= \frac{3}{2} \\ -z &= 1 \end{aligned}$	$\begin{aligned} L_2 + \frac{1}{2}L_3 &\to L_2 \\ L_1 - L_3 &\to L_1 \end{aligned}$	$\left[\begin{array}{ccc\|c} 2 & 1 & 0 & 7 \\ 0 & 1/2 & 0 & 3/2 \\ 0 & 0 & -1 & 1 \end{array}\right]$
$\begin{aligned} 2x + y &= 7 \\ y &= 3 \\ z &= -1 \end{aligned}$	$\begin{aligned} 2L_2 &\to L_2 \\ -L_3 &\to L_3 \end{aligned}$	$\left[\begin{array}{ccc\|c} 2 & 1 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right]$
$\begin{aligned} x &= 2 \\ y &= 3 \\ z &= -1 \end{aligned}$	$\begin{aligned} L_1 - L_2 &\to L_1 \\ \frac{1}{2}L_1 &\to L_1 \end{aligned}$	$\left[\begin{array}{ccc\|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right]$

Back substitution

Derivatives, gradients, Hessians, convexity

Handout

“**Handout - Review of Linear Algebra, Matrix Computations, Derivatives and Convexity**”

posted with the lecture