Deep learning by Ian goodfellow

- Chapter 2 - Linear Algebra
  - 2.1 - Scalars, Vectors, Matrices and Tensors
    - What is a scalar?

      A scalar is just a single number
    - What is a vector?

      A vector is an array of numbers.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

    - What is a matrix?

      A matrix is a 2-D array of numbers, so each element is identified by two indices instead of just one. If a real-valued matrix A has a height of m and a width of n, then we say thatA ∈ R(m×n).

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}.$$

    - What is a Tensor?

      In some cases we will need an array with more than two axes In the general case, an array of numbers arranged on a regular grid with a variable number of axes is known as a **tensor.**

- What is a transpose of a matrix?

  he transpose of a matrix is the mirror image of the matrix across a diagonal line, called the main diagonal, running down and to the right, starting from its upper left corner.

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^{\top} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

- What is broadcasting?

  In the context of deep learning, we also use some less conventional notation. We allow the addition of a matrix and a vector, yielding another matrix: C = A + b, where Ci,j= Ai,j+bj. In other words, the vector b is added to each row of the matrix. This shorthand eliminates the need to define a matrix with b copied into each row before doing the addition. This implicit copying of b to many locations is called **broadcasting.**

- 2.2 - Multiplying Matrices and Vectors

  - How do we multiply 2 matrices?

    The matrix product of matrices A and B is a third matrixC. In order for this product to be defined, A must have the same number of columns as B has rows. If A is of shape m × n and B is of shape n × p, thenCis of shape m × p.

$$C_{i,j} = \sum_k A_{i,k} B_{k,j} .$$

- How do we define dot product of 2 matrices?

  The dot product between two vectors x and y of the same dimensionality is the matrix product x^(T)y.

  We can think of the matrix productC=AB as computing Ci,j as the dot product between row i of A and column j of B.

- Some properties of matrices

  Matrix Multiplication is distributive:

$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{A}\boldsymbol{B}+\boldsymbol{A}\boldsymbol{C}.$$

  It is also associative:

$$\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C})=(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}.$$

  Matrix multiplication is not commutative:

$$\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A}$$

  does not hold true

  However, the dot product between two vectors is commutative:

$$\boldsymbol{x}^\top\boldsymbol{y}=\boldsymbol{y}^\top\boldsymbol{x}.$$

- 2.3 - Identity and Inverse Matrices
  - What is an Identity matrix?

    An identity matrix is a matrix that does not change any vector when we multiply that vector by that matrix. We denote the identity matrix that preserves n-dimensional vectors as In. Formally, **I**n ∈ Rn×n, and ∀x ∈ Rn, **I**nx = x.

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}.$$

The structure of the identity matrix is simple: all the entries along the main diagonal are 1, while all the other entries are zero.

- How do we calculate inverse of a matrix?

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n.$$

- How do we solve a linear equation using matrix inverse?

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

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

Of course, this process depends on it being possible to find A^(−1).

- 2.4 - Linear Dependence and Span
  - What is linear combination?

    For A−1 to exist, equation 2.11(Ax = b) must have exactly one solution for every value of b. It is also possible for the system of equations to have no solutions or infinitely many solutions for some values of b.

To analyze how many solutions the equation has, think of the columns of A as specifying different directions we can travel in from the origin(the point specified by the vector of all zeros), then determine how many ways there are of reaching b.

$$\boldsymbol{A}\boldsymbol{x} = \sum_i x_i \boldsymbol{A}_{:,i}.$$

In general, this kind of operation is called a linear combination.

- What is a span?

  The span of a set of vectors is the set of all points obtainable by linear combination of the original vectors.

- Define linear independence?

  A set of vectors is linearly independent if no vector in the set is a linear combination of the other vectors.

- Revisit Linear independence!!!

- 2.5 - Norms

  - What is a norm if a vector?

    Sometimes we need to measure the size of a vector.

    Lp norm is give by:

$$||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

Norms, including the Lp norm, are functions mapping vectors to non-negative values.

On an intuitive level, the norm of a vector x measures the distance from the origin to the point x.

The L2 norm, with p= 2, is known as theEuclidean norm, which is simply the Euclidean distance from the origin to the point identified by x.

- What properties should norm satisfy?

  - $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
  - $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$ (the **triangle inequality**)
  - $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

- How can we write dot product of a matrix using norm?

  The dot product of two vectors can be rewritten in terms of norms.

$$\boldsymbol{x}^\top \boldsymbol{y} = ||\boldsymbol{x}||_2 ||\boldsymbol{y}||_2 \cos\theta,$$

  where θ is the angle between x and y

- 2.6 - Special Kinds of Matrices and Vectors

  - What is a diagonal matrix?

    Diagonal matrices consist mostly of zeros and have nonzero entries only along the main diagonal.

    It is possible to construct a rectangular diagonal matrix. Non square diagonal matrices do not have inverses, but we can still multiply by them cheaply.

  - What is a symmetric matrix?

    A symmetric matrix is any matrix that is equal to its own transpose:

$$A = A^\top.$$

- What is a unit vector?

  A unit vector is a vector with unit norm:

$$||x||_2 = 1.$$

- When do we call 2 vectors to be orthogonal?

  A vector x and a vector y are orthogonal to each other if .

$$x^\top y = 0.$$

- When do we call 2 vectors to be orthonormal?

  If the vectors are orthogonal and also have unit norm, we call them orthonormal.

- 2.7- Eigen decomposition
  - What is an eigenvector?

    An eigenvector of a square matrix A is a nonzero vector v such that multi-plication by A alters only the scale of v.

$$Av = \lambda v.$$

  - What is an eigenvalue?

    The scalar λ is known as the eigenvalue corresponding to this eigenvector.

  - What is eigendecomposition of A?

$$\boldsymbol{A} = \boldsymbol{V}\operatorname{diag}(\boldsymbol{\lambda})\boldsymbol{V}^{-1}.$$

▼ 2.8

▼ 2.9 - The Moore-Penrose Pseudoinverse

Matrix inversion is not defined for matrices that are not square. Suppose we wantto make a left-inverse B of a matrix A so that we can solve a linear equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

by left-multiplying each side to obtain

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}.$$

Depending on the structure of the problem, it may not be possible to design a unique mapping from A to B. If A is taller than it is wide, then it is possible for this equation to have no solution. If A is wider than it is tall, then there could be multiple possible solutions.

TheMoore-Penrose pseudoinverseenables us to make some headway in these cases. The pseudoinverse of A is defined as a matrix

$$\boldsymbol{A}^{+} = \lim_{\alpha \searrow 0}(\boldsymbol{A}^{\top}\boldsymbol{A} + \alpha\boldsymbol{I})^{-1}\boldsymbol{A}^{\top}.$$

$$\boldsymbol{A}^{+} = \boldsymbol{V}\boldsymbol{D}^{+}\boldsymbol{U}^{\top},$$

RE-READ THIS

- ▼ 2.10 - The Trace Operator
  - ▼ What is a trace operator?

    The trace operator gives the sum of all the diagonal entries of a matrix:

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i \boldsymbol{A}_{i,i}.$$

  - ▼
- ▼ 2.11 - The Determinant

  The determinant of a square matrix, denoteddet(A), is a function that maps matrices to real scalars. The determinant is equal to the product of all the eigenvalues of the matrix. The absolute value of the determinant can be thought of as a measure of how much multiplication by the matrix expands or contracts space. If the determinant is 0, then space is contracted completely along at least one dimension, causing it to lose all its volume. If the determinant is 1, then the transformation preserves volume.

- ▼ 2.12 - Principal components analysis

  One simple machine learning algorithm, **principal components analysis(PCA)**, can be derived using only knowledge of basic linear algebra.

  Read this algo and make notes!!!