

Metrics on Matrices

1. Norm
 - a. Length / magnitude for vector
2. Dot Product with Self
3. Transpose
4. Inversion
5. Determinant
6. Factorizations and Decompositions
 - a. LU, $A = LU$
 - b. Singular Value Decomposition

Metrics between matrices

1. Angle (Ratio of dot product and product of norms)
2. Distance (Euclidean or otherwise)
3. Orthogonality

Operations on Matrices

1. Addition / Sum
2. Multiplication / Product

Types of Matrices

1. Invertible
2. Singular
 - a. Right Singular
 - b. Left Singular
3. Upper / Lower Triangular
4. Diagonal
5. Orthonormal
 - a. Unitary Matrix
6. Reduced Row Echelon Form / Elimination Matrix / Jordan Matrix
7. Cyclic Matrix
8. Identity Matrix
9. Symmetric
10. Basis Matrix
11. Cofactor Matrix
12. Fourier Matrix
13. Hadamard Matrix
14. Stiffness Matrix
15. Markov Matrix
16. Nullspace Matrix

17. Permutation Matrix
18. Projection Matrix
19. Toeplitz Matrix
20. Eigenvector Matrix
21. Eigenvalue Matrix

Consilience in Linear Algebra
(14 ways to show that a matrix is invertible)

1. A is invertible
2. The columns are independent
3. The rows are independent
4. The determinant is not zero
5. $Ax=0$ has one solution $x=0$
6. $Ax=b$ has one solution $x=A^{-1}b$
7. A has n (nonzero) pivots (after elimination)
8. A has full rank $r=n$
9. The RREF is $R=I$ (identity matrix)
10. The column space is all of $\mathbb{R}^n$
11. The row space is all of $\mathbb{R}^n$
12. All eigenvalues are nonzero
13. $A^T A$ is symmetric positive definite
14. A has n (positive) singular values

Comprehensive Concept List

- 1.1 Vectors and Linear Combinations
 - Vector
 - Linear Combination
- 1.2 Lengths and Dot Products
 - Norm / Length
 - Dot Product / Inner Product
 - Outer Product
 - Unit Vector
 - Angle between vectors
 - Cosine formula
 - Schwarz Inequality
 - Triangle Inequality
- 1.3 Matrices
 - Matrix
 - Matrix Multiplication
 - Independent / Dependent Vectors
- 2.1 Vectors and Linear Equations

- Column view
 - Row view
 - Gaussian Elimination
 - Invertibility
 - Pivot
- 2.2 The Idea of Elimination
 - Upper / Lower Triangular Matrix
- 2.3 Elimination Using Matrices
 - Identity Matrix
 - Elimination Matrices
 - Row exchange matrix
- 2.4 Rules for Matrix Operations
 - Fundamental Law of Matrix Multiplication
 - Block Matrices
 - Block Multiplication
- 2.5 Inverse Matrices
 - Square Matrix
 - Calculating Inverse through Gauss-Jordan Elimination
 - Singular Matrix
 - Diagonally dominant matrices are invertible
 - (each a_{ii} on the diagonal is larger than the sum along the rest of the row i)
- 2.6 Elimination = Factorization: $A = LU$
 - LU Decomposition
- 2.7 Transposes and Permutations
 - Transpose
 - Symmetric Matrix
 - Orthogonal Matrix
 - Permutation Matrix
- 3.1 Spaces of Vectors
 - Vector Space
 - Subspace
 - Column Space
 - Row Space
 - $\mathbb{R}^n$
- 3.2 The Nullspace of A : Solving $Ax = 0$ and $Rx = 0$
 - Nullspace
 - Reduced Row Echelon Form
 - Rank
- 3.3 The Complete Solution to $Ax = b$
 - Full column rank
 - Full row rank
- 3.4 Independence, Basis and Dimension
 - Span

- Basis
- Dimension of a space
- 3.5 Dimensions of the Four Subspaces
 - Four Subspaces
 - Column Space
 - Row Space
 - Nullspace
 - Left nullspace
 - Left nullspace
- 4.1 Orthogonality of the Four Subspaces
 - Orthogonal Complements
 - Nullspace and row space are orthogonal complements
 - Left nullspace and column space are orthogonal complements
- 4.2 Projections
 - Projection
 - Projection Matrix
- 4.3 Least Squares Approximations
 - Least squares
 - Least squares as projection
 - Normal Equation
 - Minimizing the Error (Sum of squared Errors)
 - By geometry
 - By algebra
 - By calculus
 - $\|Ax-b\|^2$
 - Pseudoinverse
 - Fitting by Parabola
- 4.4 Orthonormal Bases and Gram-Schmidt
 - Orthogonal matrix
 - Orthonormal
 - Gram-Schmidt
 - QR Factorization
- 5.1 Determinants
 - Determinant
 - Determinant is zero when a matrix has no inverse.
 - The product of the pivots is the determinant of a 2x2 matrix.
 - Determinant of the identity matrix is 1.
 - Determinant changes sign when two rows are exchanged.
 - When the rows of your matrix are points of a box, the volume of the box is the absolute value of the determinant.
 - $\det(AB) = \det(A)\det(B)$
 - $\det(A^T) = \det(A)$
- 5.2 Permutations and Cofactors

- Determinant of 3 by 3 matrices.
- $\det(A)$ = sum over all $n!$ Column permutations
 - Number of terms involved in computing the determinant is $n!$.
- Determinant by cofactors
- 5.3 Cramer's Rule, Inverses, and Volumes
 - Cramer's Rule
 - Determinant to compute volume of a box / area of a parallelogram
 - Cross Product
- 6.1 Eigenvalues and Eigenvectors
 - Eigenvector
 - Eigenvalue
 - Projections have $\lambda=1$ and 0 . Reflections have 1 and -1 . Rotations have $e^{i\theta}$ and $e^{-i\theta}$.
 - Markov matrix
 - $\det(A - \lambda I) = 0$
 - Characteristic Polynomial
- 6.2 Diagonalizing a Matrix
 - Eigenvalue matrix
 - Diagonalization via n independent eigenvectors
 - Geometric multiplicity
 - Algebraic multiplicity
- 6.3
 - Differential equations
 - Converting constant-coefficient differential equations into linear algebra
 - Stable matrix
 - Matrix exponentials
 - First / Second order equation
 - Difference Equations
- 6.4 Symmetric Matrices
 - A symmetric matrix has real eigenvalues and orthonormal eigenvectors
 - Spectral Theorem:
 - A real symmetric matrix can be diagonalized by its eigenvector matrix times its diagonal eigenvalue matrix times its eigenvector matrix transpose (inverted)
- 6.5 Positive Definite Matrices
 - Positive definite
 - all eigenvalues > 0
 - All pivots > 0
 - All upper left determinants > 0
 - Positive semidefinite
 - Allows for eigenvalue / pivot / determinant $= 0$
- 7.1 The Singular Value Decomposition
 - Singular Vectors

- Eigenvectors of $A^T A$
 - Eigenvectors of $A A^T$
- 7.2 Bases and Matrices in the SVD
 - Singular value
 - Singular value matrix sigma
 - Singular Value Decomposition
 - $A = U \Sigma V^T$
- 7.3 Principal Component Analysis (PCA by SVD)
 - Principal Component Analysis
 - Largest singular value => axis of greatest variance
 - Covariance Matrix
 - Principal components
- 7.4 The Geometry of the SVD
 - $A = U \Sigma V^T$ factors into (rotation)(stretching)(rotation)
 - The geometry of SVD matrix A transforms vectors on a circle into an ellipse.
 - Polar Decomposition
- 8.1 The Idea of a Linear Transformation
 - Linear Transformation
 - Requirements to make a transformation linear
 - Differentiation is a linear transformation
 - Integration is linear transformation (pseudoinverse of differentiation)
- 8.2 The Matrix of a Linear Transformation
 - Change of Basis
 - Choosing the Best Bases (eigen / singular vectors)
- 8.3 The Search for a Good Basis
 - Fourier Matrix
 - Jordan Matrix (The Jordan Form)
 - Circulant Matrix
 - Fourier Basis
 - Legendre Basis
 - Chebyshev Basis
 - Legendre Polynomials
 - Chebyshev Polynomials
- 9.1 Complex Numbers
 - Complex Numbers
 - Imaginary Number
 - Complex conjugate
 - Polar Form
- 9.2 Hermitian and Unitary Matrices
 - Hermitian Matrix (Conjugate transpose)
 - Complex Inner Products
 - Unitary Matrix
- 9.3 The Fast Fourier Transform

- Roots of Unity
- Fourier Transform
 - Represent a function as a sum of harmonics (in frequency space)
- Fast Fourier Transform Algorithm
- 11.1 Gaussian Elimination in Practice
 - Partial Pivoting
 - Sparse Matrices
 - Householder reflections (in QR Factorization)
- 11.2 Norms and Condition Numbers
 - Growth Factor
 - Norm of A is the square root of the maximum eigenvalue of $A^T A$!?
 - Condition Number
 - Solution Error
 - Problem Error
 - Relative Error
- 11.3 Iterative Methods and Preconditioners
 - Replacing A by a simpler matrix S and using an iterative method to solve $Ax=b$
 - Preconditioner
 - Jacobi's method
 - Gauss-Seidel method
 - Incomplete LU method
 - Iteration Matrix
 - Spectral radius
 - Successive Overrelaxation Method (SOR)
 - Multigrid
 - Conjugate Gradients
 - Power Methods
 - Inverse Power Methods
- 12. Mean, Variance and Probability
 - Mean
 - Variance
 - Probabilities
 - Sample Values
 - Sample Mean
 - Sample Variance
 - Expected Value
 - Variance
 - Standard Deviation
 - Probability Distribution
 - Uniform Distribution
 - Normal Distribution
 - Binomial Distribution
 - Cumulative Distribution

- Probability Density Function
- Integration
- Central Limit Theorem
- Monte-Carlo Estimation Methods
- 12.2 Covariance Matrices and Joint Probabilities
 - Covariance (again, in more detail)
 - Joint Probability
 - Correlation
- 12.3 Multivariate Gaussian and Weighted Least Squares
 - Independence
 - Multivariate Gaussian probability distribution
 - Weighted Least Squares
 - Kalman Filter
 - Kalman update
 - Kalman gain matrix

Ch. 1 Introduction to Vectors

Plan:

30m asking what is here, why it matters, where it can be applied.

Ontology:

1. Vectors and Linear Combinations
2. Lengths and Dot Products
3. Matrices

Axioms & Recombination

There are two operations at the heart of linear algebra, both with vectors. Addition and multiplication.

Almost every operation is a function of these operations, whether it be matrix multiplication, singular value decomposition or whatever. Everything can be broken down into vector additions and multiplications.

Notes

The length of a vector is the square root of the dot product. Which is how you quickly compute distances in KNN / Kmeans!! You use matrix multiplication, followed by a square root!

You can think of the product of norms as being equal to the dot products when the vectors go in the same direction.

Questions

1. I wonder - how stochastic is SGD? That is, how far are sampled gradients from the true gradient, on average? Can we say something about the optimal batch size for generalization, using this angle? How well does a notion of angles between matrices hold up? Is it just the average angle between every vector in the matrix? Or should it be represented as a vector of angles?
2. People don't talk about matrix angles. I wonder why?
3. How would we do deep learning without this representation of parameters / outputs? Is there a workable alternate representation?

Ch. 2. Solving Linear Equations

Ontology:

1. Vectors and Linear Equations
2. The Idea of Elimination
3. Elimination Using Matrices
4. Rules for Matrix Operations
5. Inverse Matrices
6. Elimination = Factorization: $A = LU$
7. Transposes and Permutations

Ch. 3. Vector Spaces and Subspaces

Ontology:

1. Spaces of Vectors
2. The Nullspace of A : Solving $Ax=0$ and $Rx=0$
3. The Complete Solution to $Ax=b$
4. Independence, Basis and Dimension
5. Dimensions of the Four Subspaces

Ch. 4. Orthogonality

Ontology:

1. Orthogonality of the Four Subspaces
2. Projections
3. Least Squares Approximations
4. Orthonormal Bases and Gram-Schmidt

Notes

For projection onto a subspace, it looks like you have to take a set of vectors that span that subspace and compute a projection onto each one, and then sum those projections to get your matrix.

There has to be an efficient way to compute this...

Questions

1. I see a lot of value to doing projections in the training processes of algorithms - would be great to have a notebook with a nice example of the projection of one numpy array onto another.
 - a. These are the projections of a matrix onto another matrix. So I guess there's an adaptive range / manifold for each layer? What if it's different at each layer (it probably is)?
2. Can we find a projection matrix that will project a matrix onto our adaptive range?
3. Why does numpy not have a projection function?
 - a. Why do people almost exclusively talk about / look at orthogonal projections?
4. Looks like you need to do an inversion to do an oblique projection.
5. Could use random projections to speed up k-means for high dimensional data, computing distance on a lower dimensional space