Deep Learning

Linear Algebra

* Norm - size of a vector. L^p norm is given by.

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

Norms are functions mapping vectors to non-negative values. A norm is any function that satisfies.

- $f(x) = 0 \Rightarrow x = 0$
- $f(x+y) \leq f(x) + f(y)$
- $\forall \alpha \in R,\ f(\alpha x) = |\alpha|\, f(x)$.

L^2 norm - Euclidean norm (Euclidean distance from origin to x).

squared L^2 norm $= x^T x$.

computationally convenient, because

$\frac{dL^2}{dx}$ = depend on whole vector

$\frac{d \text{ squared } L^2}{dx}$ - depends only on element x.

In ML we would want to discriminate between exactly 0 & non-zero elements. function growing at same rate in all locations. L^1 norm

$$||x||_1 = \sum_i |x_i|.$$

L^∞ - max norm. absolute value of element with largest magnitude

$$||x||_\infty = \max_i |x_i|.$$

size of matrix - Frobenius fro norm

$$||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2}.$$

$$x^T y = ||x||_2 \, ||y||_2 \cos\theta.$$

Symmetric matrix

$$A = A^T.$$

unit vector = unit norm

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

vec orthogonal: vectors x & y are orthogonal if $x^T y = 0$. Orthogonal + unit norm - orthonormal

orthogonal matrix: square matrix whose rows & columns are mutually orthonormal

$$A^T A = A A^T = I.$$

This implies $A^{-1} = A^T$.

Eigendecomposition

decomposing matrix into eigenvectors & eigen values.

eigenvector - It is a non zero vector of square matrix A if multip by A alters only scale of v.

$$Av = \lambda v.$$

λ - eigenvalue corresponding to the eigenvector. $v^T A = \lambda v^T$. - left eigen vectors.

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

eigendecomposition is unique only if all the eigenvalues are unique. The matrix is singular only if any of the eigenvalues are 0.

positive definite - matrix with all positive eigen values. positive semidefinite - +ve or 0.

positive semidefinite $\forall x, x^T A x \geq 0$.

positive definite $x^T A x = 0 \Rightarrow x = 0$.

<u>SVD</u> - decompose into singular values & singular vectors. SVD is more generally applicable.

$$A = U D V^T.$$

A - $m \times n$ matrix U - $m \times m$ matrix

D - $m \times n$ matrix, V - $n \times n$ matrix.

U, V - orthogonal matrices.

D - diagonal matrix, not necessarily square.

elements along diagonal of D - singular values of matrix A. columns of U - left singular vectors. V - right singular vectors.

left singular vectors of A - eigenvectors of AA^T. right singular vectors of A - eigenvectors of A^TA.

SVD - most useful property is that matrix inversion can be generalized to non-square matrices.

---

* Moore - Penrose Pseudoinverse

Matrix inversion is not defined for non square matrices.

$$Ax = y$$

left multiplying $x = By$.

Pseudo-inverse of A is

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

$A^T = V D^T U^T$. D^+ is pseudo-inverse of D taken by taking reciprocal of it's non-zero elements & taking transpose of matrix.

When A has more columns,

$x = A^+ y$ with minimum euclidean norm $||x||_2$ among all possible solutions.

A - more rows than columns, no solution. Here pseudoinverse gives x for which Ax is close to y in terms of Euclidean norm $||Ax - y||_2$.

Trace operator

Gives the sum of all diagonal elements of matrix.

Frobenius form alternative way

$$||A||_F = \sqrt{Tr(AA^T)}$$

Trace operator is invariant to transpose

$$Tr(A) = Tr(A^T)$$

$$Tr(ABC) = Tr(CAB) = Tr(BCA)$$

$$Tr\left(\prod_{i=1}^{n} F^{(i)}\right) = Tr\left(F^{(n)} \prod_{i=1}^{n-1} F^{(i)}\right).$$

invariance to cyclic permutation also holds good if resulting product has different shape.

$$Tr(AB) = Tr(BA)$$

scalar is its own trace $a = Tr(a)$

Principal component analysis

collection of m points - $\{x^{(1)} \dots x^{(m)}\}$

- Apply lossy compression to points.