Mathematics for Machine Learning

Linear Algebra

Linear Algebra
- System of linear equation
  - vector – Obj → moves around space
  - vector space – mapping fn between vector space

Operations – Vector
- Represent
  - Geometric Object – Polar Coordinate
  - List of attributes – in terms of basis vector – Co-ordinate System
- Properties
  - Addition – Associative
  - Scalar Multiplication – Distributive
- Attributes – Size / Length
  - Geometric Obj – length
  - Co-ordinate System – Pythagoras theorem
- Dot Product / Inner Product $r \cdot s = r_i s_i + r_j s_j$ $r \cdot r = |r|^2$
  - Commutative
  - Distributive
- Cosine Rule ① $c^2 = a^2 + b^2 + 2ab\cos\theta$ $\cos\theta = \frac{\text{adjacent}}{\text{hypothesis}}$
  - ② $|r-s|^2 = |r|^2 + |s|^2 - 2r \cdot s$
  - ① + ② → $r \cdot s = ab\cos\theta$
  - $= |r|\,|s|\cos\theta$
  - adjacent → shadow / Projection = $\frac{r \cdot s}{|r|}$
  - Scalar Projection
  - Vector Projection $\frac{r \cdot s}{|r|} \cdot \frac{r}{|r|} = \frac{r \cdot s}{r \cdot r} \cdot r$

Graph Steps
① Fit curve – given parameter
② Goodness of fit ↑ – Parameter space – Cost
③ ↓ Cost wrt parameter

Optimization
- Optimal value of parameter

Fitting equation → Data / solving equation

Reference Frame
- Vector – describe space → Co-ordinate System
- basis vector
  - angle
  - length
  - different
  - independent from co-ordinate system
  - Transform ← dot product
  - Projection → regular & equally spaced grid
- set of n vectors
  - linearly independent → n dimension space
  - span space
  - map / wrap / fold space

Linear Algebra
- value
- notation – mathematical object – vector
- system of manipulating

Matrix – object that operates on vector
Simultaneous equation → matrix

easy – n vectors
- Orthogonal
- Unit length

Matrix Transformation Properties
- $Ar = r'$
- $A(nr) = nr'$
- $A(r+s) = Ar + As$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} r_i \\ r_j \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\left[r_i\begin{bmatrix}1\\0\end{bmatrix} + r_j\begin{bmatrix}0\\1\end{bmatrix}\right] = r_i\begin{bmatrix}a\\c\end{bmatrix} + r_j\begin{bmatrix}b\\d\end{bmatrix}$$

matrix × vector; vector → basis vector; vector transformed to new space defined by matrix

Type
- Identity Matrix $\begin{bmatrix}1&0\\0&1\end{bmatrix}$
- Reflection mirror
  - x-axis $\begin{bmatrix}1&0\\0&-1\end{bmatrix}$
  - y-axis $\begin{bmatrix}-1&0\\0&1\end{bmatrix}$
  - $x=y$ $\begin{bmatrix}0&1\\1&0\end{bmatrix}$
- Rotation $\begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}$
- Stretching $\begin{bmatrix}k&0\\0&1\end{bmatrix}$
- Shearing $\begin{bmatrix}1&k\\0&1\end{bmatrix}$
- Squeezing $\begin{bmatrix}k&0\\0&1/k\end{bmatrix}$

Composition / Combination of Matrix transformation $A_1(A_2 r) = (A_1 A_2) r$ order dependence

Matrix Inverse $A^{-1}A = I$
- Gaussian elimination → Row elimination → Back substitution → triangular Matrix → inverse
- Use Determinant –
  - A (matrix) → stretch original basis vector → to new basis vector space
  - area increase by factor → determinant
  - det = 0 → linearly dependent → singular matrix → can't be inversed

Matrix makes linear mapping

Matrix Product represent
- Einstein Convention
- using
- non-square matrix

$$(ab)_{ik} = \sum_j a_{ij} b_{jk} = a_{ij} b_{jk}$$

Dot Product – column matrix
- $u \cdot v = u^T v$
- matrix product
- $u \cdot v = v \cdot u$
- Geometrically similar △

Changes in basis
- world – Bear / My
- Bear's basis vector in my world (B) × Bear's vector (Bv) = My vector (Mv) → $B^{-1} \times M_v = B_v$
- Bear's Vector → Transform → My Vector
- My Vector → Transform → Bear Vector

Basis Vectors → Orthogonal → Dot product → Projection

My Vector
- Bear's axis in my world = Bear's Vector
  - all the axis → one by one
  - orthogonal vector
  - unit ① vector → normalized by length

Transformation in Bear's world

B^{-1} × transformation × $B \times \begin{pmatrix}x\\y\end{pmatrix}$ (vector in my world)
- ↑ Transformed vector in my world
- Transformed vector in bear's world

Orthogonal Matrix $A = [(a_1)(a_2)\cdots(a_n)]$
- basis set
- unit length + orthogonal to each other → orthonormal matrix
- $a_i \cdot a_j = 0$ $i \neq j$
- $a_i \cdot a_j = 1$ $i = j$
- Transpose $A^T_{ij} = A_{ji} \rightarrow A^T A = I$
- $A^T = A^{-1}$
- $AA^T = I$ → rows are also orthonormal matrix
- factor → area → $|A| = \pm 1$ → flip space
- Projection of vector → Dot product

Convert to orthonormal basis vector
- linearly independent → Determinant
- Gram Schmidth process

$V = \{v_1, v_2, \ldots v_n\}$
- $v_1 \rightarrow u_1 = v_1$, $e_1 = v_1/\|v_1\|$ → unit vector
- $v_2 \rightarrow u_2 = v_2 - (v_2 \cdot e_1)e_1$, $e_2 = u_2/\|u_2\|$
- $v_3 \rightarrow u_3 = v_3 - (v_3 \cdot e_1)e_1 - (v_3 \cdot e_2)e_2$, $e_3 = u_3/\|u_3\|$

orthogonal unit basis set
orthonormal basis set