Machine Learning for Neuroscience

Slides and notebooks: `https://github.com/PBarnaghi/ML4NS`

Mathematical Notations

Symbols and meaning:

$\sum$: sum or sum up; e.g., $\sum_1^n x = x_1 + x_2 ... + x_n$

$\boldsymbol{\frac{d}{dx}}$: Derivative function; $\frac{df(x)}{dx}$ specifies the rate of change for function $f(x)$ with respect to changes in its parameter x

$\boldsymbol{\frac{\partial}{\partial x}}$: Partial derivative; when a function has multiple parameters, a partial derivative specifies the rate of change for that function with respect to one of its parameters while other parameters are kept as constants; e.g., $f(x, y) = 2xy + 3y$ then $\frac{\partial f}{\partial x} = \frac{\partial 2xy+3y}{\partial x} = 2y$

$\boldsymbol{a} \wedge \boldsymbol{b}$: Logical AND

$\boldsymbol{a} \vee \boldsymbol{b}$: Logical OR

$\boldsymbol{a} \neg \boldsymbol{b}$: Logical NOT

$\boldsymbol{\infty}$: Infinity

$\boldsymbol{lim}$: In mathematics limit of function a is the value that a function approaches as its input approaches a vlue; for example $\lim f(x)_{x \to 0}$ means the value of $f(x)$ when x approaches 0. If we assume $f(x) = 2x + 3$ then $\lim f(x)_{x \to 0} = \lim(2x + 3)_{x \to 0} = 3$

$\boldsymbol{\to}$: tends towards to; for example $x \to \infty$

$\boldsymbol{\propto}$: Proportional; for example, if $y = \alpha x$ then $y \propto x$

$\boldsymbol{n}$!: Factorial function; e.g., $n! = n \times (n - 1) \times (n - 2) ... \times 1$;
$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

$\approx$: Approximately equal to

$\triangleq$: Defined as

$\boldsymbol{argmx_x}$*f(x)*: Argmax or the Arguments of Maxima are the value(s) of x that maximises the function *f(x)*

$\binom{\boldsymbol{n}}{\boldsymbol{k}}$: n chooses k which is equal to $n!/k!(n-k)!$

$\boldsymbol{exp(x)}$: Exponential function; e^x

$\boldsymbol{A}$: Matrix A

$\boldsymbol{A}^{-1}$: Inverse of a matrix

$\boldsymbol{A}^T$: Transpose of a matrix; e.g. if $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ then $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

$\boldsymbol{a}$ or $\boldsymbol{x}$: refer to a vector; e.g., $x = [1, 2, 3]$

$\boldsymbol{a}^T$ or $\boldsymbol{x}^T$: refer to transpose of a vector; e.g. $x^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$\boldsymbol{X}_{ij}$: Element (i,j) in matrix X where i is the row number of j is the column number

$\boldsymbol{I}$ or $\boldsymbol{I}_d$: Identity matrix; e.g. A 3×3 identify matrix $\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\|\boldsymbol{x}\|_{\boldsymbol{1}}$: L1 norm of a vector or Manhattan distance $\sum_{j=1}^{d} x_j^2$

$\|\boldsymbol{x}\|_{\boldsymbol{2}}$: L2 norm of a vector or Euclidean distance $\sqrt{\sum_{j=1}^{d} x_j^2}$

$\boldsymbol{x} \otimes \boldsymbol{y}$: Tensor product x and y; e.g. if $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $y = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ then

$$x \otimes y = \begin{bmatrix} 1 \times 4 \\ 1 \times 5 \\ 2 \times 4 \\ 2 \times 5 \\ 3 \times 4 \\ 3 \times 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 8 \\ 10 \\ 12 \\ 15 \end{bmatrix}$$

Notations is probability theory:

$\boldsymbol{X \perp Y}$: means X is independent of Y

$\mathbb{E}[\boldsymbol{X}]$: Expected value of X

$\mathbb{H}[\boldsymbol{X}]$: Expected value of distribution of p(X)

$\boldsymbol{L(y, o)}$ or $\boldsymbol{Loss(y, o)}$: Loss function/value for the output of o when the actual output/value is y

$\boldsymbol{\mu}$: Mean of a scalar distribution

$\boldsymbol{\sigma}$: Standard deviation of a scalar distribution

$\boldsymbol{\sigma}^2$: Variance

$\boldsymbol{\pi}$: In the probability theory it refers to the stationary of a Markov chain

$\boldsymbol{sigm(x)}$: Sigmoid function $sigm(x) = 1/1 - e^{-x}$

Notations in machine learning:

C: Number of Classes or Class set

X: Design matrix or an input dataset to a model

$\mathscr{D}$: Training data

$\mathscr{N}$: Number of samples in a dataset

$\mathscr{N}_c$: Number of samples in Class c

x: Input vector (one sample)

D: Dimentionality of data or the number of features

$\mathscr{D}_{\mathscr{T}}$: Test data

$\boldsymbol{J(\theta)}$: Cost function

k(x,y: Kernel function

K: Kernel matrix

T: Transition matrix of Markov chain

$\boldsymbol{\theta}$: Parameters of a model or parameter vector

$\boldsymbol{W}$: Weight matrix in a regression model or a neural network

End Notes

Mathmatical notations, Machine learning: A Probabilistic Perspective, Kevin Murphy, MIT Press, 2013; pages: 1013-1018, see: `http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf`

Tensor products, https://www.math3ma.com/blog/the-tensor-product-demystified