

## Machine Learning for Neuroscience

Slides and notebooks: <https://github.com/PBarnaghi/ML4NS>

Mathematical Notations

### Symbols and meaning:

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

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

$\frac{\partial}{\partial x}$ : 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  $\frac{\partial f}{\partial x} = \frac{\partial 2xy + 3y}{\partial x} = 2y$

$a \wedge b$ : Logical AND

$a \vee b$ : Logical OR

$a \neg b$ : Logical NOT

$\infty$ : Infinity

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

$\rightarrow$ : tends towards to; for example  $x \rightarrow \infty$

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

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

$\approx$ : Approximately equal to

$\triangleq$ : Defined as

**argmax<sub>x</sub>** $f(x)$ : Argmax or the Arguments of Maxima are the value(s) of  $x$  that maximises the function  $f(x)$

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

**exp**( $x$ ): Exponential function;  $e^x$

**A**: Matrix  $A$

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

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

**a** or **x**: refer to a vector; e.g.,  $x = [1, 2, 3]$

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

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

**I** or **I<sub>d</sub>**: Identity matrix; e.g. A  $3 \times 3$  identify matrix  $\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\|\mathbf{x}\|_1$ : L1 norm of a vector or Manhattan distance  $\sum_{j=1}^d x_j^2$

$\|\mathbf{x}\|_2$ : L2 norm of a vector or Euclidean distance  $\sqrt{\sum_{j=1}^d x_j^2}$

$\mathbf{x} \otimes \mathbf{y}$ : 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 in probability theory:

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

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

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

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

$\mu$ : Mean of a scalar distribution

$\sigma$ : Standard deviation of a scalar distribution

$\sigma^2$ : Variance

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

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

### Notations in machine learning:

$\mathbf{C}$ : Number of Classes or Class set

$\mathbf{X}$ : Design matrix or an input dataset to a model

$\mathcal{D}$ : Training data

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

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

$\mathbf{x}$ : Input vector (one sample)

$D$ : Dimensionality of data or the number of features

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

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

$k(\mathbf{x}, \mathbf{y})$ : Kernel function

$\mathbf{K}$ : Kernel matrix

$\mathbf{T}$ : Transition matrix of Markov chain

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

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

## End Notes

Mathematical notations, Machine learning: A Probabilistic Perspective, Kevin Murphy, MIT Press, 2013; pages: 1013-1018, see: [http://noiselab.ucsd.edu/ECE228/Murphy\\_Machine\\_Learning.pdf](http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf)

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