

## Notation

### Numbers and Arrays

Syntax	Description
$a$	A scalar (integer or real)
$\mathbf{a}$	A vector
$\mathbf{A}$	A matrix
$\mathbf{A}$	A tensor
$\mathbf{I}_n$	Identity matrix with $n$ rows and $n$ columns
$\mathbf{I}$	Identity matrix with dimensionality implied by context
$\mathbf{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position $i$
$\text{diag}(\mathbf{a})$	A square, diagonal matrix with diagonal entries given by $\mathbf{a}$
$a$	A scalar random variable
$\mathbf{a}$	A vector-valued random variable
$\mathbf{A}$	A matrix-valued random variable
$\theta$	Parameters of a model
$f(\theta, \mathbf{x})$	A function (model) with parameters $\theta$ and data $\mathbf{x}$
$\mathbf{A} \odot \mathbf{B}$	Element-wise (Hadamard) product of $\mathbf{A}$ and $\mathbf{B}$

### Indexing

Syntax	Description
$a_i$	Element $i$ of vector $\mathbf{a}$ , with indexing starting at 1
$A_{i,j}$	Element $i, j$ of matrix $\mathbf{A}$

### Datasets and Distributions

Syntax	Description
$\mathbf{X}$	The design matrix with dimensionality $n \times p$ with $n$ samples with $p$ features.
$\mathbf{x}^{(i)}$	The $i$ -th training example.
$\mathbf{y}^{(i)}$	The label-vector for the $i$ -th training example.
$y^{(i)}$	The label for the $i$ -th training example.

## Probability Theory

Syntax	Description
$P(x)$	A probability distribution over a discrete variable.
$p(x)$	A probability distribution over a continuous variable or over a variable whose type has not been specified.
$\mathbb{E}_{x \sim P}[f(x)]$ or $\mathbb{E}f(x)$	Expectation of $f(x)$ with respect to $P(x)$
$\mathcal{N}(\mathbf{x}; \mu, \Sigma)$	Gaussian distribution over $\mathbf{x}$ with mean $\mu$ and covariance $\Sigma$
$x \sim \mathcal{N}(\mu, \sigma)$	Gaussian distribution over $x$ with mean $\mu$ and variance $\sigma$

## Calculus

Syntax	Description
$\nabla_{\mathbf{w}} J$	Gradient of $J$ with respect to $\mathbf{w}$
$\frac{\partial J}{\partial w}$	Partial derivative of $J$ with respect to $w$

## Functions

Syntax	Description
$\log x$	The natural logarithm of $x$ .
$\ \mathbf{x}\ _p$	$L^p$ norm of $\mathbf{x}$
$\ \mathbf{x}\ $	$L^2$ norm of $\mathbf{x}$

## Deep Learning

Syntax	Description
NCHW	The input format of images in PyTorch. N: number of images (batch size), C: number of channels, H: height, W: width