CS 1810 Notation Glossary

CS 1810 Staff

March 2, 2025

Numbers and Arrays

a	A scalar (real number)
$\boldsymbol{a}$	A vector (by default, a column)
$\boldsymbol{A}$	A matrix
$\boldsymbol{I}_n$	Identity matrix with n rows and n columns
$\boldsymbol{I}$	Identity matrix with dimensionality implied by context
$\boldsymbol{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position i
$\text{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$

Indexing

a_i	Element i of vector $\boldsymbol{a}$, with indexing starting at 1
a_{-i}	All elements of vector $\boldsymbol{a}$ except for element i
$A_{i,j}$	Element i, j of matrix $\boldsymbol{A}$

Data

x_d	Dimension d of input data point $\boldsymbol{x} \in \mathbb{R}^D$
$\boldsymbol{x}^{(n)}$ or $\boldsymbol{x}^n$	The n-th input data point from a dataset of N total observations.
$\boldsymbol{X}$	The $N \times D$ dimensional model matrix with row n corresponding to $\boldsymbol{x}^{(n)}$
$y^{(n)}$	The target/response for observation n
$\boldsymbol{y}$	The N-dimensional vector of targets

Probability and Statistics

$\mathbb{P}(A)$	Probability of event A
$p(x)$	Probability distribution (PMF/PDF) p for random variable x
$x \sim p(x)$	Random variable x has probability distribution p
$\mathbb{E}_{D \sim p(D)}[\cdot]$	Expectation over D, which has distribution p
$\mathbf{1}_A$	Indicator of event A
$\mathcal{N}(x; \mu, \sigma^2)$	The PDF of a Normal with mean μ and variance σ^2, evaluated at x
$\boldsymbol{\theta}$	General notation for parameter vector

Supervised Learning

$\boldsymbol{w}$	Weights for linear regression
$\boldsymbol{\epsilon}$	Noise terms
$p(D; \boldsymbol{\theta})$	Likelihood (frequentist) for data D and parameters $\boldsymbol{\theta}$
$p(D\|\boldsymbol{\theta})$	Likelihood (Bayesian)
ϕ	Basis transform
$\hat{y}$	Prediction for the true y
C_k	Class k, out of K classes
$\sigma(z)$	Sigmoid function
$\boldsymbol{W}$	Weights in neural nets. J often used for first hidden layer dimension, then K, etc.
$K(\boldsymbol{x}, \boldsymbol{x'})$	Kernel function
$\boldsymbol{\xi}$	Slack variables for SVMs
$\boldsymbol{\alpha}$	Lagrangian variables for SVMs

Unsupervised Learning

C_k	Cluster k, out of K clusters. Often one-hot encoded, so K-dimensional
Z_{nk}	For K-Means: indicator of whether observation n is in cluster k
n_k	Size of the cluster k
$\boldsymbol{\mu}_k$	Center of cluster k (note that this is a vector).
$\boldsymbol{z}^{(n)}$	Latent variable for observation n, usually one-hot encoded so K-dimensional
$q(\boldsymbol{z}^{(n)})$	Proxy distribution for latent variables
$\boldsymbol{u}_k$	Principal component k
$\boldsymbol{U}$	A $D \times K$ matrix where column k is $\boldsymbol{u}_k$
$\boldsymbol{x}^{(n)}$	Document n out of N, which is a list of length L, where each element is a word
$\boldsymbol{\theta}_n$	Parameter for document n's distribution over all K topics
$\boldsymbol{\phi}_k$	Parameter for topic k's distribution over all D words in the dictionary
$x_l^{(n)}$	Word l in document n
$z_l^{(n)}$	Topic from which $x_l^{(n)}$ is drawn

HMMs, MDPs, RL

$\boldsymbol{z}_t$	Hidden state at time t
$\boldsymbol{x}_t$	Observed state at time t
$\alpha_t(\boldsymbol{z}_t)$	Equal to $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t, \boldsymbol{z}_t)$, used in Forward-Backward algorithm
$\beta_t(\boldsymbol{z}_t)$	Equal to $p(\boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_T \| \boldsymbol{z}_t)$, used in Forward-Backward algorithm
γ	Discount rate
S, A, r	State, action, reward
$r(s, a)$	Reward from taking action a at state a
$p(s'\|s, a)$	Transition distribution of ending up in state s' after taking action a from state s
π	Policy
$V^{\pi}(s)$	Value of policy π at state s
$Q^{\pi}(s, a)$	Action value function

α_t	Learning rate for RL methods
V^*, Q^*	Optimal value and action-value functions
ϵ	Probability of exploration in an ϵ-greedy algorithm

References

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.