Abstract

This document describes the notational conventions we plan to follow for the tutorial slides.

1 Probability

- Probability densities are denoted by lower case letters and probability mass functions are denoted by upper case letters. Defaults: P and p are used for model densities, Q and q for approximations to the posterior or proposal distributions.
- Random variables whose outcomes are data are denoted by upper case Roman letters, while outcomes are denoted by the corresponding lower case letters. Defaults: X and x for observed data, Z and z for latent data and Y and y for data labels (observed or latent).
- The set of outcomes of a random variable X is denoted by $\mathcal{X}$. If we take Ω to be the event space, a random variable is thus a function $X : \Omega \to \mathcal{X}$.
- Random variables whose outcomes are parameters are denoted by upper case Greek letter while outcomes are denoted by the corresponding lower case letters. Defaults: Θ and θ for model parameters and Λ and λ for inference parameters.
- Non-random hyperparameters are denoted by lower case Greek letters from the beginning of the alphabet. Default: α.
- By default, all random variables are understood to be vectors.
- The expectation of a (function of) a random variable $X \sim p$ is denoted $\mathbb{E}_p\left[f(X)\right]$ by default or $\mathbb{E}\left[f(X)\right]$ when it is clear which distribution is used.
- The entropy of a random variable X is denoted by $\mathbb{H}\left(X\right)$.
- The relative entropy, or Kullback-Leibler divergence, from a distribution q to a distribution p is denoted as $\mathrm{KL}\left(q \,||\, p\right)$.
- The univariate normal distribution with mean μ and variance σ^2 is denoted as $\mathcal{N}\left(\mu, \sigma^2\right)$.
- The multivariate normal distribution with mean μ and covariance matrix Σ is denoted as $\mathcal{N}\left(\mu, \Sigma\right)$.

2 Linear Algebra

- Vectors are denoted by Roman lower case letters, Matrices (and higher-order tensors) are denoted by upper case Roman letters. To avoid the proliferation of letters, we use indeces whenever possible. For example, two weight matrices may be distinguished as W_{task1} and W_{task2}. The letter W is standardly used to denote weight matrices in neural networks.

- Matrix multiplication is denoted by $\times$ or by writing to matrices next to each other. Element-wise multiplication (Hadamard product) is denoted by $\odot$.
- The norm of a vector x is denoted $||x||$ and the Frobenius norm of a matrix W is similarly denoted $||W||$. Unless otherwise indicated, the norm is understood to be the Euclidean or L_2-norm.