Guide to Notation for CS 274A

Probabilistic Learning: Theory and Algorithms, CS 274A
Professor Padhraic Smyth

This list provides a summary of the general notation we will be using in class and in homeworks. Note that the notation in various texts, such as the Bishop text, will not necessarily always follow the same conventions as we will use in class.

- X: a random variable (upper case).
- x: a generic value of a random variable (lower case).
- $\underline{x}$ or $\mathbf{x}$: a d-dimensional vector of values (e.g., a vector of values of a random variable). Vectors are considered to be column vectors, i.e., of dimension $d \times 1$. Note that in the Bishop text **boldface** is generally used for vectors and matrices. We will use "underline" in class for vectors, since it is easier to write on the board than boldface.
- x_j: the jth component value of the vector $\underline{x}$.
- $\underline{w}^T$: T indicates the transpose of a vector w, i.e., the vector w written in $1 \times d$ "row" format. (transpose can also be defined for matrices, e.g., see Appendix C in the text).
- $\underline{w}^T\underline{x}$: the "inner product" of two vectors $\underline{w}$ and $\underline{x}$ of the same length or dimension, i.e., $\sum_{i=1}^{d} w_i x_i$.
- $g(\underline{x})$: a scalar-valued function of the vector $\underline{x}$.
- $\underline{A}$ or $\mathbf{A}$: a matrix of values with some dimension $n \times d$, i.e., n rows and d columns. Note that a vector can be thought of as a special type of matrix with $d = 1$ (i.e., a single column matrix). The value of the cell located at row i and column j is referred to as a_{ij}
- $\underline{D}$: we will often use $\underline{D}$ to refer to a data set which is represented in the form of a matrix. For example, we can have d measurements/features/attribute (corresponding to the columns of the matrix) on each of n rows (the n objects or individuals for which we have measurements in our data).
- Matrix-vector multiplication, e.g., $\underline{y} = \underline{A}\ \underline{x}$. If $\underline{A}$ has dimension $n \times d$ and $\underline{x}$ has dimension $d \times 1$, then $\underline{y}$ has dimension $n \times 1$.

- $P(A = a), p(y)$: a probability distribution and a density function respectively for the random variables A (discrete) and Y (real-valued) respectively. $P(a)$ is shorthand for $P(A = a)$. In class, when writing on the board, I will try to use P for distributions and p for densities. If in doubt, look at the argument of the probability function—if the variable (e.g., here A) takes a finite number of values, then we have a distribution: if it takes values on the real-line (e.g., here Y) then we have a density).

- $P(\underline{x}), p(\underline{x})$: a scalar-valued distribution or density function of a d-dimensional vector of variables taking values $\underline{x}$.

- $P(x|y), p(x|y)$: conditional distribution (or density) of variable X given that variable Y takes value y (generalizes for vector arguments in the obvious way).

- $P(b, c|y, z), p(b, c|y, z)$ joint conditional distribution (or density) of variables B and C given that variables Y and Z have values y and z respectively (generalizes for vector arguments in the obvious way).

- $E[x]$: the expectation of a random variable X with respect to the probability distribution $P(x)$ or density $p(x)$ (unless $E[x]$ is specifically defined with respect to some other distribution or density, e.g., $E_{p(y|x)}[x]$ is the expectation of x with respect to the density $p(x|y)$). $E[\underline{x}]$ is a d-dimensional vector where each component is $E[x_j], 1 \leq j \leq d$.

- θ: a scalar parameter.

- $\underline{\theta}$: a $p \times 1$ vector of parameters, $(\theta_1, \ldots, \theta_p)^T$.

- $\hat{\theta}$: an estimate of parameter θ. We will discuss specific types of estimates in class such as the maximum likelihood (ML) estimate for θ, $\hat{\theta}_{ML}$, as well as other (more Bayesian) estimates.

- $\prod_{i=1}^{n}$: the product from $i = 1$ to $i = n$

- $\sum_{i=1}^{n}$: the sum from $i = 1$ to $i = n$

- Σ: denotes a symmetric $d \times d$ covariance matrix (will be defined in class—see also the discussion of the Gaussian density in Appendix B in the Bishop text). Note that the symbol for Σ (covariance matrix) and $\sum$ (sum) are virtually identical: which is which should be clear from the context.