Guide to Notation for CS 274A

Padhraic Smyth Department of Computer Science University of California, Irvine January 2023

This list provides a summary of the general notation we will be using in class and in homeworks. This notation may vary a little in lectures and class notes depending on the topic we are discussing, but should be clear in general from the context. Also note that the notation in various texts may be a little different.

- X (or other upper -case letter): a random variable (upper case).
- x (or other lower-case letter): a generic value of a random variable (lower case).
- <u>x</u> or x: a *d*-dimensional vector of values (e.g., a vector of values of a random variable). In this class vectors are considered to be column vectors, i.e., of dimension $d \times 1$ unless stated otherwise. We will typically use "underline" during the quarter, since it is easier to write by hand than boldface, but you may also see occasional use of boldface (e.g., its common notation in textbooks and papers).
- \underline{w}^T : the superscript T here indicates the transpose of a vector w, i.e., the vector written in $1 \times d$ "row" format. (transpose can also of course be defined for matrices).
- $\underline{w}^T \underline{x}$: the inner product of vectors \underline{w} and \underline{x} defined as $\sum_{i=1}^d w_i x_i$.
- $g(\underline{x})$: some scalar-valued function g of the vector \underline{x} .
- <u>A</u> or A: a matrix of scalars with some number of rows and columns, e.g., an $n \times d$ with 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 matrix with a single column).
- <u>D</u> or **D**: will be used to refer to a data set in the form of a matrix, i.e., d measurements (corresponding to variables or features) for each of n rows (each row corresponding to an object or individual we have measurements for).
- Matrix-vector multiplication, e.g., <u>y</u> = <u>A</u> <u>x</u>. If <u>A</u> has dimension n × d and <u>x</u> has dimension d × 1, then y has dimension n × 1.

- P(a), p(x): a probability distribution and a probability density function respectively for the random variables A (discrete) and X (real-valued) respectively. P(a) is shorthand for P(A = a). Note that in handwritten notes in class it may not always be clear whether p is lower case or upper case. 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 X) then we have a density). We may also use sometimes other letters like f(x) or q(x) to indicate densities at times—in general it should be clear from the context.
- $P(\underline{x}), p(\underline{x})$: a scalar-valued distribution or density function of a *d*-dimensional vector of variables taking values \underline{x} (depending on whether \underline{x} has discrete or real-valued components).
- P(x|y), p(x|y): the conditional distribution (or density) of a variable X given that variable Y takes value y (generalizes to vector arguments <u>x</u> and/or y in the obvious way).
- P(b, c|y, z), p(b, c|y, z) joint conditional distribution (or density, depending on whether variables *B* and *C* are discrete or continuous random variables) of *B* and *C* given that variables *Y* and *Z* have values *y* and *z* respectively (again generalizes to 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[x] is a d-dimensional vector where each component is E[x_i], 1 ≤ j ≤ d.
- θ: a scalar parameter, e.g., representing an unknown parameter (such as the mean) for a parametric probability distribution or density. Note that we will also use other letters/symbols to indicate parameters, such as α, β, etc, depending on the context or common usage for a particular type of 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 estimate, $\hat{\theta}_{ML}$, as well as other types of estimates (e.g., 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 × d covariance matrix (will be defined in class). Note that the symbol for Σ (covariance matrix) and ∑ (sum) are virtually identical: which is which should be clear from the context.