Notations and useful equations

September 6, 2023

Here are notations and some useful equations for the MDM 2023 course. The list can be updated when the course proceeds.

1 Basic notations

The notations try to balance between two principles: 1. One font type is used for one entity type, when possible, and 2. Familiar or conventional symbols are used, even if they mismatched the font type rule. E.g., function and measure names have variable notations, both uppercase and lowercase, Roman and Greek letters, single letters and strings, and it is impossible to use just one font type for them. Similarly, some notation types are used in different meanings in different sources or even in one source (the course textbook is not systematic in notations). Still, the font types help to guess what the entity is or is not; e.g., if you see X=x in slides, you can guess X is a variable and x its value; if you see a calligraphic or bold letter you can guess it is never a single variable or value.

$A, B, C, X, Y, A_i, \ldots$	variables (uppercase letters) typically A, B, C are binary and X, Y numerical
$Dom(A)$	domain of variable A
$a_1, a_2, a_3, \ldots \in Dom(A)$	values of variable A (lowercase letters)
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{Q}$	sets of variables (bold capital letters)
$\mathbf{x} = (x_1, ..., x_k)$	vector (bold lowercase)
$\mathcal{D}$	data set
$\lvert\mathcal{D}\rvert = n$	data size (number of data points)
$\mathcal{F}$	collection of frequent patterns (calligraphic letters for collections or sets of more complex elements)
i, j, k, m	integers; k often used for the total number of features

K	number of clusters or nearest neighbours
μ	mean
σ and σ^2	standard deviation and variance
d	distance

Note: statisticians suggest to use μ and σ only for the population mean and standard deviation and something else for their sample-based estimates. However, other fields use μ and σ frequently for the sample-based estimates.

2 Intervals

Let a and b numerical values, $a \leq b$, that define a range. Notatons that value x is in this range:

$[a, b]$	closed interval
$x \in [a, b]$	$a \leq x \leq b$
$]a, b[$	open interval (also (a, b) is used for this)
$x \in]a, b[$	$a < x < b$
$]a, b]$ and $[a, b[$	half open intervals
$x \in]a, b]$	$a < x \leq b$
$[a, a] = \{a\}$	singleton set
$]a, a[$	empty set

3 Set theory symbols

Let $\mathbf{X}$ and $\mathbf{Y}$ be sets. Here are common symbols used in this course:

$\emptyset$	empty set
$\mathbf{X} \cup \mathbf{Y}$	union of $\mathbf{X}$ and $\mathbf{Y}$
$\mathbf{X} \cap \mathbf{Y}$	intersection of $\mathbf{X}$ and $\mathbf{Y}$
$\mathbf{X} \cap \mathbf{Y} = \emptyset$	intersection is empty, $\mathbf{X}$ and $\mathbf{Y}$ disjoint
$\mathbf{X} \setminus \mathbf{Y}$	difference: elements in $\mathbf{X}$ but not in $\mathbf{Y}$
$\mathbf{Y} \subseteq \mathbf{X}$	$\mathbf{Y}$ is a subset of $\mathbf{X}$ and $\mathbf{X}$ is superset of $\mathbf{Y}$, possibly $\mathbf{Y} = \mathbf{X}$
$\mathbf{Y} \subsetneq \mathbf{X}$	$\mathbf{Y}$ is a proper subset of $\mathbf{X}$ and $\mathbf{X}$ is a proper superset of $\mathbf{Y}$ (i.e., always $\mathbf{Y} \neq \mathbf{X}$) (sometimes $\subset$ is used for this)
$\mathcal{P}(\mathbf{X})$	power set of $\mathbf{X}$, i.e., all subsets of $\mathbf{X}$ including $\emptyset$ and $\mathbf{X}$ itself

Note that the curly braces may sometimes be omitted, when clear from the context. E.g., frequent set $\{A, B, C\}$ can be notated ABC.

4 Probability notations and basic definitions

If A and B be binary variables, the truth values are notated compactly as A for $A = 1$ and $\neg A$ for $A = 0$, when clear from the context. Probability of A (i.e., event $A = 1$) is notated $P(A)$. Probability of the conjunction of A and B (i.e., event where $A = 1$ and $B = 1$), is notated by $P(A, B)$ or simply $P(AB)$.

Note: For simplicity, we don't usually distinguish the true probability and its sample-based (maximum-likelihood) estimate (sometimes notated $\tilde{P}$).

Events $A = a$ and $B = b$ are **statistically independent**, if $P(A = a, B = b) = P(A = a)P(B = b)$. (In the vase of binary vriables, $a, b \in \{0, 1\}$.) Variables A and B are statistically independent, if for all $a \in Dom(A)$, $b \in Dom(B)$, $P(A = a, B = b) = P(A = a)P(B = b)$. (For binary variables, $Dom(A) = Dom(B) = \{0, 1\}$.)

Events $A = a$ and $B = b$ are **conditionally independent**, given $C = c$, if $P(A = a, B = b|C = c) = P(A = a|C = c)P(B = b|C = c)$. Variables A and B are conditionally independent, given C, if $P(A = a, B = b|C = c) = P(A = a|C = c)P(B = b|C = c)$ for all $a \in Dom(A)$, $b \in Dom(B)$, $c \in Dom(C)$.

5 Number of pairwise distances

If data contains n points, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, there are $\frac{n(n-1)}{2}$ pairwise distances (between any two data points). The idea is that from $\mathbf{x}_1$ you can calculate distance to the rest $n - 1$ points, from $\mathbf{x}_2$ to $n - 2$ points (not again to $\mathbf{x}_1$), from $\mathbf{x}_3$ to $n-3$ points, etc. This equals to $n-1+n-2+\ldots+2+1 = \frac{n(n-1)}{2}$.

Note that here distances are calculated just once (distance between $\mathbf{x}$ and $\mathbf{y}$ does not have direction, $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$) and distance from the point to itself is excluded ($d(\mathbf{x}, \mathbf{x}) = 0$) is not a pairwise distance).

6 On logarithms

Basic rules:

$\log(x) + \log(y) = \log(xy)$
$\log(x) - \log(y) = \log(\frac{x}{y})$
$y \log(x) = log(x^y)$
$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$

These can be chained. E.g., from the first rule you can derive $\sum_i \log(x_i) = \log(\prod_i x_i)$

Note: In the definition of entropy (and thus mutual information), there is a **convention that $\mathbf{0}\log(\mathbf{0}) = \mathbf{0}$.** (Note that $\log(0)$ itself is not defined.)