Preliminaries

Garrett Thomas

July 11, 2018

1 About

This document is part of a series of notes about math and machine learning. You are free to distribute it as you wish. The latest version can be found at http://gwthomas.github.io/notes. Please report any errors to gwthomas@stanford.edu.

2 Sets

A **set** is a collection of objects which is determined entirely by its contents. The elements of a set have no ordering or multiplicity within the set. If S is a set, we write $x \in S$ to denote that the object x is a member of the set S, and $x \notin S$ to denote that x is not a member of S. The set which contains no elements is referred to as **the empty set** and denoted $\varnothing$.

If A and B are sets, and every element of A is also an element of B, we say that A is a **subset** of B (or equivalently, that B is a **superset** of A) and write $A \subseteq B$. Two sets A and B are equal if both $A \subseteq B$ and $B \subseteq A$, which is to say that they contain exactly the same elements.

If A is a set, the **powerset** of A, denoted $\mathcal{P}(A)$, is the set of all subsets of A.

2.1 Specifying sets

Sets can be written down in a number of ways. If there are only a few elements, one can simply list them out, e.g. $\{\text{cat}, 3, a\}$. For sets too large to enumerate, ellipsis notation $(\dots)$ indicates that the reader should be able to infer the pattern, e.g. $\{2, 4, 6, 8, \dots\}$ or $\{2, 4, 6, \dots, 100\}$. The trailing ellipses in the first of these indicates that the pattern continues infinitely.

It is also common to use **set-builder** notation to specify a set by means of one or more conditions. For example, to specify the set formed by keeping only the positive elements of some set A, we might write $\{x \in A : x \text{ is positive}\}$.

One can also describe sets with natural language, as long as the description is unambiguous, e.g. "the set of all humans who were born before the year 2000".

2.2 Constructing sets from existing sets

The **intersection** of A and B, denoted $A \cap B$, is the set of elements which lie in both A and B:

$$A \cap B \triangleq \{x : x \in A, x \in B\}$$

A and B are said to be **disjoint** if they have no elements in common, i.e. $A \cap B = \varnothing$.

The **union** of A and B, denoted $A \cup B$, is the set of elements which lie in A or B (or both):

$$A \cup B \triangleq \{x : x \in A \text{ or } x \in B\}$$

Clearly $A \cap B \subseteq A, B \subseteq A \cup B$ for any sets A and B.

These binary operations have natural generalizations to combinations of many sets. Let $\{A_\alpha\}_{\alpha \in I}$ be a collection of sets indexed by some set I. Then the intersection and union of this collection are

$$\bigcap_{\alpha \in I} A_\alpha \triangleq \{x : x \in A_\alpha \text{ for all } \alpha \in I\}$$
$$\bigcup_{\alpha \in I} A_\alpha \triangleq \{x : x \in A_\alpha \text{ for at least one } \alpha \in I\}$$

The **difference** of A and B is the set of elements which lie in A but not in B:

$$A \setminus B \triangleq \{x : x \in A, x \notin B\}$$

If $A \subseteq B$, the **complement** of A (in B) is defined as $A^{\mathrm{c}} \triangleq B \setminus A$. Note that the notation for complements does not explicitly include the containing set B; this set should be clear from context.

The interplay between complements, intersections, and unions is well-captured by **De Morgan's laws**:

$$(A \cap B)^{\mathrm{c}} = A^{\mathrm{c}} \cup B^{\mathrm{c}} \qquad\qquad (A \cup B)^{\mathrm{c}} = A^{\mathrm{c}} \cap B^{\mathrm{c}}$$

or, more generally,

$$\left(\bigcap_{\alpha \in I} A_\alpha\right)^{\mathrm{c}} = \bigcup_{\alpha \in I} A_\alpha^{\mathrm{c}} \qquad\qquad \left(\bigcup_{\alpha \in I} A_\alpha\right)^{\mathrm{c}} = \bigcap_{\alpha \in I} A_\alpha^{\mathrm{c}}$$

3 Common sets of numbers

The **natural numbers** are the set $\mathbb{N} \triangleq \{0, 1, 2, \dots\}$.

The **integers** are the set $\mathbb{Z} \triangleq \{\dots, -2, -1, 0, 1, 2, \dots\}$.

The **rational numbers** are the set $\mathbb{Q} \triangleq \{\frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{Z}, q \neq 0\}$.

The **real numbers** are denoted $\mathbb{R}$. We will not define them here but trust that the reader is familiar with their basic properties.

The **complex numbers** are the set $\mathbb{C} \triangleq \{a + ib : a, b \in \mathbb{R}\}$, where i is the imaginary unit which satisfies $i^2 = -1$.

For shorthand, we define the notation $[n] \triangleq \{1, 2, \dots, n\}$ when n is a positive integer.

4 Tuples

An n**-tuple** (where n is a positive integer) is an ordered list of n objects, denoted with parentheses, e.g. $(x_1, x_2, \dots, x_n)$. If the number of items is unimportant, we will simply say **tuple**. A 2-tuple is referred to as an **ordered pair**, and a 3-tuple as a **triple**.

The **Cartesian product** of n sets $A_1, \dots, A_n$ is the set of all n-tuples whose entries come from those sets:

$$A_1 \times \cdots \times A_n \triangleq \{(x_1, \dots, x_n) : x_1 \in A_1, \dots, x_n \in A_n\}$$

For example, $\{1, 2, 3\} \times \{4, 5\} = \{(1, 4), (2, 4), (3, 4), (1, 5), (2, 5), (3, 5)\}$. If A is a set, the notation A^n means the Cartesian product of n copies of A, i.e.

$$A^n \triangleq \underbrace{A \times \cdots \times A}_{n \text{ times}}$$

5 Functions

A **function** is a rule which associates to each element of a certain set, known as the **domain** of the function, exactly one element of another set, known as the **codomain** of the function. We write $f : X \to Y$ to denote that f is a function from X to Y; here X is the domain (denoted $\operatorname{dom} f$), and Y the codomain. The element of Y associated to $x \in X$ is written $f(x)$.

Perhaps the simplest example of a function is the **identity** function on a set X, denoted id_X, which maps every element to itself:

$$\begin{aligned} \mathrm{id}_X : X &\to X \\ x &\mapsto x \end{aligned}$$

The **image** of $A \subseteq X$ under f, denoted $f(A)$, is the set of elements in Y that are mapped to by elements of A:

$$f(A) \triangleq \{f(x) : x \in A\}$$

Conversely, the **preimage** of $A \subseteq Y$ under f, denoted $f^{-1}(A)$, is the set of elements of X which map to some element of A:

$$f^{-1}(A) \triangleq \{x \in X : f(x) \in A\}$$

The **range** of f is the image of the entire domain of f:

$$\operatorname{range} f \triangleq f(\operatorname{dom} f)$$

Functions can be chained together if their domains and codomains are appropriately matched up. If $f : X \to Y$ and $g : Y \to Z$, then their **composition** $g \circ f$ is defined as

$$\begin{aligned} g \circ f : X &\to Z \\ x &\mapsto g(f(x)) \end{aligned}$$

If two functions $f : X \to Y$ and $g : Y \to X$ satisfy the relationships $g \circ f = \mathrm{id}_X$ (i.e. $g(f(x)) = x$ for all $x \in X$) and $f \circ g = \mathrm{id}_Y$ (i.e. $f(g(y)) = y$ for all $y \in Y$), then g is said to be the **inverse** of f, and we write $g = f^{-1}$.[1] Due to the symmetry involved here it should be clear that $f = g^{-1}$ is an equivalent statement.

A function is said to be **surjective** if its range is the entirety of its codomain, i.e. $f(X) = Y$. It is **injective** if $f(x) = f(\tilde{x})$ implies $x = \tilde{x}$ for all $x, \tilde{x} \in X$. A function that is both surjective and injective is said to be **bijective**.

Proposition 1. *A function is bijective if and only if it has an inverse.*

[1] Note the overloading of the notation f^{-1}! It is unfortunate but standard.

Proof. Suppose $f : X \to Y$ is bijective. As a consequence of surjectivity, for each $y \in Y$ there exists at least one $x \in X$ such that $f(x) = y$. However injectivity tells us that there is at most one such x for each y, since if $f(x_1) = y = f(x_2)$ then we must have $x_1 = x_2$. Hence each $y \in Y$ has exactly one $x \in X$ satisfying $f(x) = y$. Define $g : Y \to X$ as the function which maps each $y \in Y$ to the corresponding $x \in X$. Then we have $g(f(x)) = x$ for all $x \in X$ and $f(g(y)) = y$ for all $y \in Y$, so $g = f^{-1}$.

Conversely, suppose f has an inverse $f^{-1} : Y \to X$. Then for any $y \in Y$ we have $f(f^{-1}(y)) = y$, so $y \in \text{range}(f)$, and hence f is surjective. Furthermore if $x, \tilde{x} \in X$ and $f(x) = f(\tilde{x})$, then

$$x = f^{-1}(f(x)) = f^{-1}(f(\tilde{x})) = \tilde{x}$$

so f is injective. □

6 Sequences

A **sequence** is a countable ordered list whose entries come from a common set. We write $(x_n) \subseteq X$ to denote that (x_n) is a sequence whose entries lie in the set X. Formally, we consider a sequence as a function from some interval of integers (usually the natural numbers) into X, and observe that x_n is shorthand for this function evaluated at n.

A sequence is said to be a **subsequence** of another sequence if it can be formed from this second sequence by dropping zero or more entries, without reordering any. More formally, (y_n) is a subsequence of (x_n) if there exist indices $i_1 < i_2 < \dots$ such $y_j = x_{i_j}$ for all j.