Lecture 1 - Maths for Machine Learning

Dahua Lin

The Chinese University of Hong Kong

Roadmap

Mathematics lies at the heart of machine learning research. In addition to *linear algebra* and *statistics*, the development of many learning methods is also closely related to *functional analysis*, *modern probability theory*, and *convex optimization*.

Complete coverage of all these subjects is obviously beyond the scope of this course. This lecture aims to give an overview of several important math concepts that are useful in the course.

Concepts to Be Covered

- Metrics
- Basics of topology
  - *open* and *closed sets*, *continuous function*, and *compact space*
- Basics of functional analysis:
  - *norm*, *inner product*, *Banach space*, *Hilbert space*
  - *functional*, *operator*, and *bilinear form*
- Basics of modern probability theory:
  - *measure space*, and *Lebesgue integration*
  - *random variables*, *expectation*, and *convergence of laws*.
- Convex set and convex functions

Metrics

- Measurement of *distances* or *deviation* lies at the heart of many learning problems.
- Given an arbitrary set Ω, a real-valued function $d : \Omega \times \Omega \mapsto \mathbb{R}$ is called a *metric*, if it satisfies:
  - *(non-negativity)* $d(x, y) \geq 0,\ \forall x, y \in \Omega$
  - *(coincidence axiom)* $d(x, y) = 0 \Leftrightarrow x = y$
  - *(symmetry)* $d(x, y) = d(y, x),\ \forall x, y \in \Omega$
  - *(triangle inequality)* $d(x, y) + d(y, z) \geq d(x, z),\ \forall, x, y, z \in \Omega$
- The *triangle inequality* can be further generalized:

$$d(x_1, x_n) \leq d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{n-1}, x_n).$$

- A set Ω together with a metric d defined thereon is called a *metric space*, denoted by (Ω, d).

Examples of Metrics

- *Euclidean metric* $d_{euc} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$:

$$d_{euc}(\mathbf{x}, \mathbf{y}) \triangleq \sqrt{\sum_{i=1}^{m} \left(x^{(i)} - y^{(i)}\right)^2}$$

- *Rectilinear metric* $d_{rec} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$:

$$d_{rec}(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^{m} \left|x^{(i)} - y^{(i)}\right|$$

Examples of Metrics

- *Euclidean metric* $d_{euc} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$:

$$d_{euc}(\mathbf{x}, \mathbf{y}) \triangleq \sqrt{\sum_{i=1}^{m} \left(x^{(i)} - y^{(i)}\right)^2}$$

- *Rectilinear metric* $d_{rec} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$:

$$d_{rec}(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^{m} \left| x^{(i)} - y^{(i)} \right|$$

- One can define a metric over an arbitrary set S through an *injective map* $f : S \to \Omega$ to a metric space (Ω, d_Ω):

$$d_S(x, y) = d_\Omega(f(x), f(y)), \ \forall x, y \in S$$

It can be easily verified that d_S is a metric on S.

Topology Defined by Metrics

- *Topological space* is a more general concept than *metric space*
  - With *metrics*, *topological concepts* can be defined in a more intuitive way

Topology Defined by Metrics

- *Topological space* is a more general concept than *metric space*
  - With *metrics*, *topological concepts* can be defined in a more intuitive way

- *Open ball* of radius r: $B_\epsilon(x) = \{y \in \Omega : d(y, x) < \epsilon\}$
- x is called an *interior point* of S if there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset S$. The *interior* of S, denoted by S° or $\mathrm{int}(S)$, is the set of all interior points of S is called.
- x is called a *boundary point* of S, if for any $\epsilon > 0$, $B_\epsilon(x)$ overlaps with both S and $\Omega - S$. The *boundary* of S, denoted by ∂S, is the set of all *boundary points* of S.

Topology Defined by Metrics

- *Topological space* is a more general concept than *metric space*
  - With *metrics*, *topological concepts* can be defined in a more intuitive way

- *Open ball* of radius r: $B_\epsilon(x) = \{y \in \Omega : d(y, x) < \epsilon\}$
- x is called an *interior point* of S if there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset S$. The *interior* of S, denoted by S° or $\mathrm{int}(S)$, is the set of all interior points of S is called.
- x is called a *boundary point* of S, if for any $\epsilon > 0$, $B_\epsilon(x)$ overlaps with both S and $\Omega - S$. The *boundary* of S, denoted by ∂S, is the set of all *boundary points* of S.

- Consider a metric space (Ω, d), and a subset $S \subset \Omega$:
  - S is called an *open set*, if $S = S^\circ$
  - S is called a *closed set*, if $\Omega - S$ is open

Properties of Open and Closed Sets

- S is *open* if $\partial S \cap S = \emptyset$.
- S is *closed* if $\partial S \subset S$.
- The union of any collections of open sets is open.
- The intersection of *finitely many* open sets is open.
- The intersection of any collections of closed sets is closed.
- The union of *finitely many* closed sets is closed.

Questions

- Consider an arbitrary metric space (Ω, d), are Ω and $\emptyset$ open or closed?

Questions

- Consider an arbitrary metric space (Ω, d), are Ω and $\emptyset$ open or closed?
- Let (Ω, d) be a finite metric space and $x \in \Omega$, is $\{x\}$ open or closed?

Questions

- Consider an arbitrary metric space (Ω, d), are Ω and $\emptyset$ open or closed?
- Let (Ω, d) be a finite metric space and $x \in \Omega$, is $\{x\}$ open or closed?
- Consider *Euclidean space* $\mathbb{R}^2$ and $S = (a, b) \times \{0\}$, is S open or closed? what is ∂S?

Closure

- Let S be a subset of a metric space (Ω, d), then the *(topological) closure* of S, denoted by $\bar{S}$ or $\mathrm{cl}(S)$, is defined to be the *minimum* closed set that contains S, namely the intersection of all closed sets that contain S.
- S is closed if and only if $\bar{S} = S$.
- $\bar{S} = S \cup \partial S$.
- $\partial S = \bar{S} - S^\circ$.
- Examples:
  - $\mathrm{cl}((a, b)) = [a, b]$
  - $\mathrm{cl}(B_\epsilon(x)) = \{y \in \Omega : d(x, y) \leq \epsilon\}$.

Convergence and Cauchy Sequences

- Let $x_1, x_2, \ldots$ be a sequence in a metric space (Ω, d), then $x_* \in \Omega$ is called a *limit* of this sequence if

$$\forall \epsilon > 0,\ \exists N : \forall m > N, d(x_m, x_*) < \epsilon.$$

- A sequence is said to be *convergent* if it has a *limit*.
- The *limit* of a convergent sequence is unique.

Convergence and Cauchy Sequences

- Let $x_1, x_2, \dots$ be a sequence in a metric space (Ω, d), then $x_* \in \Omega$ is called a *limit* of this sequence if

$$\forall \epsilon > 0,\ \exists N : \forall m > N, d(x_m, x_*) < \epsilon.$$

- A sequence is said to be *convergent* if it has a *limit*.
- The *limit* of a convergent sequence is unique.

- A sequence $x_1, x_2, \dots$ is called a *Cauchy sequence* if

$$\forall \epsilon > 0,\ \exists N : \forall m, n > N, d(x_m, x_n) < \epsilon.$$

Convergence and Cauchy Sequences

- Let $x_1, x_2, \ldots$ be a sequence in a metric space (Ω, d), then $x_* \in \Omega$ is called a *limit* of this sequence if

$$\forall \epsilon > 0,\ \exists N : \forall m > N, d(x_m, x_*) < \epsilon.$$

- A sequence is said to be *convergent* if it has a *limit*.
- The *limit* of a convergent sequence is unique.

- A sequence $x_1, x_2, \ldots$ is called a *Cauchy sequence* if

$$\forall \epsilon > 0,\ \exists N : \forall m, n > N, d(x_m, x_n) < \epsilon.$$

- A *convergent sequence* must be a *Cauchy sequence*
- **Question:** Is a *Cauchy sequence* always convergent?

Complete Metric Space

- Consider a real valued sequence $x_n = \left(1 + \frac{1}{n}\right)^n$.
  - We have learned in Calculus that $x_n \to e$ as $n \to \infty$.

Complete Metric Space

- Consider a real valued sequence $x_n = \left(1 + \frac{1}{n}\right)^n$.
  - We have learned in Calculus that $x_n \to e$ as $n \to \infty$.
- Note that (x_n) is a rational sequence. Is this sequence convergent in the *rational space* $\mathbb{Q}$?

Complete Metric Space

- Consider a real valued sequence $x_n = \left(1 + \frac{1}{n}\right)^n$.
  - We have learned in Calculus that $x_n \to e$ as $n \to \infty$.

- Note that (x_n) is a rational sequence. Is this sequence convergent in the *rational space* $\mathbb{Q}$?

- **Convergence is not an intrinsic property of a sequence, which also depends on the space in which the sequence lies.**
- A metric space (Ω, d) is called a *complete metric space* if *every Cauchy sequence converges*.
  - $\mathbb{R}$ is complete, but $\mathbb{Q}$ is *not*.
  - $\mathbb{R}$ can be constructed by *completing* $\mathbb{Q}$.

Complete Metric Space

- Consider a real valued sequence $x_n = \left(1 + \frac{1}{n}\right)^n$.
  - We have learned in Calculus that $x_n \to e$ as $n \to \infty$.
- Note that (x_n) is a rational sequence. Is this sequence convergent in the *rational space* $\mathbb{Q}$?
- **Convergence is not an intrinsic property of a sequence, which also depends on the space in which the sequence lies.**
- A metric space (Ω, d) is called a *complete metric space* if *every Cauchy sequence converges*.
  - $\mathbb{R}$ is complete, but $\mathbb{Q}$ is *not*.
  - $\mathbb{R}$ can be constructed by *completing* $\mathbb{Q}$.
- **Question:** Is the integer space $\mathbb{Z}$ with $d(x, y) \triangleq |x - y|$ a *complete metric space*?

Closed, Bounded, and Convergent

Let S be a subset of a complete metric space (Ω, d):

- Being *closed* means *"containing all the limits"*:
  - Let (x_n) be a sequence in S and $x_n \to x$, then $x \in \bar{S}$
  - S is *closed* if and only if S is *complete*, meaning *all convergent sequences in S converge within S*.

Closed, Bounded, and Convergent

Let S be a subset of a complete metric space (Ω, d):

- Being *closed* means *"containing all the limits"*:
  - Let (x_n) be a sequence in S and $x_n \to x$, then $x \in \bar{S}$
  - S is *closed* if and only if S is *complete*, meaning *all convergent sequences in S converge within S*.
- The *diameter* of S, denoted by $\delta(S)$, is defined to be the largest distance between points in S, as

$$\delta(S) \triangleq \sup_{x,y \in S} d(x, y).$$

- S is said to be bounded if $\delta(S) < \infty$.

Compactness

- In elementary Calculus, we learned that *"every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence"*.
  - Does this hold in general metric spaces?

Compactness

- In elementary Calculus, we learned that *"every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence"*.
  - Does this hold in general metric spaces?
- Consider a metric space (Ω, d). Then Ω is said to be *compact* if any open cover of Ω has a finite subcover.

Compactness

- In elementary Calculus, we learned that *"every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence"*.
  - Does this hold in general metric spaces?
- Consider a metric space (Ω, d). Then Ω is said to be *compact* if any open cover of Ω has a finite subcover.
- A metric space (Ω, d) is *compact* if and only if either of the following holds:
  - Ω is *complete* and *totally bounded*, namely, for every $\epsilon > 0$, there exists a finite cover of Ω by ϵ-balls.
  - Ω is *sequentially compact*, *i.e.* every sequence in Ω has a convergent subsequence.

Compactness

- In elementary Calculus, we learned that *"every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence"*.
  - Does this hold in general metric spaces?
- Consider a metric space (Ω, d). Then Ω is said to be *compact* if any open cover of Ω has a finite subcover.
- A metric space (Ω, d) is *compact* if and only if either of the following holds:
  - Ω is *complete* and *totally bounded*, namely, for every $\epsilon > 0$, there exists a finite cover of Ω by ϵ-balls.
  - Ω is *sequentially compact*, *i.e.* every sequence in Ω has a convergent subsequence.
- Compact sets are always *closed and bounded*.
  - But, the converse is *generally false*.
  - $S \subset \mathbb{R}^m$ is compact *if and only if* S is bounded and closed.

Continuous Function

- Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is called a *continuous function*, if $f^{-1}(V) = \{x \in X : f(x) \in V\}$ is open whenever $V \subset Y$ is open.
- f is continuous, if and only if either of the following holds:
  - (*Weierstrass definition*) $\forall \epsilon > 0, \forall x \in X, \exists \delta > 0, d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \epsilon$
  - (*Sequential continuity*) $f(x_n) \to f(x)$ whenever $x_n \to x$.

Continuous Function

- Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is called a *continuous function*, if $f^{-1}(V) = \{x \in X : f(x) \in V\}$ is open whenever $V \subset Y$ is open.
- f is continuous, if and only if either of the following holds:
  - (*Weierstrass definition*) $\forall \epsilon > 0, \forall x \in X, \exists \delta > 0, d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \epsilon$
  - (*Sequential continuity*) $f(x_n) \to f(x)$ whenever $x_n \to x$.
- f is called *uniformly continuous*, if $\forall \epsilon > 0, \exists \delta > 0, \forall x, x' \in X, d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \epsilon$.
- f is called *Lipschitz continuous*, if there exists $L > 0$, such that $d_Y(f(x), f(x')) \le L \cdot d_X(x, x'), \forall x, x' \in X$. Here, L is called the *Lipschitz constant*.
- *Lipschitz continuous* $\Rightarrow$ *Uniform continuous* $\Rightarrow$ *Continuous*.

Continuous Function on Compact Domain

- Let f be a continuous function, then $f(S) = \{f(x) : x \in S\}$ is compact whenever S is compact.
- Let $f : S \to \mathbb{R}$, then f assumes *maximum* and *minimum* within S when S is compact.
- A *continuous function* is *uniformly continuous* over a compact domain.
- A *continuously differentiable function* is *Lipschitz continuous* over a compact domain.

Dense Set and Separable Space

Consider a metric space Ω:

- A subset $S \subset \Omega$ is said to be *dense in* Ω, if $\bar{S} = \Omega$.
- A subset S is *dense in* Ω, if and only if either of the following holds:
  - Each open subset of Ω contains at least one element in S.
  - Each $x \in \Omega$ is a limit of some sequence in S.
- $\mathbb{Q}$ is dense in $\mathbb{R}$, $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

Dense Set and Separable Space

Consider a metric space Ω:

- A subset $S \subset \Omega$ is said to be *dense in* Ω, if $\bar{S} = \Omega$.
- A subset S is *dense in* Ω, if and only if either of the following holds:
  - Each open subset of Ω contains at least one element in S.
  - Each $x \in \Omega$ is a limit of some sequence in S.
- $\mathbb{Q}$ is dense in $\mathbb{R}$, $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

- Ω is called a *separable space*, if Ω contains a countable dense subset.
- Let f be a continuous function on Ω, and S is *dense*, then f is *uniquely* determined by $f|_S$.

Dense Set and Separable Space

Consider a metric space Ω:

- A subset $S \subset \Omega$ is said to be *dense in* Ω, if $\bar{S} = \Omega$.
- A subset S is *dense in* Ω, if and only if either of the following holds:
  - Each open subset of Ω contains at least one element in S.
  - Each $x \in \Omega$ is a limit of some sequence in S.
- $\mathbb{Q}$ is dense in $\mathbb{R}$, $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

- Ω is called a *separable space*, if Ω contains a countable dense subset.
- Let f be a continuous function on Ω, and S is *dense*, then f is *uniquely* determined by $f|_S$.

- **Question:** Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, and it has $f(x + y) = f(x) \cdot f(y)$ and $f(1) = a$. Is f completely determined by these conditions? If so, what is f?

Contraction and Fixed Points

Consider a function $f : \Omega \to \Omega$.

- $x \in \Omega$ is called a *fixed point* of f if $f(x) = x$.
  - Some functions may not have fixed points, *e.g.* $x \mapsto x + a$
  - Some functions may have infinitely many fixed points, *e.g.* $x \mapsto x$
- A function $f : \Omega \to \Omega$ defined on a metric space (Ω, d) is called a *contraction* if it is *Lipschitz continuous* with $L < 1$.
- *(Banach fixed point theorem)* Let f be a *contraction* on a metric space (Ω, d), then f has a unique fixed point.
- Let f be a contraction with fixed point x_*. Consider an iterative sequence $x_0, x_1 = f(x_0), x_2 = f(x_1), \ldots$, then $x_n \to x_*$ as $n \to \infty$.
  - *(Prior error estimate):* $d(x_m, x) \leq \frac{L^m}{1-L} d(x_0, x_1)$.
  - *(Posterior error estimate):* $d(x_m, x) \leq \frac{L}{1-L} d(x_{m-1}, x_m)$.

Norms and Banach Spaces

- Things become much more interesting when *vector spaces* meet with *metrics*.

Norms and Banach Spaces

- Things become much more interesting when *vector spaces* meet with *metrics*.
- Consider a *vector space* Ω, a function $\|\cdot\| : \Omega \mapsto \mathbb{R}$ is called a *norm* if:
  - $\|\mathbf{x}\| \geq 0, \forall \mathbf{x} \in \Omega$
  - $\|\mathbf{x}\| = 0$ iff $x = 0$. ($\|\cdot\|$ is called a *seminorm* without this)
  - $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|, \ \forall \alpha \in \mathbb{R}, \mathbf{x} \in \Omega$.
  - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \ \forall \mathbf{x}, \mathbf{y} \in \Omega$.
- A vector space together with a norm, as $(\Omega, \|\cdot\|)$, is called a *normed space*.
  - A *normed space* is always a *metric space*, where the *norm* induces a *metric* as $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.
- A *complete normed space* is called a *Banach space*.

Properties of Norms

- The *norm* function $x \mapsto \|x\|$ is Lipschitz continuous:

$$|\|x\| - \|y\|| \leq \|x - y\|$$

- The metric d induced by the norm:
  - *translation-invariance*: $d(x + a, y + a) = d(x, y)$
  - $d(\alpha x, \alpha y) = |\alpha| \cdot d(x, y)$

Examples of Norms

Consider a real vector space $\mathbb{R}^m$. For each $p \geq 1$, we can define L_p-norm as:

$$\|\mathbf{x}\|_p \triangleq \left(\sum_{i=1}^{m} |x^{(i)}|^p \right)^{1/p}$$

It can be easily verified that $\| \cdot \|_p$ is a *norm*. When p is set to ∞, L_∞-norm is still a *norm*, defined as

$$\|\mathbf{x}\|_\infty \triangleq \max_{i=1}^{m} |x^{(i)}|$$

We will extend these norms to the *space of functions*.

Convergence of Vectors and Basis

- Let $x_1, x_2, \ldots$ be a sequence of vectors in a Banach space, we say (x_n) *converges* to x if $\|x_n - x\| \to 0$ as $n \to \infty$.
  - This is sometimes called *convergence in norm*.
- Consider an infinite series $s = x_1 + x_2 + \cdots$ in a Banach space:
  - Let $s_n = x_1 + \ldots + x_n$. The *series* s is said to be *convergent* if the *sequence* $s_1, s_2, \ldots$ converges.
  - s is said to be *absolutely convergent* if the series $\|x_1\| + \|x_2\| + \cdots$ converges.
- *Absolute convergence* implies *convergence (in norm)*.

Basis of Banach Space

- In elementary linear algebra, we learned that a *basis* of a vector space is a linearly independent set of vectors that spans the space.
  - If a vector space has a finite basis, it is called *finite-dimensional*. The cardinalities of all bases of a *finite-dimensional* space are the same, called the *dimension* of the space.

Basis of Banach Space

- In elementary linear algebra, we learned that a *basis* of a vector space is a linearly independent set of vectors that spans the space.
  - If a vector space has a finite basis, it is called *finite-dimensional*. The cardinalities of all bases of a *finite-dimensional* space are the same, called the *dimension* of the space.
- Let (e_n) be a sequence in a Banach space Ω, it is called a *(Schauder) basis* of Ω if for each $x \in \Omega$, there exists a unique sequence of real values (α_n) such that

$$\|x - (\alpha_1 e_1 + \cdots + \alpha_n e_n)\| \to 0, \qquad \text{as } n \to 0$$

- A Banach space Ω with a Schauder basis must be *separable*.

Basis of Banach Space

- In elementary linear algebra, we learned that a *basis* of a vector space is a linearly independent set of vectors that spans the space.
  - If a vector space has a finite basis, it is called *finite-dimensional*. The cardinalities of all bases of a *finite-dimensional* space are the same, called the *dimension* of the space.
- Let (e_n) be a sequence in a Banach space Ω, it is called a *(Schauder) basis* of Ω if for each $x \in \Omega$, there exists a unique sequence of real values (α_n) such that

$$\|x - (\alpha_1 e_1 + \cdots + \alpha_n e_n)\| \to 0, \qquad \text{as } n \to 0$$

- A Banach space Ω with a Schauder basis must be *separable*.
- It took about forty years to get the answer – *No*. Enflo constructed a separable Banach space with no Schauder basis in 1973.

Inner Products and Hilbert Spaces

- An inner product on a real vector space Ω is a mapping $(x, y) \mapsto \langle x, y \rangle$ that satisfies:
  - $\langle x, x \rangle \geq 0$
  - $\langle x, x \rangle = 0$ iff $x = 0$
  - *(symmetry)* $\langle x, y \rangle = \langle y, x \rangle$
  - *(bilinearity)* $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
- A vector space together with an inner product is called an *inner product space*.
- An *inner product space* is always a *normed space*, where the *norm* is induced by the *inner product*, as $\|x\| = \sqrt{\langle x, x \rangle}$.
- A *complete inner product space* is called a *Hilbert space*.
  - A *Hilbert space* is always a *Banach space*.
- $\mathbb{R}^m$ is a *Hilbert space* with the inner product defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{m} x^{(i)} y^{(i)}$, which induces the L_2-norm and *Euclidean metric*.

Properties of Inner Products

- *(Parallelogram equality)* $\|x+y\|^2 + \|x-y\|^2 = \|x\|^2 + \|y\|^2$ – this only holds for norms induced by inner products.
- *(Cauchy–Schwarz inequality)* $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$, or equivalently, $(\langle x, y\rangle)^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$.
- *(Continuity)* if $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.

Orthogonality and Projection

Let Ω be a Hilbert space:

- $x, y \in \Omega$ are said to be *orthogonal* to each other, denoted by $x \perp y$, if $\langle x, y \rangle = 0$.
- A subset $M \subset \Omega$ is called an *orthogonal set* if elements of M are mutually orthogonal. Moreover, if each element has a unit norm, then M is called an *orthonormal set*.
- *(Pythagorean theorem)* $x \perp y \Leftrightarrow \|x\|^2 + \|y\|^2 = \|x + y\|^2$.

Orthogonality and Projection

Let Ω be a Hilbert space:

- $x, y \in \Omega$ are said to be *orthogonal* to each other, denoted by $x \perp y$, if $\langle x, y \rangle = 0$.
- A subset $M \subset \Omega$ is called an *orthogonal set* if elements of M are mutually orthogonal. Moreover, if each element has a unit norm, then M is called an *orthonormal set*.
- *(Pythagorean theorem)* $x \perp y \Leftrightarrow \|x\|^2 + \|y\|^2 = \|x + y\|^2$.

- Let $x \in \Omega$ and $S \subset \Omega$, then the *distance* between x and S is defined to be $d(x, S) \triangleq \inf_{y \in S} d(x, y)$.
- Let S be a non-empty convex closed subset of Ω. Then there exists a unique element $y_* \in S$ such that $d(x, y_*) = d(x, S)$. This element y_* is called the *projection* of x onto S, denoted by $\mathrm{proj}_S(x)$.

Orthogonality and Projection

Let Ω be a Hilbert space:

- $x, y \in \Omega$ are said to be *orthogonal* to each other, denoted by $x \perp y$, if $\langle x, y \rangle = 0$.
- A subset $M \subset \Omega$ is called an *orthogonal set* if elements of M are mutually orthogonal. Moreover, if each element has a unit norm, then M is called an *orthonormal set*.
- *(Pythagorean theorem)* $x \perp y \Leftrightarrow \|x\|^2 + \|y\|^2 = \|x + y\|^2$.

- Let $x \in \Omega$ and $S \subset \Omega$, then the *distance* between x and S is defined to be $d(x, S) \triangleq \inf_{y \in S} d(x, y)$.
- Let S be a non-empty convex closed subset of Ω. Then there exists a unique element $y_* \in S$ such that $d(x, y_*) = d(x, S)$. This element y_* is called the *projection* of x onto S, denoted by $\mathrm{proj}_S(x)$.

- **Question:** Why does S need to be closed and convex?

Orthogonality and Projection (cont'd)

Let Ω be a Hilbert space and S be a subspace of Ω:

- Given $x \in \Omega$ and $y = \mathrm{proj}_S(x)$, then $x - y \perp S$, meaning that $x - y \perp z, \forall z \in S$.

Let $P = \mathrm{proj}_S$ be a projection, where S is non-empty:

- $\|Px\| \leq \|x\|$
- P is *idempotent*, namely, $P^2 = P$, *i.e.* $P^2x = P(Px) = Px$.

Direct Sum and Orthogonal Complement

- A vector space Z is said to be the *direct sum* of two subspaces X and Y, denoted by $X \oplus Y$, if each $z \in Z$ can be expressed *uniquely* as $x + y$ with $x \in X$ and $y \in Y$.
- Let S be a subspace of Ω. The *orthogonal complement* of S, denoted by $S^\perp$, is defined to be $S^\perp \triangleq \{y \in \Omega : y \perp S\}$.
  - $S^\perp$ is also a subspace of Ω.
- *(Projection theorem)* $S \oplus S^\perp = \Omega$
  - Each $x \in \Omega$ can be expressed uniquely as $x = y + z$ with $y \in S$ and $z \in S^\perp$. Particularly, we have $y = \mathrm{proj}_S(x)$ and $z = x - y$.
- $S^{\perp\perp} = S$.

Orthonormal Basis of Hilbert Space

- Let Ω be a Hilbert space, a subset $M \subset \Omega$ is called a *total subset* of Ω if $\mathrm{span}(M)$ is dense.
  - M is a *total subset* of Ω if and only if $x \perp M \Rightarrow x = 0$.
  - **Question:** Why do we only require $\mathrm{span}(M)$ to be dense instead of $\mathrm{span}(M) = \Omega$?

Orthonormal Basis of Hilbert Space

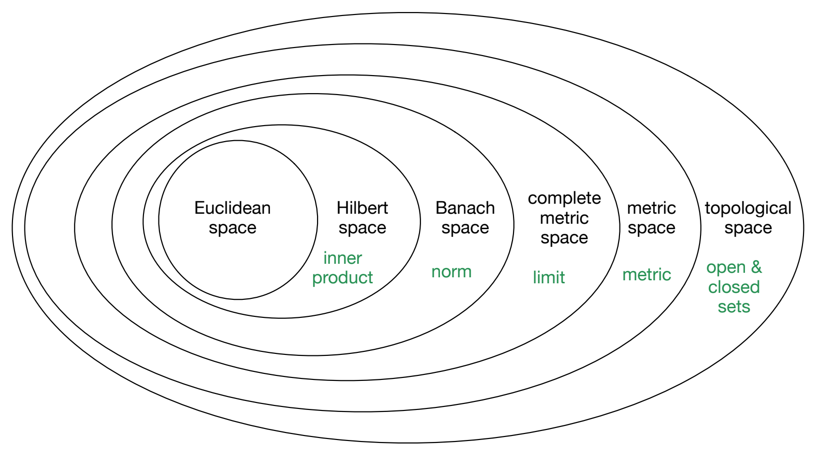
- Let Ω be a Hilbert space, a subset $M \subset \Omega$ is called a *total subset* of Ω if $\mathrm{span}(M)$ is dense.
  - M is a *total subset* of Ω if and only if $x \perp M \Rightarrow x = 0$.
  - **Question:** Why do we only require $\mathrm{span}(M)$ to be dense instead of $\mathrm{span}(M) = \Omega$?

- A *total orthonormal set* is called an *orthonormal basis*.
- *(Parseval theorem)* An orthonormal sequence (e_k) is *total* if and only if

$$\sum_k |\langle x, e_k\rangle|^2 = \|x\|^2, \quad \forall x \in \Omega$$

  In this case, each $x \in \Omega$ can be uniquely expressed as $x = \sum_k \langle x, e_k\rangle e_k$.
- Every Hilbert space has a *total orthonormal set*. Every *separable* Hilbert space has a *total orthonormal sequence*.

Spaces



Euclidean space
Hilbert space
inner product
Banach space
norm
complete metric space
limit
metric space
metric
topological space
open & closed sets

Linear Operators and Functionals

- Let X and Y be two linear spaces, a function: $T : X \to Y$ is called a *linear operator* if it preserves *linear dependency*, that is,

$$T(\alpha x_1 + \beta x_2) = \alpha T(x_1) + \beta T(x_2), \ \forall x_1, x_2 \in X, \alpha, \beta \in \mathbb{R}$$

  In particular, when Y is $\mathbb{R}$, T is called a *linear functional*.
- The set $\mathscr{N}(T) \triangleq \{x : T(x) = 0\}$ is a subspace of X, called the *null space* of T.

Bounded Operators

Let X and Y be normed spaces, and $T : X \to Y$ be a linear operator:

- T is said to be *bounded*, if $\exists c > 0, \forall x \in X, \|Tx\| \leq c\|x\|$.
- The *(operator) norm* of T is defined as

$$\|T\|_{op} \triangleq \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}$$

  Hence, $\|Tx\| \leq \|T\|_{op}\|x\|$ always holds.
- Given a linear operator T, the following statements are equivalent:
  - T is bounded
  - T is continuous
  - $\|T\|$ is finite
- All linear operators with finite dimensional domain are bounded.

Space of Bounded Operators and Dual Space

- All *bounded* operators $T : X \to Y$ form a *normed space*, denoted by $B(X, Y)$.
  - When Y is a Banach space, $B(X, Y)$ is a Banach space.
- All *bounded* functionals $f : X \to \mathbb{R}$ form a *normed space*, called the *dual space* of X, denoted by X^*, where the norm is called the *dual norm*, defined as:

  $$\|f\|_* = \sup_{x \neq 0} \frac{|f(x)|}{\|x\|}$$

  Note X^* is always a Banach space, no matter whether X is or not.
- X^* is generally *not* isomorphic to the X.

Function Space and Uniform Norm

- The set of all continuous real valued functions defined on a *compact space* Ω forms a normed vector space, denoted by $C(\Omega)$, where the norm is defined by $\|x\| = \sup_{v \in S} x(v)$, which is called the *uniform norm* (or *Chebyshev norm*).
  - When Ω is a closed interval $[a, b]$, $C(\Omega)$ is often written as $C[a, b]$.
- Given any compact space Ω, $C(\Omega)$ is a *Banach space*.
- There are other ways to define norms of functions, which we will discuss later.

Function Space and Uniform Norm

- The set of all continuous real valued functions defined on a *compact space* Ω forms a normed vector space, denoted by $C(\Omega)$, where the norm is defined by $\|x\| = \sup_{v \in S} x(v)$, which is called the *uniform norm* (or *Chebyshev norm*).
  - When Ω is a closed interval $[a, b]$, $C(\Omega)$ is often written as $C[a, b]$.
- Given any compact space Ω, $C(\Omega)$ is a *Banach space*.
- There are other ways to define norms of functions, which we will discuss later.

- **Question:** Let $C'[a, b]$ be a subspace of $C[a, b]$ that comprises all *continuously differentiable functions* on $[a, b]$. Let $m = (a + b)/2$. We define a functional as $f_m : x \mapsto x'(m)$ which takes the derivative at m.
  - Is f_m a linear functional?
  - Is f_m bounded?

Representation of Functionals

- Every Hilbert space is isomorphic to its *dual space*.

Representation of Functionals

- Every Hilbert space is isomorphic to its *dual space*.
- *(Riesz's theorem)* Every *bounded* linear functional f on a Hilbert space Ω can be represented in terms of inner product, namely, $f(x) = \langle x, z \rangle$, where z is uniquely determined by f and has $\|z\| = \|f\|_*$.

Representation of Functionals

- Every Hilbert space is isomorphic to its *dual space*.
- *(Riesz's theorem)* Every *bounded* linear functional f on a Hilbert space Ω can be represented in terms of inner product, namely, $f(x) = \langle x, z \rangle$, where z is uniquely determined by f and has $\|z\| = \|f\|_*$.
- **Question:** How can you find z given f?

Representation of Functionals

- Every Hilbert space is isomorphic to its *dual space*.
- *(Riesz's theorem)* Every *bounded* linear functional f on a Hilbert space Ω can be represented in terms of inner product, namely, $f(x) = \langle x, z \rangle$, where z is uniquely determined by f and has $\|z\| = \|f\|_*$.
- **Question:** How can you find z given f?
- In a separable Hilbert space, z can be expressed as a series:

$$z = \sum_k f(e_k) e_k$$

Bilinear Forms

- Let X and Y be Hilbert spaces. Then the function $h : X \times Y \to \mathbb{R}$ is called a *bilinear form* if it is linear *w.r.t* each argument with the other fixed.
- Let h be a *bilinear form* on $X \times Y$, then h is said to be *bounded* if there exists $c > 0$ with $|h(x, y)| \leq c\|x\|\|y\|, \forall x \in X, y \in Y$. In such case, the norm of h is defined to be

$$\|h\| = \sup_{x \neq 0, y \neq 0} \frac{|h(x, y)|}{\|x\|\|y\|}$$

  Hence, we always have $|h(x, y)| \leq \|h\|\|x\|\|y\|$.
- *(Generalized Riesz's theorem)* Let h be a *bounded* bilinear form on $X \times Y$. Then h has a representation as $h(x, y) = \langle Sx, y\rangle$, where S is a bounded linear operator uniquely determined by h and have $\|S\|_{op} = \|h\|$.

Measure Theory

- *Measure theory* studies *assigning values to subsets*.
- Measure theory is the corner stone of many math subjects
  - Modern approach to integration is based on measure theory.
  - Modern probability theory is based entirely on measure theory.
- Intuitively, the *measure* of a set can be intepreted as the *size*, *e.g.* the length of an interval.

Measure Theory

- *Measure theory* studies *assigning values to subsets*.
- Measure theory is the corner stone of many math subjects
  - Modern approach to integration is based on measure theory.
  - Modern probability theory is based entirely on measure theory.
- Intuitively, the *measure* of a set can be intepreted as the *size*, *e.g.* the length of an interval.
- Why is this so challenging?

Lengths of Real Subsets

- How long are these subsets: (a, b), $[a, b]$, and $\{a\}$?

Lengths of Real Subsets

- How long are these subsets: (a, b), $[a, b]$, and $\{a\}$?
- What is the length of $\mathbb{N}$?

Lengths of Real Subsets

- How long are these subsets: (a, b), $[a, b]$, and $\{a\}$?
- What is the length of $\mathbb{N}$?
- Now let's try to compute the length of (a, b) by *decomposing it into infinitely many points*:

$$\sum_{x \in (a,b)} \operatorname{len}(x) = \sum_{x \in (a,b)} 0 = \cdots ?$$

Lengths of Real Subsets

- How long are these subsets: (a, b), $[a, b]$, and $\{a\}$?
- What is the length of $\mathbb{N}$?
- Now let's try to compute the length of (a, b) by *decomposing it into infinitely many points*:

$$\sum_{x \in (a,b)} \mathrm{len}(x) = \sum_{x \in (a,b)} 0 = \cdots ?$$

- Sizes (*e.g.* lengths, areas, and volumes) are *countably additive*.
- Let $\mathcal{C}$ be a collection of subsets of Ω, then a function $f : \mathcal{C} \to \mathbb{R}$ is said to be *countably additive*, if for any *finite* or *countable* sequence of *disjoint* subsets $A_1, A_2, \ldots \in \mathcal{A}$, it has

$$f\left(\bigcup_i A_i\right) = \sum_{i=1}^{n} f(A_i)$$

.

The Vitali Set

- Let's consider a very interesting subset of $[0,1]$, called *Vitali set*, constructed as follows:
  - We say x and y are equivalent, if $x - y \in \mathbb{Q}$.
  - $[0,1]$ can then be partitioned into multiple equivalent classes.
  - There are *incountably infinite* such equivalent classes
  - By *axiom of choices*, we can form a set V, which contains *one* representative from each equivalent class.

The Vitali Set

- Let's consider a very interesting subset of $[0, 1]$, called *Vitali set*, constructed as follows:
  - We say x and y are equivalent, if $x - y \in \mathbb{Q}$.
  - $[0, 1]$ can then be partitioned into multiple equivalent classes.
  - There are *incountably infinite* such equivalent classes
  - By *axiom of choices*, we can form a set V, which contains *one* representative from each equivalent class.
- **Question:** the length of V?

The Vitali Set

- Let's consider a very interesting subset of $[0, 1]$, called *Vitali set*, constructed as follows:
  - We say x and y are equivalent, if $x - y \in \mathbb{Q}$.
  - $[0, 1]$ can then be partitioned into multiple equivalent classes.
  - There are *incountably infinite* such equivalent classes
  - By *axiom of choices*, we can form a set V, which contains *one* representative from each equivalent class.

- **Question:** the length of V?

- Try to derive the length of V:
  - enumerate all rational numbers within $[0, 1]$ as $q_1, q_2, \ldots$, and let $V_k = V + q_k$
  - Obviously, $V_1, V_2, \ldots$ are disjoint
  - Let $U = \bigcup_k V_k$, then $[0, 1] \subset U \subset [-1, 2]$, and therefore $1 \leq \mathrm{len}(U) \leq 3$.
  - However, if $\mathrm{len}(V) = 0$, then $\mathrm{len}(U) = 0$, or if $\mathrm{len}(V) > 0$, then $\mathrm{len}(U) = \infty \ldots$

σ-algebra

- $\mathcal{S}$, a nonempty collection of subsets of Ω, is called a *ring*, if it is closed under *union* and *set difference*.
  - A ring must contain $\emptyset$.
  - A ring is closed under *intersection*.
- A ring $\mathcal{S}$ is called a *algebra*, if it is also closed under *complement*.
  - An *algebra* must contain Ω.
- An *algebra* is called a σ-algebra, if it is also closed under *countable union*.
  - A σ-algebra is closed under countable fold of any elementary set operations.

Measure and Measure Space

- A countably additive function μ from a σ-algebra $\mathcal{S}$ into $[0, \infty]$ is called a *measure*. Then $(\Omega, \mathcal{S}, \mu)$ is called a *measure space*. All the elements of $\mathcal{S}$ are called *measurable sets*.

Measure and Measure Space

- A countably additive function μ from a σ-algebra $\mathcal{S}$ into $[0, \infty]$ is called a *measure*. Then $(\Omega, \mathcal{S}, \mu)$ is called a *measure space*. All the elements of $\mathcal{S}$ are called *measurable sets*.
- Consider a *measure space* $(\Omega, \mathcal{S}, \mu)$:
  - $\mu(\emptyset) = 0$.
  - $A \subset B \Rightarrow \mu(A) \leq \mu(B)$.
  - Let $A_1, A_2, \ldots \in \mathcal{S}$ be disjoint, then $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.
  - μ is called a *finite measure* if $\mu(\Omega) < \infty$.
  - μ is called a *σ-finite measure* if Ω is covered by countably many subsets of finite measure.
  - Continuity:
    - $A_1 \subset A_2 \subset \ldots \Rightarrow \mu(A_i) \uparrow \mu\left(\bigcup_i A_i\right)$.
    - $A_1 \supset A_2 \supset \ldots \Rightarrow \mu(A_i) \downarrow \mu\left(\bigcap_i A_i\right)$.

Open Intervals and Their Lengths

To define a measure over $\mathbb{R}$. Let's first consider the simplest cases – the intervals.

- The length of an open interval is defined as $\mathrm{len}((a, b)) = b - a$.
- The the length of a finite union of disjoint open intervals is defined to be the sum of the lengths of individual intervals.
- The collection of all finite unions of disjoint open intervals constitutes a *ring*.
    - The length is well defined over this ring, and it satisfies finite additivity. – Hence, it is called a *pre-measure*.
- Now, we want to extend the length to as many subsets as possible.

Borel σ-algebra

Let's first extend the open interval ring to a σ-algebra:

- Let $\mathcal{C}$ be a collection of subsets. Then the smallest σ-algebra that contains $\mathcal{C}$ is called the *σ-algebra generated by $\mathcal{C}$*.
- Let (Ω, d) be a *metric space*, then the σ-algebra generated by the open subsets of Ω is called the *Borel σ-algebra* of Ω, denoted by $\mathcal{B}$
  - Elements of the Borel σ-algebra are called *Borel sets*.
  - All the open sets, closed sets, and their finite and countable unions and intersections are all *Borel sets*.
  - Finite or countable sets are all *Borel sets*.

Measure Extension

The next step is to extend the *measure*:

- *(Hahn–Kolmogorov theorem):* Let $\mathcal{C}$ be a ring that covers Ω, $\mu' : \mathcal{C} \to [0, \infty]$ be a *pre-measure*, and $\mathcal{S}$ be the σ-algebra generated by $\mathcal{C}$, then μ' can be extended to a measure $\mu : \mathcal{S} \to [0, \infty]$. If μ is σ-finite, then the extension is unique.
- The length function over the open interval ring can be *uniquely extended* to a measure over the Borel σ-algebra over $\mathbb{R}$, called *Borel measure*.

Lebesgue Measure

We are not done yet . . .

- Given a measure space $(\Omega, \mathcal{S}, \mu)$, a subset $A \subset \Omega$ is called a *null set* if there exists $B \in \mathcal{S}$ with $\mu(B) = 0$ such that $A \subset B$.
- This *measure space* is called a *complete measure space* if every null set is measurable.
- Let $\mathcal{N}(\mu)$ denote the collection of all the *null sets*. Then the σ-algebra generated by $\mathcal{S} \cup \mathcal{N}(\mu)$ is *complete*.
- The measure can be straightforwardly extended. The resultant measure space is called the *completion* of $(\Omega, \mathcal{S}, \mu)$.

Lebesgue Measure

We are not done yet . . .

- Given a measure space $(\Omega, \mathcal{S}, \mu)$, a subset $A \subset \Omega$ is called a *null set* if there exists $B \in \mathcal{S}$ with $\mu(B) = 0$ such that $A \subset B$.
- This *measure space* is called a *complete measure space* if every null set is measurable.
- Let $\mathcal{N}(\mu)$ denote the collection of all the *null sets*. Then the σ-algebra generated by $\mathcal{S} \cup \mathcal{N}(\mu)$ is *complete*.
- The measure can be straightforwardly extended. The resultant measure space is called the *completion* of $(\Omega, \mathcal{S}, \mu)$.

- The *Borel measure space* is generally not *complete*. The *completion* of the *Borel measure space* is called the *Lebesgue measure space*, and the extended measure is called *Lebesgue measure*.

Lebesgue Integral: Roadmap

We derive the *Lebesgue integral* in a way that begins with simple functions and then extends the definition to more complicated functions:

Lebesgue Integral: Roadmap

We derive the *Lebesgue integral* in a way that begins with simple functions and then extends the definition to more complicated functions:
Let $(\Omega, \mathcal{S}, \mu)$ be a measure space:

- Indicator functions – the integral of an indicator function 1_A of a *measurable* set $A \in \mathcal{S}$ is defined to be the *measure* of A, as

$$\int 1_A d\mu = \mu(A)$$

Lebesgue Integral: Roadmap

We derive the *Lebesgue integral* in a way that begins with simple functions and then extends the definition to more complicated functions:
Let $(\Omega, \mathcal{S}, \mu)$ be a measure space:

- Indicator functions – the integral of an indicator function 1_A of a *measurable* set $A \in \mathcal{S}$ is defined to be the *measure* of A, as

$$\int 1_A d\mu = \mu(A)$$

- Simple functions – linear combinations of indicator functions:

$$\int \sum_k \alpha_k 1_{A_k} d\mu = \sum_k \alpha_k \int 1_{A_k} d\mu = \sum_k \alpha_k \mu(A_k)$$

Lebesgue Integral: Roadmap

We derive the *Lebesgue integral* in a way that begins with simple functions and then extends the definition to more complicated functions:
Let $(\Omega, \mathcal{S}, \mu)$ be a measure space:

- Indicator functions – the integral of an indicator function 1_A of a *measurable* set $A \in \mathcal{S}$ is defined to be the *measure* of A, as

$$\int 1_A d\mu = \mu(A)$$

- Simple functions – linear combinations of indicator functions:

$$\int \sum_k \alpha_k 1_{A_k} d\mu = \sum_k \alpha_k \int 1_{A_k} d\mu = \sum_k \alpha_k \mu(A_k)$$

- Non-negative functions – through approximation
- Signed functions – decompose as $f = f_+ - f_-$.

Lebesgue Integral (cont'd)

- Let s be a *non-negative* function on Ω, then we define

$$\int f d\mu = \sup \left\{ \int s d\mu : 0 \leq s \leq f, s \text{ simple} \right\}$$

Lebesgue Integral (cont'd)

- Let s be a *non-negative* function on Ω, then we define

$$\int f d\mu = \sup \left\{ \int s d\mu : 0 \leq s \leq f, s \text{ simple} \right\}$$

- Signed function: decompose f as $f = f_+ - f_-$, with $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$, then

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu$$

  When both $\int f_+ d\mu$ and $\int f_- d\mu$ are infinite, $\int f d\mu$ is *undefined*.

Properties of Lebesgue Integral

- $f \mapsto \int f d\mu$ is a linear functional.
- If $f = g$ *a.e.* then $\int f d\mu = \int g d\mu$ (Note: a predicate holds *almost everywhere* means that it holds over Ω, except for a null set)
- *(Monotonicity)* $f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$.
- *(Monotone convergence theorem)* Let (f_k) be a sequence of *non-negative* functions, and $f_k \uparrow f$ pointwisely, then $\int f_k d\mu \uparrow \int f d\mu$.
- *(Dominated convergence theorem)* If $f_k \to f$ pointwisely, (f_k) is *dominated* by g, *i.e.* $\forall k\ |f_k| \leq g$, and $\int g d\mu < \infty$, then $\int f_k d\mu \to \int f d\mu$.

Examples

- The *counting measure* over a *finite* or *countable* space Ω is defined as the *cardinality* of subsets.
- Let μ be a *counting measure* over a countable space $\Omega = \{x_k\}$, then

$$\int f d\mu = \sum_k f(x_k)$$

Examples

- The *counting measure* over a *finite* or *countable* space Ω is defined as the *cardinality* of subsets.
- Let μ be a *counting measure* over a countable space $\Omega = \{x_k\}$, then

$$\int f d\mu = \sum_k f(x_k)$$

- Let f be a function over $[a, b]$, then when f is *Riemann integrable*, f must be *Lebesgue integrable* (*w.r.t.* the *Lebesgue measure*), and in such a case, the *Lebesgue integral* is equal to the *Riemann integral*.

Examples

- The *counting measure* over a *finite* or *countable* space Ω is defined as the *cardinality* of subsets.
- Let μ be a *counting measure* over a countable space $\Omega = \{x_k\}$, then

$$\int f d\mu = \sum_k f(x_k)$$

- Let f be a function over $[a, b]$, then when f is *Riemann integrable*, f must be *Lebesgue integrable* (*w.r.t.* the *Lebesgue measure*), and in such a case, the *Lebesgue integral* is equal to the *Riemann integral*.
- **Question:** Consider $1_{\mathbb{Q}}$:
  - Is it *Riemann integrable*?
  - Compute the *Lebesgue integral* of $1_{\mathbb{Q}}$ with respect to the *Lebesgue measure* μ.

L^p space

Given a measure space $(\Omega, \mathcal{S}, \mu)$:

- For any $p \geq 1$, the L^p space of Ω, denoted by $\mathcal{L}^p(\Omega)$, is the vector space comprised of all *measurable* functions on Ω such that $\int |f|^p d\mu < \infty$, where the norm is defined by

$$\|f\|_p \triangleq \left(\int |f|^p d\mu \right)^{1/p}$$

  Here, $\|\cdot\|_p$ is called the L^p *norm*.
- **Question:** Is $\|\cdot\|_p$ actually a *norm*?

L^p space

Given a measure space $(\Omega, \mathcal{S}, \mu)$:

- For any $p \geq 1$, the L^p space of Ω, denoted by $\mathcal{L}^p(\Omega)$, is the vector space comprised of all *measurable* functions on Ω such that $\int |f|^p d\mu < \infty$, where the norm is defined by

$$\|f\|_p \triangleq \left(\int |f|^p d\mu \right)^{1/p}$$

  Here, $\| \cdot \|_p$ is called the L^p *norm*.
- **Question:** Is $\| \cdot \|_p$ actually a *norm*?
- Functions that are equal *almost everywhere* are indistinguishable by integration. Let $L^p(\Omega)$ denotes the vector space with all *essentially equivalent* functions merged into an element. Then L^p is a normed vector space with the L^p.

L^p space (cont'd)

- What about when $p = \infty$?

L^p space (cont'd)

- What about when $p = \infty$?
- The L^∞ norm is defined as the *essential supreme* of $|f|$, as

$$\|f\|_\infty = \inf\{c : |f(x)| \le c,\ a.e.\}$$

- For any $p \in [1, +\infty]$, L^p is complete, implying that L^p is a Banach space.
- The dual space of L^p is isomorphic to L^q with $p^{-1} + q^{-1} = 1$. For each bounded functional on $h \in L^p$, there exists $g \in L^q$, such that

$$h(f) = \int fgd\mu,\ \forall f \in L^p$$

L^p space (cont'd)

- What about when $p = \infty$?
- The L^∞ norm is defined as the *essential supreme* of $|f|$, as

$$\|f\|_\infty = \inf\{c : |f(x)| \leq c, \ a.e.\}$$

- For any $p \in [1, +\infty]$, L^p is complete, implying that L^p is a Banach space.
- The dual space of L^p is isomorphic to L^q with $p^{-1} + q^{-1} = 1$. For each bounded functional on $h \in L^p$, there exists $g \in L^q$, such that

$$h(f) = \int fgd\mu, \ \forall f \in L^p$$

- L^2 is a Hilbert space, where the inner product is defined as

$$\langle f, g \rangle \triangleq \int fgd\mu$$

Absolute Continuity

Let μ and ν be two measures over a measurable space $(\Omega, \mathcal{S})$

- ν is said to be *absolutely continuous* with respect to μ, denoted by $\nu \prec \mu$, if $\nu(A) = 0$ whenever $\mu(A) = 0$.
- μ and ν are said to be *singular*, denoted by $\mu \perp \nu$, if there exists $A \in \mathcal{S}$ such that $\mu(A) = 0$ and $\nu(\Omega \backslash A) = 0$.
- *(Lebesgue Decomposition)* there exists a unique decomposition of ν as $\nu = \nu_{ac} + \nu_s$ such that $\nu_{ac} \prec \mu$ and $\nu_s \perp \mu$.

Radon-Nikodym

(Radon-Nikodym theorem) Let ν be a finite measure that is *absolutely continuous* with respect to μ. Then there exists an *essentially* unique $h \in L^1(\Omega, \mathcal{S}, \mu)$ such that

$$\nu(A) = \int_A h d\mu, \ \forall A \in \mathcal{S}$$

Here, ν is called the *Radon-Nikodym density* of ν with respect to μ. In this case, we also have

$$\int f d\nu = \int f h d\mu, \ \forall f \in L^1(\Omega, \mathcal{S}, \nu)$$

Probability Measure

- A *probability measure* P is a measure on $(\Omega, \mathcal{S})$ with $P(\Omega) = 1$.
- The σ-algebra $\mathcal{S}$ can be interpreted as an *event space*:
  - Each element $A \in \mathcal{S}$, which is a subset of Ω, can be considered as an *event*.
  - $A \cap B$ means *both A and B happen
  - $A \cup B$ means either A or B happens
  - A^c means A does not happen
  - $A \cap B = \emptyset$ means that A and B are *mutually exclusive*.
- $P(A)$ can be considered as the *probability* of the event A.
- Obviously, P satisfies all the requirement of probability functions in classical probability theory.

Probability Measure

- A *probability measure* P is a measure on $(\Omega, \mathcal{S})$ with $P(\Omega) = 1$.
- The σ-algebra $\mathcal{S}$ can be interpreted as an *event space*:
  - Each element $A \in \mathcal{S}$, which is a subset of Ω, can be considered as an *event*.
  - $A \cap B$ means *both A and B happen
  - $A \cup B$ means either A or B happens
  - A^c means A does not happen
  - $A \cap B = \emptyset$ means that A and B are *mutually exclusive*.
- $P(A)$ can be considered as the *probability* of the event A.
- Obviously, P satisfies all the requirement of probability functions in classical probability theory.
- Two events A and B are said to be *independent* if $P(A \cap B) = P(A)P(B)$.
- We say two collections of events $\mathcal{A}$ and $\mathcal{B}$ are *independent* when $P(A \cap B) = P(A)P(B)$ $\forall A \in \mathcal{A}, B \in \mathcal{B}$.

Random Variables

- A *(real-valued) random variable* X is defined to be a *measurable function* $X : \Omega \to \mathbb{R}$, meaning that for any Borel set $A \in \mathcal{B}(R)$, $X^{-1}(A)$ is measurable, *i.e.* $X^{-1}(A) \in \mathcal{S}$.
- The *expectation* of a random variable X is defined to be

$$E[X] = \int X dP$$

- The *expectation* is a *bounded linear functional*.

Random Variables

- A *(real-valued) random variable* X is defined to be a *measurable function* $X : \Omega \to \mathbb{R}$, meaning that for any Borel set $A \in \mathcal{B}(R)$, $X^{-1}(A)$ is measurable, *i.e.* $X^{-1}(A) \in \mathcal{S}$.
- The *expectation* of a random variable X is defined to be

$$E[X] = \int X dP$$

- The *expectation* is a *bounded linear functional*.

- Each random variable X induces a sub-σ-algebra, denoted by $X^{-1}(\mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$.
- Two random variables X and Y are said to be *independent* if $X^{-1}(\mathcal{B})$ and $Y^{-1}(\mathcal{B})$ are independent.

Probability Density

- The probability that the value of X falling in A is then given by $P(X^{-1}(A))$.
  - The function $P \circ X^{-1}$ that maps each Borel set $A \subset \mathbb{R}$ to a *probability value* itself is a *probability measure* over $\mathbb{R}$, which is called the *law* of X, denoted by $X_* P$.
- The Radon-Nikodym density of $X_* P$ with respect to the Lebesgue measure over $\mathbb{R}$, denoted by f, is called the *probability density* of X. So we have $f d\mu = dX_* P$.

(Semantic) Correspondence

measure theory	probability theory
σ-algebra	event space
sub-σ-algebra	a family of events
measure P	probability function
measurable function X	random variable
$X_* P$	law of X
integration	expectation
Radon-Nikodym density of $X_* P$	probability density

Convex Analysis

- Convex optimization plays a crucial role in many machine learning problems
- Prof. *Stephen Boyd* has a very famous book *"Convex Optimization"*, which provides an excellent treatment of this subject
  - The PDF version is available for download in Prof. Boyd's website
  - You should at least read the first five chapters

Convex Analysis

- Convex optimization plays a crucial role in many machine learning problems
- Prof. *Stephen Boyd* has a very famous book *"Convex Optimization"*, which provides an excellent treatment of this subject
  - The PDF version is available for download in Prof. Boyd's website
  - You should at least read the first five chapters
- Important Concepts in convex analysis
  - convex set
  - convex function
  - conjugate function
  - Lagrange dual
  - Strong duality and weak duality
- We will begin to use these concepts in Lecture 4.