Notes of Statistical Learning

Nanyi, UIBE

September 30, 2022

This is not to b123e read as a PDF file. Ideally, you will look at 123123the *.tex file and the *.pdf file side-by-side123 to read the comments and see the commands that created the PDF file.

Contents

Chapter 1

Measure, Integration and Analysis

1.1 Real Measures

1.2 Integration

1.3 Decomposition Theorems

1.4 Fourier Analysis

Chapter 2

Banach Spaces and Hilbert Spaces

Normed vector spaces and Banach spaces capture the notion of distance. In this chapter we introduce inner product spaces, which capture the notion of angle. The concept of orthogonality plays a particularly important role in inner product spaces.

Hilbert spaces are named in honor of David Hilbert(1862-1943), who helped develop parts of the theory in the early twentieth century.

In this chapter, we will see a clean description of the bounded linear functionals on a Hilbert space. We will also see that every Hilbert space has an orthonormal basis, which make Hilbert spaces look much like standard Euclidean spaces but with infinite sums replacing finite sums.

2.1 Metric Spaces

Definition 2.1. A complete normed vector space is called Banach space.

Theorem 2.1. V is a Banach space if and only if converge absolutely impies converge.

$$V \text{ is a Banach space } \iff \sum_{k=1}^{\infty} \|g_k\| < \infty \text{ implies } \sum_{k=1}^{\infty} g_k \text{ converges.}$$

Theorem 2.2 ($B(V, W)$ is a Banach space if W is a Banach space)**.**
Suppose V is a normed vector space and W is a Banach space. Then $B(V, W)$ is a Banach space.

Theorem 2.3 (continuity is equivalent to boundedness for linear maps)**.** 123

2.2 Inner Product Spaces

2.3 Orthogonality

2.4 Conditional Expectation as Projection

Chapter 3

Linear Maps on Hilbert Spaces

3.1 Adjoint

Definition 3.1 (adjoint;T^*)**.**
Suppose V and W are Hilbert spaces and $T : V \to W$ is a bounded linear map. The adjoint of T is the function $T^* : W \to V$ such that

$$\langle Tf, g\rangle = \langle f, T^*g\rangle$$

for every $f \in V$ and every $g \in W$.

To see why the definiton above makes sense, fix $g \in W$. Consider the linear functional on V defined by $\varphi_g : f \mapsto \langle Tf, g\rangle$. This linear functional is bounded and hence by the Riesz Representation Theorem, there exist a unque element h_g of V such that

$$\varphi_g(f) = \langle Tf, g\rangle = \langle f, h_g\rangle,\ \|h_g\| = \|\varphi_g\|.$$

The procedure induces a linear map $g \mapsto h_g$, which is called the adjoint of T ,denoted by T^*.

Theorem 3.1 (Properties of adjoint)**.**
Suppose V and W are Hilbert spaces and $T \in B(V, W)$, T^* is the adjoint T.

- $T^* \in B(W, V)$; $(T^*)^* = T$; $\|T^*\| = \|T\|$
- $(S + T)^* = S^* + T^*$; $(aT)^* = \bar{a}T^*$
- $I^* = I$
- $(ST)^* = T^*S^*$

Theorem 3.2 (null space and range of T^*)**.**
Suppose V, W are Hilbert spaces and $T \in B(V, W)$, Then
(a) null $T^* =$ (range T)$^\perp$
(b) $\overline{\text{range } T^*} = (\text{null } T)^\perp$
(c) null $T =$ (range T*)$^\perp$
(d) $\overline{\text{range } T} = (\text{null } T^*)^\perp$

Corollary 3.1.1. T^* is injective if and only if $\overline{\text{range } T} = W$

Definition 3.2 (Invertible)**.** 123123

Theorem 3.3. Suppose T is an operator on a Banach space V, then

$$(I - T)^{-1} = \sum_{k=0}^{\infty} T^k \ if \ \|T\| < 1.$$

Theorem 3.4 (Inverible operators form an open set)**.**
Suppose V is a Banach space. Then $\{T \in B(V) :$ T is invertible$\}$ is an open subset of $B(V)$.

Theorem 3.5. Suppose V is a Hilbert space and $T \in B(V)$. Then the followings are equivalent
(a) T is left invertible.
(b) $\exists\ a \in (0, \infty)$ such that $\|f\| \leq a\|Tf\|$ for all $f \in V$
(c) T is injective and has closed range.
(d) T^*T is invertible.

Theorem 3.6. Suppose V is a Hilbert space and $T \in B(V)$. Then the followings are equivalent
(a) T is right invertible.
(b) T is surjective.
(c) TT^* is invertible.

3.2 Spectrum

Definition 3.3. Suppose T is a bounded operator on a Banach space V.

- A number α is called an eigenvalue of T is $T - \alpha I$ is not injective.
- A nonzero vector $f \in V$ is called an eigenvector of T corresponding to an eigenvalue $\alpha \in F$ if

$$f \in \text{null } T - \alpha I$$

- The spectrum of T is denoted by $\mathrm{sp}(T)$ and is defined by

$$\mathrm{sp}(T) = \{\alpha \in F : T - \alpha I \text{ is not invertible}\}.$$

Remark 1. If V is a finite-dimensional Banach space and $T \in B(V)$, the fundamental theorem of linear maps induces that $T - \alpha I$ is not injective if and only if $T - \alpha I$ is not invertible. Thus if T is an operator on a finite-dimensional Banach space, then the spectrum of T equals eigenvalues of T.

Theorem 3.7 ($\mathrm{sp}(T)$ is closed)**.**
Suppose T is a bounded operator on a Banach space V, then $\mathrm{sp}(T)$ is a closed subset of V.

Proof. 123 □

Theorem 3.8 ($T - \alpha I$ is invertible for $|\alpha|$ large)**.**
Suppose T is a bounded operator on a Banach space. Then
(a) $\mathrm{sp}(T) \subset \{\alpha \in \mathbb{F} : |\alpha| < \|T\|\}$
(b) $T - \alpha I$ is invertible for all $\alpha \in F$ with $|\alpha| > \|T\|$
(c) $\lim_{\alpha \to \infty} \|(T - \alpha I)^{-1}\| = 0$

Proof. We begin by proofing (b). Suppose $|\alpha| > \|T\|$. Then

$$T - \alpha I = -\alpha \left(I - \frac{T}{\alpha} \right).$$

Because $\|T/\alpha\| < 1$, which implied that $T - \alpha I$ is invertible. □

Theorem 3.9. Suppose T is a bounded operator on a complex Hilbert space V. Then the function

$$\varphi(\alpha; T, f, g) : \alpha \mapsto \langle (T - \alpha I)^{-1} f, g \rangle$$

is analytic on $\mathbb{C}/\mathrm{sp}(T)$ for all $f, g \in V$.

Theorem 3.10 (Spectrum is nonempty)**.** 123

Definition 3.4 (self-adoint)**.** A bounded operator T on a Hilbert space is called self-adjoint if $T^* = T$.

Remark 2. On a finite-dimensional Hilbert space V, the matrix of T^* with respect to an orthonormal basis is the conjugate tranpose of that of T. Thus the matrix of a self-adjoint operator T has the following form:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \overline{a_{12}} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & a_{nn} \end{bmatrix}$$

Remark 3 (self-adjoint and complex conjugate)**.** Some insight into the adjoint can be obtained by thinking the operation $T \mapsto T^*$ on $B(V)$ as analogous to the operation $z \mapsto \overline{z}$ on $\mathbb{C}$. The following facts illustrate this analogy.

Definition 3.5 (Isometry and Unitary Operator)**.**
Suppose T is a bounded operator on a Hilbert space V.

- T is called an isometry if $\|Tf\| = \|f\|$ for every $f \in V$.
- T is called unitary if $T^*T = TT^* = I$.

Theorem 3.11 (Unitary operators and their adjoints are isometries)**.**
Suppose T is a bounded operator on a Hilbert space V. Then the followings are equivalent:
(a) T is unitary.
(b) T is an isometry and surjection.
(c) T and T^* are both isometries.
(d) T^* is unitary.
(e) T is invertible and $T^{-1} = T^*$.
(f) $\{Te_k\}_{k\in\Gamma}$ is an orthonormal basis of V for every orthonormal basis $\{e_k\}_{k\in\Gamma}$.
(g) $\{Te_k\}_{k\in\Gamma}$ is an orthonormal basis of V for some orthonormal basis $\{e_k\}_{k\in\Gamma}$.

Chapter 4

Matrix Calculus

4.1 Basic Notations

suppose $\mathbf{x}$ and $\mathbf{a} \in \mathbb{R}^n$, and $(\mathbb{R}^n, \langle\rangle)$ is a Hilbert space, then we have

$$\langle \mathbf{x}, \mathbf{a} \rangle = \sum_{i=1}^{n} a_i x_i = \mathbf{a}^\mathsf{T}\mathbf{x}$$

Also, suppose f is map from $V(\mathbb{R}^n)$ to $W(\mathbb{R}^m)$. We apply the following notations as usual.

- f if $W = \mathbb{F}^1$
- $\mathbf{f}$ if $W = \mathbb{F}^m$
- $\mathbf{F}$ if $W = \mathbb{F}^{m\times n}$
- x if $V = \mathbb{F}^1$
- $\mathbf{x}$ if $V = \mathbb{F}^n$
- $\mathbf{X}$ if $V = \mathbb{F}^{m\times n}$

4.2 Derivatives of first-order

We start with m = 1 and n $\neq$ 1. Define $\frac{\partial}{\partial \mathbf{x}} : F \to F^{n\times 1}$ by

$$\frac{\partial}{\partial \mathbf{x}} f := \begin{bmatrix} \frac{\partial}{\partial x_1} f \\ \frac{\partial}{\partial x_2} f \\ \vdots \\ \frac{\partial}{\partial x_n} f \end{bmatrix}, \quad \frac{\partial}{\partial \mathbf{x}^\mathsf{T}} f := (\frac{\partial}{\partial \mathbf{x}} f)^\mathsf{T} = \begin{bmatrix} \frac{\partial}{\partial x_1} f & \frac{\partial}{\partial x_2} f & \cdots & \frac{\partial}{\partial x_n} f \end{bmatrix} \tag{4.1}$$

Similarly, for m ! = 0 and $n! = 0$. Define $\frac{\partial}{\partial \mathbf{x}}$: by

$$\frac{\partial}{\partial \mathbf{x}^{\mathsf{T}}}\mathbf{f} := 1 = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1 & \frac{\partial}{\partial x_2} f_1 & \cdots \frac{\partial}{\partial x_n} f_1 \\ \frac{\partial}{\partial x_1} f_2 & \frac{\partial}{\partial x_2} f_2 & \cdots \frac{\partial}{\partial x_n} f_2 \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_1} f_m & \frac{\partial}{\partial x_2} f_m & \cdots \frac{\partial}{\partial x_n} f_m \end{bmatrix} \tag{4.2}$$

4.3 Differentiation

Definition 4.1 (Derivatives)**.** suppose f is a map from $(\mathbb{R}, d_{\mathbb{R}})$ to (V, d_V). Then the derivative of f at x_0 is define by

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

A direct consequence of this definition is $f'(x_0) \in V$.

Definition 4.2 (Differentiation)**.** Suppose f is a map and $x_0 \in \mathbb{R}$. In order to approximate $f(x) - f(x_0)$ with a linear map. We define

$$\mathrm{d}f(x_0) : \mathbb{R} \to \mathbb{R},\ h \mapsto f'(x_0)h$$

if f is a map from $\mathbb{R}$ to $\mathbb{R}$. Similarly, $\mathbf{f}$ can be defined by

$$\mathrm{d}f(\mathbf{x_0}) : \mathbb{R}^n \to \mathbb{R},\ \mathbf{h} \mapsto \langle \mathbf{h}, \frac{\partial}{\partial \mathbf{x_0}} f \rangle.$$

Also,

$$\mathrm{d}\mathbf{f}(\mathbf{x_0}) : \mathbb{R}^n \to \mathbb{R}^m,\ \mathbf{h} \mapsto \begin{bmatrix} \langle \mathbf{h}, \frac{\partial}{\partial \mathbf{x_0}} f_1 \rangle \\ \langle \mathbf{h}, \frac{\partial}{\partial \mathbf{x_0}} f_2 \rangle \\ \vdots \\ \langle \mathbf{h}, \frac{\partial}{\partial \mathbf{x_0}} f_m \rangle \end{bmatrix}$$

However, the definition above does not capture the essence of differentiation. In fact, differentiation operator is a map from $B(V, W) \to B(V, W)$(given x_0), where V and W are normed vector spaces. Now, suppose $x_0, v \in V$.

Definition 4.3 (Differentiation operator)**.** Suppose V and W are normed vector spaces, x_0 and v are vectors in V. If there exists a linear map S from V to W(an element of $B(V, W)$) such that

$$\lim_{v \to 0} \frac{\|T(x_0 + v) - T(x_0) - Sv\|}{\|v\|} = 0$$

Then we say T is differentiable at x_0 and S is the differentiation of T at x_0. For historical reasons, the differentiaion S is usually written as $dT(x_0)$.

Theorem 4.1 (Jacobian Matrix)**.** Suppose f is a map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then the matrix of differentiation of f at $x_0(\mathrm{d}f(x_0))$ with respect to the standard orthonormal basis is

$$\mathcal{M}(\mathrm{d}f(x_0)) = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x_0) & \frac{\partial}{\partial x_2} f_1(x_0) & \cdots \frac{\partial}{\partial x_n} f_1(x_0) \\ \frac{\partial}{\partial x_1} f_2(x_0) & \frac{\partial}{\partial x_2} f_2(x_0) & \cdots \frac{\partial}{\partial x_n} f_2(x_0) \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x_0) & \frac{\partial}{\partial x_2} f_m(x_0) & \cdots \frac{\partial}{\partial x_n} f_m(x_0) \end{bmatrix}$$

$\mathcal{M}(\mathrm{d}f(x_0))$ is also known as the Jacobian Matrix of $\mathbf{f}$.

Theorem 4.2 (Chain rules)**.** Suppose $f : U \to V$, $g : V \to W$. If f is differentiable at x_0 and g is differentiable at $f(x_0)$, then $g \circ f$ is differentiable at x_0 and

$$(\mathrm{d}(g \circ f))(x_0) = (\mathrm{d}g)(f(x_0)) \circ df(x_0).$$

Taking matrice of the equality above, we obtain

$$\mathcal{M}((\mathrm{d}(g \circ f))(x_0)) = \mathcal{M}((\mathrm{d}g)(f(x_0)) \circ df(x_0)) = \mathcal{M}((\mathrm{d}g)(f(x_0)))\mathcal{M}(\mathrm{d}f(x_0)).$$

Example 4.1 (dx)**.** In calculus, the notion dx is often used but never being strictly defined, thus this example attempts to given an accurate explaination of Total Differential Formula

$$\mathrm{d}f(x_0) = f'(x_0)\mathrm{d}x,\ \mathrm{d}f(\mathbf{x_0}) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\mathbf{x_0})\mathrm{d}x_i.$$

At first, we consider dx. By definition,

$$\mathrm{d}f(x_0):\ \mathbb{R} \to \mathbb{R},\ h \mapsto h.$$

Where f is a indentical map. From the equality above, we know that dx is also a indentical map. Easy to verify that

$$\mathrm{d}f(x_0) = f'(x_0)\mathrm{d}x.$$

This equality relationship often wriiten as $\frac{\mathrm{d}f(x_0)}{\mathrm{d}x} = f'(x_0)$.

Next, we concentrate on the case of multivariate. In this case, dx_i has different meanings in constract to $\mathrm{d}x$.

$$\mathcal{M}(\mathrm{d}f(\mathbf{x_0})) = \left[\tfrac{\partial}{\partial x_1}f(\mathbf{x_0}), \tfrac{\partial}{\partial x_2}f(\mathbf{x_0}), \cdots, \tfrac{\partial}{\partial x_n}f(\mathbf{x_0})\right].$$

x123

$$\mathcal{M}(\mathrm{d}x) := \mathcal{M}(\mathrm{d}I(\mathbf{x_0})) = \left[1, 1, \cdots, 1\right].$$

$\mathrm{d}x_i$ maps a vector its i-th slot. To be specific, suppose $v_k : k = 1, 2, \cdots, n$,is a basis of V. Suppose $v = \sum_{i=1}^{n} a_i v_i$, $\mathrm{d}x_i(v) = a_i$, and hence

$$\mathcal{M}(\mathrm{d}x_1) := \left[1, 0, \cdots, 0\right]$$

$$\mathrm{d}x = \sum_{i=1}^{n} \mathrm{d}x_i$$

$$\mathcal{M}(\sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\mathbf{x_0})\mathrm{d}x_i) = \left[\tfrac{\partial}{\partial x_1}f(\mathbf{x_0}), \tfrac{\partial}{\partial x_2}f(\mathbf{x_0}), \cdots, \tfrac{\partial}{\partial x_n}f(\mathbf{x_0})\right].$$

and thus,

$$\mathrm{d}f(\mathbf{x_0}) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\mathbf{x_0})\mathrm{d}x_i$$

At last, suppose $f : V \to W$, where V and W are normed vector spaces. We also suppose that

Example 4.2. Find the differentiation of $f(t, s) = t^2 + s^3$ at (t, s).

$$\mathrm{d}f(t, s) = 2t\mathrm{d}t + 3s^2\mathrm{d}s$$

where $\mathcal{M}(\mathrm{d}t) = \left[1, 0\right]$ and $\mathcal{M}(\mathrm{d}s) = \left[0, 1\right]$.

Definition 4.4 (maps in the form of matrix)**.** Suppose $\mathcal{F} : \mathbb{F}^{m\times n} \to \mathbb{F}^{m\times n}$

$$\mathcal{F}(\mathrm{X}) = \begin{bmatrix} f_{11}(x_{11}) & f_{12}(x_{12}) & \cdots & f_{1n}(x_{1n}) \\ f_{21}(x_{21}) & f_{22}(x_{22}) & \cdots & f_{2n}(x_{2n}) \\ \vdots & \vdots & & \vdots \\ f_{m1}(x_{m1}) & f_{m2}(x_{m2}) & \cdots & f_{mn}(x_{mn}) \end{bmatrix}$$

$$\mathrm{d}\mathcal{F}(X^0) := \begin{bmatrix} \mathrm{d}f_{11}(x^0_{11}) & \mathrm{d}f_{12}(x^0_{12}) & \cdots & \mathrm{d}f_{1n}(x^0_{1n}) \\ \mathrm{d}f_{21}(x^0_{21}) & \mathrm{d}f_{22}(x^0_{22}) & \cdots & \mathrm{d}f_{2n}(x^0_{2n}) \\ \vdots & \vdots & & \vdots \\ \mathrm{d}f_{m1}(x^0_{m1}) & \mathrm{d}f_{m2}(x^0_{m2}) & \cdots & \mathrm{d}f_{mn}(x^0_{mn}) \end{bmatrix}$$

$$\mathrm{d}X = \begin{bmatrix} dx_{ij} \end{bmatrix}$$

suppose $f : \mathbb{F}^{m\times n} \to \mathbb{F}$

$$\frac{\mathrm{d}f}{\mathrm{d}X} := \begin{bmatrix} \frac{\partial f}{\partial x_{ij}} \end{bmatrix}$$

Chapter 5

Linear Models

324

aaa

32

Chapter 6

Genelised Linear Models