Port II:

Calculus

ML motivation

HL is all about ophimizing functions to fit the training data, and we typically use gradients to de this. So we need to know everything about differential coloulus in 12^d.

To be able to define all of this, we first need to look at sequences and convegence.

hud if you would be be a Bayesian, you need to interprate all the

Al heyessed: couvrence et a learning alpor Mini.

Cauchy requence



A point
$$x \in \mathbb{R}$$
 is raked accumulation point of the
sequence $(x_u)_u$ if
 $\forall E > O \forall N \in \mathbb{N}$ $\exists u > N : |x_u - x| < E$



1) lu IR, ou replace the abrabuk value with a varue: Il ru-rll.

N-200

_

A sequence
$$(x_u)_u$$
 counsques to $x \in \mathbb{R}^d$ if
 $\forall E 70 \exists N \forall u > N : |x_u - x| < E$
Notation: $\lim_{x_u} x_u = x$, $x_u = x_u$

First observations

- · a sequece can have many acc. points (or none at all)
- · even if the sequence har just one acc. point, it is not necc. a Canchy sequence.
- · If (Ku)n converges to x, then x is the only acc. point

and the sequence it Courty.

Example

٠

•

Exan ples

- A is the maximum of EO, 1]. It is also the representation of EO, 1].
- . Jo, 1 [does not have a maximum element.
- · 5 is an upper bound of Jo, 1[. 1 is also an upper bound of Jo, 1[.
- 1 is the representation of JO,1E.

Bounded requince

A requence (Xn)nein CR is called bounded if there wirt a, b G R such that xy G Each J for all net.

Linsup and liminf

For a require
$$(x_u)_u \subset \mathbb{R}$$
 we define:
liminf $x_u := \lim_{n \to \infty} (\inf_{m \ge n} x_m)$
 $n \to \infty$ $m \ge n$

Observations

B-Y S WX : NCNY NE OC34

Open and elored pets

Def Let (X, d) be a metric space, and densh for
$$x \in X$$
, $z \ge 0$
 $B_z(x) = \{y \in X \mid d(x,y) \le \varepsilon\}$
Def Arubat $u \in x$ of a metric space is called cloud if

Def Arubat
$$U \subset X$$
 of a metric proce is called cloud if all
Coundry-requesces course and have their limit point in U.
A set $U \subset X$ is called open if
 $\forall u \in U \ \exists E \ ? O \ : B_E(u) \subset U.$

I ropologies, open, closed not an abo defined if a metric does not exist ...

Examples

- . Let [0,1] is cloud
- . set Jo, nE is open:



• A cut ll cau be neither open vor closed: E0,1E Open vs. claud

Proposition: Complements of eper sets are closed. Complements of closed sets are open.

Unbriss, closur Def A point u e U is an inhist point of U if Hure wists a E>O s.H. $B_E(u) \subset U$.

The (topolopical) closure of a set U is defined as
the set of points that can be appreciated by
Cauchy requences in U:

$$w \in \overline{U} \subset \forall E \ge \forall E \ge 0 \exists z \in U : d(w,z) < E$$

Boundary

$$\begin{array}{l} \text{The } \left(\begin{array}{cccc} \text{to pological} \right) & \text{boundary of } e & \text{set } \mathcal{U} & \text{if defined} \\ \text{ar } & \text{the not } \mathcal{U} & \mathcal{U}^{0} & & \text{librature ust alward} \\ & X = \begin{bmatrix} 0, 1 \end{bmatrix} & & \text{some hines one also} \\ & X = \begin{bmatrix} 0, 1 \end{bmatrix} & & \text{some hines one also} \\ & X = \begin{bmatrix} 0, 1 \end{bmatrix} & & & \text{if } u & u^{0} \\ & X & & \text{if } u & u^{0} \end{array} \\ & X^{\circ} = \begin{bmatrix} 0, 1 \end{bmatrix} & & & & & \\ & X^{\circ} = \begin{bmatrix} 0, 1 \end{bmatrix} & & & & \\ & & & & & \\ & X^{\circ} = \begin{bmatrix} 0, 1 \end{bmatrix} & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & & \\ & & & & \\ &$$

Oense sets

.



Continuous function



Alternative définitions

$$B \subset \mathcal{C} \quad \text{open} \implies f^{-1}(B) := \{x \in X \mid f(x) \in B\} \quad \text{open}$$

in $\mathcal{C} \quad \text{in } X$
in $\mathcal{C} \quad \text{in } X$



lutaition: bounded divivative

Lipschitz continuous

A function
$$f: X \to Y$$
 is called Lipschitz continuous
with Lipschitz constant L if
 $\forall x, y \in X : d(f(x), f(y)) \leq L \cdot d(x, y)$
 $utuition: bounded derivation$



Intruediak value Krearen

Meorem: If f: [a, b] -> R is continuous, then f attains all values between fra, and frb): $\forall_Y \in [f(a), f(b)] \exists x \in [a, b]; f(x) = y$.





hurse function

Invertible function



. Continuity of the invertable directly for cont. of f.

Pointwire convergence

Example





Uniform convegence

Intuition



=> Not uniformly cours. O
Alknatic definition

$$f_{n} \rightarrow f \quad \text{uniformly} \quad iff \quad \|f_{n} - f h_{\infty} \rightarrow 0,$$

$$\mathcal{C}(\mathfrak{d}) := \left\{ f: \mathfrak{d} \rightarrow |\mathfrak{R}| \quad f \quad \text{continuous} \right\} \quad \text{is vertor space}$$

$$\|\cdot\|_{\infty} \quad \text{is a norm} \quad \mathfrak{d} \rightarrow \mathcal{C}(\mathfrak{d})$$

$$\|f\|_{\infty} := \sup_{x \in \mathfrak{d}} \quad |f_{rxs}|$$

Runarh: uniform converse =) pointwin

Uniform convegence provers continuition

Proof

Couridu K, KED. Suppose some E>O is given.

Obcore Kist for every
$$u \in \mathbb{N}$$
,
(for) - f(x) | $\leq |f_{cr} - f_{u}(r)| + |f_{u}(r) - f_{u}(r)| + |f_{u}(r) - f_{r}|$

replace y -> y

Uniform convergence => In ein such that for all x, x. e D $\left|f_{u}(x) - f(x)\right| \leq \frac{2}{3}$ $|f_u(y) - f(y)| \leq \frac{\varepsilon}{2}.$ pour consider the function for. By asr. it is continuous, a those exists 5>0 such that 1x0-x125 => Ifu(x0)-fu(x)12 E/3. Together we then get that for given E >0 there exists of >0 rule that for all x $|x_0-x| \leq \delta = 5$ $|f(x_0)-f(x)| \leq \frac{2}{3} + \frac{2}{3} + \frac{2}{7} = \delta.$ So f is continuous at xo.

Derivative definition

Def
$$U \subset R$$
 an interval, $f : U \rightarrow R$. The function f
is called differentiable at a $\in U$ if
 $f'(a) := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ exists.
We often writh $f'(a) = \frac{d}{dx} Ca$

Juustration





Differnhable functions

For
$$D \subset \mathbb{R}$$
 we denote
 $\mathcal{C}^{n}(D) := \int f: D \to \mathbb{R} \mid f \underbrace{\operatorname{cout}}_{\operatorname{differentiable}} \mathcal{C}^{n}(D)$

We can reprose the process of taking derivatives:

$$f' = \frac{df}{dx} \quad i \qquad f'' = \frac{df'}{dx}$$

$$V = \frac{df}{dx} \quad i \qquad f'' = \frac{df'}{dx}$$

$$V = \frac{f(n)}{dx} \quad denote \quad He \quad u = He \quad derivative \quad (if \quad aister).$$

$$e^n(D) := \{f: D \Rightarrow R \mid f \quad n \quad himes \quad continuously \quad differentialle\}$$

Heren
Let
$$f$$
 be differentiable at α . Then there exists a constant c_{α}
such that on a small ball around α we have
 $|f(x) - f(\alpha)| \leq c_{\alpha} \cdot |x - \alpha|$
In particular, f is continuous at α .

Intermediate value then for derivatives <u>Heorem</u> $f \in C^{1}([a,b])$. Then there with $\overline{s} \in [a,b]$ such that $\frac{f(6) - f(a)}{b - a} = f'(\overline{s}).$



Exchanging his and derivative

$$\frac{\text{Theorem}}{\text{fn}: [a_{1}b_{3}] = R, \text{ fn } \in \mathbb{C}^{n}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ fn } \in \mathbb{C}^{n}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ fn } \in \mathbb{C}^{n}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ fn } \in \mathbb{C}^{n}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ fn } \in \mathbb{C}^{n}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ fn } \in \mathbb{C}^{n}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ for } e^{-R}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ for } e^{-R}[a_{1}b_{3}]. \text{ If Here limit}$$

$$f_{n}: [a_{1}b_{3}] = R, \text{ for } e^{-R}[a_{1}b_{3}]. \text$$

Construction of the Ricmann-integral

Gusider a function
$$f: [a, 5] \rightarrow \mathbb{R}$$
, arrune
fluit f is bounded
 $(\exists l, u \in \mathbb{R} \forall x \in [a, 5]: l \leq f(x) \leq u).$
Consider $k_0, k_1, ..., k_n$ with
 $a = x_0 \leq k_1 \leq k_2 \dots \leq x_n = b.$
Thex points introduce a partition
of $[a, b]$ into u intervals
 $I_k := [x_k - 1, x_k].$

Th



•

Define the lower sum

$$s\left(f, \left\{x_{0}, x_{1}, \dots, x_{n}\right\}\right) = \sum_{k=\Lambda}^{n} |I_{k}| \cdot m_{k}$$

and he upper sum

$$S(f_1 \{x_{3}, x_{1}, \dots, x_{n}\}) = \sum_{k=1}^{n} |I_k| M_k$$





Now define

$$J_{*} := sup (S(f, partition))$$
partitions

We call
$$f$$
 Riemann-inhproble if $J_{*} = J^{*}$. Then we denote
 $J_{*} = J^{*} = : \int_{a}^{b} f(f) df$.

Konopone resp. continuous fets are Riemann-integrable

Many fits are not Riemann-integrable
Example:
$$f^{(x)=} \begin{cases} n & x \in \mathbb{R} \\ 0 & o Knowin \end{cases} = \frac{1}{\mathbb{R}} \\ \mathbb{R} \end{cases}$$

For any inhered
$$I_{h} = [x_{h}, x_{h+1}]$$

 $H_{h} = +\Lambda$
 $m_{h} = 0$
 $M_{m} = 0$

I

· Hard to extend to other spaces.

Fundamental Meanin of calculus

Fundamental théorem of calculus

$$\frac{\text{Theorem I}}{\text{continuous at F \in [a,b] \rightarrow R} (\text{Riemann}) - integrable and}{\text{continuous at F \in [a,b]. Let c \in [a,b]. Then the function
SEC F(x) := $\int_{C}^{x} f(t) dt$
is differentiable at 3 and $F'(\overline{5}) = f(\overline{5}).$
If $f \in C([a,b])$, then $F \in C^{1}([a,b])$ and
 $F'(x) = f(x)$ for all $x \in [a,b].$

$$\frac{\text{Theorem I}}{\int_{C}^{x}} F: [a,b] \rightarrow R \text{ continuously differentiable}, \text{ then}$$$$

Algebraic version of the Hum (informal)
Informal, algebraic version:
The integral operator I: CEarb]
$$\longrightarrow C_{ncn}^{1}(Earb)$$

with $C_{ncn}^{1} Earb] := \{f \in C^{1}Earb] : f(c) = 0\}$
is an isomorphism (linear, bijection) and its
inverse is the differential operator.

Proof Part I

$$\frac{Proof I}{h}: \quad \text{Used by prove that } F \text{ is diff. at } S.$$

$$Gounder \quad A(h) := \frac{F(S+h) - F(S)}{h}$$

$$= \frac{A}{h} \left(\int_{C} f(ch) dt - \int_{C} f(ch) dt \right)$$

$$= \frac{A}{h} \int_{C} f(ch) dt \quad \text{Wut b prove: conveges to } f(S)$$

$$= \frac{A}{h} \int_{S} f(ch) dt \quad \text{as } h \to 0$$

Would be prime:

$$\frac{A(h) - (f(x)) \stackrel{!}{\longrightarrow} 0}{f(h) dh} = \frac{A}{h} \int_{S}^{T+h} \int_{S}^{T+h} \frac{dh}{h} \int_{S}^{T+h} \frac{dh$$

Formally: given E>O we can find
$$h>O$$
 such that
 $f(t) - f(t) < E$ by $t \in [t, t+1]$.

Meu:

$$\frac{1}{h} \int_{3}^{3+h} f(t) - f(t) dt \leq \frac{1}{h} \int_{3}^{3+h} f(t) - f(t) dt$$

$$\leq \frac{1}{h} \int_{3}^{3+h} \epsilon dt = \frac{1}{h} \cdot \epsilon \int_{3}^{3+h} dt = \frac{1}{h} \cdot \epsilon \cdot h = \epsilon.$$

$$\int_{3}^{3+h} \epsilon dt = \frac{1}{h} \cdot \epsilon \int_{3}^{3+h} dt = \frac{1}{h} \cdot \epsilon \cdot h = \epsilon.$$

D Theorem I

Proof par II

Prosf II :

By (ii) we know that
$$H'(x) = F'(x) - G'(x) = 0$$
 for all x
Uncern It is a courbant function.
We know that $H(x) = F(a) - G(a) = F(a)$, thus
(iii) $H(x) \equiv F(a)$ - $Means: "constant"$
Courisdur $x = b$.
 $F(a) \stackrel{\text{ris}}{=} H(b) \stackrel{\text{def}}{=} F(b) - G(b) \stackrel{\text{def}}{=} \frac{b}{a}$
 $= F(b) - \int_{a}^{b} F'(b) \frac{def}{a}$
 $= F(b) - \int_{a}^{b} F'(b) \frac{def}{b} \frac{def}{a}$
 $= F(b) - \int_{a}^{b} F'(b) \frac{def}{b} \frac{def}{a}$
 $= F(b) - \int_{a}^{b} F'(b) \frac{def}{b} \frac{d$



couverps in the usual purpear N-> 00.

theorem (Radius of convergence)

For every power perios
$$p(x) = \sum_{n=0}^{\infty} a_n x^n$$
 there exists a $n=0$

A first cample

Radius of coursepuce:

$$V = \lim_{\substack{n \ n \ n \ n}} \left| \frac{a_n}{a_{n+1}} \right| = \lim_{\substack{n \ n \ n \ n}} \left| \frac{n}{a_{n+1}} \right|^c = \lim_{\substack{n \ n \ n \ n}} \left(\frac{n}{a_{n+1}} \right)^c = 1$$

(independently of c)

A first acample (cont.)

$$\frac{G_{nn} \quad (z - A)}{\pi} : \sum_{n=1}^{A} x^{n} \quad low \quad couv. \ radius \ r=1$$

• For $|x| \ge A$ if diverges.

• For $|x| \ge A$ if couverpres.

• For $|x| \ge A$ no pureal violement, but we can analyze it more closely:

• For $x = +A$ the veries diverges brownen

 $\sum_{n=1}^{A} x^{n} = \sum_{n=1}^{A} A^{n} = \sum_{n=1}^{A} A^{n} \Rightarrow \infty$

• For $x = -A$ it couverges:

 $\sum_{n=1}^{A} (-A)^{n} = -A + \frac{A}{2} - \frac{A}{3} + \frac{A}{4}$

 $= -(A - \frac{A}{2} + \frac{A}{3} - \frac{A}{4} - \dots) = -\log(2)$

A first example (cont.)

$$\frac{\operatorname{Carr} C=0}{\operatorname{Carr} C=0} : \sum_{n=1}^{\infty} u^{n} x^{n} = \sum_{n=1}^{\infty} dimps \quad pr \quad |x| = r$$

$$\operatorname{Gouvregunce radius is while $r=\Lambda$.
$$\cdot \quad \text{For } |x| \leq \Lambda \quad \text{diverges.}$$

$$\cdot \quad \text{For } |x| \geq \Lambda \quad \text{diverges.}$$

$$\cdot \quad \text{For } |x| \geq \Lambda \quad \text{diverges.}$$

$$\cdot \quad \text{For } |x| = \Lambda :$$

$$\cdot \quad x = +\Lambda : \sum_{i=\Lambda}^{\infty} x^{n} = \sum_{i=\Lambda}^{\infty} \Lambda = N \rightarrow \infty \quad \text{diverges.}$$$$

•
$$\chi = -\Lambda : \sum_{x} = -\Lambda + \Lambda - \Lambda + \Lambda - \Lambda + \dots$$

does not courry

More examples

• Exponential perior:

$$exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
 has $r = \infty$
 $u = 0$ $\frac{a_n}{n!} = \frac{n/n!}{n!} = u = n$

•
$$\sum_{n=0}^{\infty}$$
 $n! \times n$ has $r=0$: $\left|\frac{a_n}{a_{n+1}}\right| = \frac{h!}{(n+n)!} = \frac{1}{n+1} \rightarrow 0$.

 \mathfrak{O}
From power perios to Taylor perios, intuition
Observation: Given power perios
$$f^{(x)=} = \sum_{n=0}^{\infty} a_n (x-a)^n$$
.
Let's take its durivative:
 $f'(x) = (a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + ...)^1$
would under = $a_1 + 2a_2(k-a) + 3a_3(x-a)^2 + ...$
 $f = \sum_{n=A}^{\infty} n \cdot a_n (x-a)^{n-A}$
 $f^{(k)}(x) = \sum_{n=k}^{\infty} a_n (n \cdot (n-A) \cdot (n-2) \cdot ... \cdot (n-k+A)) (x-a)^n$
In particular, we have for $x = a$
 $f^{(k)}(a) = a_k k^{\frac{1}{2}}$ or, shaked otherwise $a_k = \frac{f^{(k)}(a)}{k!}$

$$\frac{\text{Theorem}: \text{Let } f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n \text{ with } r \ge 0. \text{ Then for } x \text{ with}}{|x-a| \ cr \ we \ have}$$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n$$

Question Does it work the oker way round? That is,
given any function (possibly with nice assumptions),
can we simply build the price
$$\sum_{n=1}^{p(n)} (x-a)^n$$
 and
"hope" that it couverpts to the function?
 $= f(x)$???



Taylor series

$$\frac{\text{Theorem}}{f \in \mathcal{L}^{n+n}(J)} := \int_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^{k} \qquad \text{Taylor series} \\ \prod_{n} (x,a) := \int_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^{k} \qquad \text{Taylor series} \\ \prod_{n \in O} (x,a) := \int_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (t) dt \qquad \text{Remainder ferm} \\ \prod_{n \in A} (x,a) := \int_{n}^{\infty} \frac{(x-t)^{n}}{n!} f^{(n+1)}(t) dt \qquad \text{Remainder ferm} \\ \text{Then } f^{(x)} = \prod_{n \in A, a} t \quad \mathbb{R}_{n} (x, a)$$



prové



· Consider
$$F(t) = \frac{(x-t)^{n+n}}{(n+n)!} f^{(n+n)}(t)$$

Taylor with Lagrange remainder

Theorem :

$$f \in \mathcal{C}^{n-en}(J), \quad a, x \in J. \text{ Then there exists some}$$

$$\overline{s} \in J \quad such \quad \text{Heat}$$

$$R_n(x, a) = \frac{(x-a)^{n+n}}{(n+n)!} \quad f^{(n+n)}(\overline{s})$$

Proof

Proof Let
$$J = [a_1 b]$$
.
• Counder two functions $F_1 \ G \in \mathcal{C}^{n+d}([a_1b])$. Assume
Heat $F(a) = G(a = 0, and G' \neq 0 or [a_1b].$ (**)
Now:
 $\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F(b)}{G(b)} = \frac{F^1(S)}{G'(S)}$ for some $S \in [a_1b]$
Assume that F' and G' also satisfy (*). We can itrade ...
We would obtain
 $\frac{F(b)}{G(b)} = \frac{F(a)}{G(a-a)} (S)$ for some $S \in [a_1b]$

Proof (cout.)

• Now chose
$$F(x) = f(x) - T_n(x_i a) = R_n(x_i a)$$

 $G(x) = (x-a)^{n+A}$

• For all
$$k$$
 in $OGK \leq u$ are have by construction that
 $f^{(k)}(a) = \overline{I}_{n}^{(k)}(a)_{j}$ so in particular
 $f^{(k)}(a) = O_{j}$ and we have $G^{(k)}(a) = O_{j}$.

Proof (cout.)

• For n=A we now have

$$F^{(n+A)}(x) = f^{(n+A)}(x), \quad G^{(n+A)}(x) = (n+A)!$$

$$g_{\gamma}(\#) \text{ we obtain}$$

$$F(x) = R_n(x,a) \stackrel{\#}{=} G(x) \cdot \frac{F^{(n+A)}(s)}{G^{(n+A)}(s)} =$$

$$= \frac{(x-\alpha)}{(x+\alpha)!} f^{(\alpha+\alpha)} (\overline{S}).$$

Taylor converpuse

Theorem
$$f \in \mathcal{C}^{\infty}(\mathfrak{I})$$
, $x_{i} \in \mathfrak{I}$. Define
 $T(\mathfrak{x}) := \lim_{n \to \infty} T_{n}(\mathfrak{x}) = \sum_{n=0}^{\infty} \frac{f^{(n)}(\mathfrak{a})}{\mathfrak{n}!} (\mathfrak{x}-\mathfrak{a})^{n}$.
Then we have $f(\mathfrak{x}) = \overline{T}(\mathfrak{a})$ if $K_{n}(\mathfrak{x},\mathfrak{a}) \xrightarrow{n+\infty} 0$.
For example, this is the case if there exist constants
 $\mathfrak{a}_{i}(\mathfrak{c} > 0)$ such that
 $|f^{(n)}(\mathfrak{c})| \leq \mathfrak{a} \cdot \mathfrak{C}^{n}$ $\forall \mathfrak{f} \in \mathfrak{I}$, $\mathfrak{b} \mathfrak{u} \in \mathfrak{N}$.

Follows directly from the Laprangian remainder.

Examples

• Exponential review:

$$exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
 pour series with $r=00$,
 $exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ exp always coincides with the Taylor series.

•
$$f(x) = \log(\Lambda + x)$$
, Taylor series about $\alpha = 0$
Can prove: Convergence radius of Taylor series is $r=1$
For x outride of $J-\Lambda$, ΛE Taylor series does not
wake server at all.

Examples (cont.)

•
$$f(x) = \begin{cases} exp(-A/x^2) & if x \neq 0 \\ 0 & k = 0 \end{cases}$$

Has the funny property that $\forall u \in h$: $f^{(u)}(0) = 0$
Consider the Taylor series derived about $a = 0$.
All terms will be 0, so $\forall u$: $T_u(x) = 0$, $r \neq \infty$
but of cause f is $u \neq = 0$, so we get
 $\forall x \neq 0$, $T_u(x) \neq f(x)$.
Faylor perior converges everywhere, but ust to the fet f !



Goal

$$\begin{bmatrix} R \\ C_1 \end{bmatrix} \begin{bmatrix} c_2 & -\infty & vol(R) \\ \vdots & c_1 \cdot c_2 \end{bmatrix}$$

And want "vice" mathematical properties.

Earlier opproaches

Now: generalization of Kir yproach

$$\mu(G) = \mu(G) + \mu(A)$$

$$\mu(A) = \mu(G) - \mu(G) + \mu(G)$$

· Need 6-algebra as underlying structure.

$$\begin{aligned} & \text{Onto Lebergue measure} \\ & \text{Fit the "natural volume" of rectangles:} \\ & \text{R} = [a_n, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n] \quad C \text{R}^n \\ & \text{IRI} := \prod_{i=1}^n (b_i - a_i) \\ & \text{IRI} := \prod_{i=1}^n (b_i - a_i) \end{aligned}$$

Definition of outer Lebergue measure:
Let A C R^{LL} be additionary. We define
$$\lambda(A) := \inf \left\{ \sum_{i=1}^{\infty} |R_i| \mid A \subset \bigcup_{i=1}^{\infty} R_i \mid R_i \text{ rechangle } \right\}$$

We cours A by a countroble union of rectangles, Hen hale inf.
Observe : $\lambda(A) \in \mathbb{R} \cup \{\infty\}$.

Measurable set

Outer measure as measure ...

Examples :

$$\lambda (\{x\}) = 0$$
$$\lambda (R) = \infty$$

• A C R countrable. The
$$\lambda(x) = 0$$
. In particular,
R is measurable and has $\lambda(Q) = 0$.

Proof shetch

For E > 0, define for all ai eA the inhural [xi, yi [such that

$$x_{i} = \alpha_{i} - \frac{\varepsilon}{2^{i+1}} \quad j \quad y_{i} = \alpha_{i} + \frac{\varepsilon}{2^{i+1}}$$



$$A \subset \bigcup_{i=1}^{\infty} E_{i,ri}E_{i=1}$$

 $= \sum_{i=1}^{\infty} \lambda(A) \leq \sum_{i=1}^{\infty} \lambda(\sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{\infty} \sum_{i=1}^{\infty} \sum_{i=1}^{\infty} \sum_{i=1}^{n} \sum_{i=$

Taking the inf. our all country shows that $\lambda(A) = 0$.

Summer: ZZB (up to sets of measur O).



Construction (quitabhact!)

Consider [0, 1[. Define au equivalence relation on [s, 1[as follows: x ~ y : (=> x - y 6 Q $\frac{\pi}{4}, \frac{\pi}{4} + \frac{1}{2} + \frac{\pi}{4}, \frac{\pi}{4} + \frac{759}{800}$ would be equivalent Couride the equivalence classes $\prod_{i} \neq Q = \left\{ \prod_{i} \neq q \mid q \in Q \right\}$ St a V2 + Q We pick a representation of each of the classes, and denote by N the set of all such representations.



Proof

Proof by contradiction:
Assume N is measurable. We now construct the
following entr: For
$$q \in [3, 1[$$

 $Nq := ((q + N) \cup (q - 1 + N)) \cap [3, 1[$



• if N is meanwords, here
$$q \in N$$
 is measurable to $G \subseteq O(A)$
and $\lambda(Nq) = \lambda(N)$

•
$$[o_1 \cap C = \bigcup N_q \\ q \in [S_1 \cap R]$$

•
$$N_q \cap N_p \neq \emptyset = \sum N_p = N_q$$

Consequently, $\bigcup N_q$ is disjoint.
• G -additivity:
 $\lambda (E_0, 1E) = \lambda (\bigcup N_q) = \sum_{q \in Eqn J \cap R} \lambda (N_q)$
 1

Proof (cont.)

h

Riemaun:





Def A function
$$f:(X, F) \rightarrow (Y, G)$$
 between two measurable
spaces is called measurable if pre-images of measurable who
are measurable:
 $\forall G \in G : f^{-1}(G) \in F$
 $L_{=}: \{x \in X \mid f(x) \in G\}$

Lebesgue-integral for single fits $2if \phi : \mathbb{R}^{n} \to \mathbb{R}$ is called a <u>single function</u> if Hure exist measurable sets $S_{i} \subset \mathbb{R}^{n}_{i}$ a is \mathbb{R} such that $\phi = \sum_{i=1}^{n} a_{i} M_{SSi}^{2}$ $S_{i} = \phi^{n}(a_{i})$

For such a simple function we can define its belongue integral at $\int \Phi \, dA := \sum_{i=1}^{n} a_i A(S_i)$



Lebesgue in hyprol for non-negative fet
For a non-negative function
$$f^{\dagger}: \mathbb{R}^{n} \rightarrow \mathbb{E}_{0}, \infty\mathbb{E}$$
 in define its
Lebesgue integral
 $\int f^{\dagger} d\lambda = \sup \left\{ \int \phi d\lambda \mid \phi \leq f, \phi simple \right\}$
(might be ∞)
 $\int \phi d\lambda = \int \phi d\lambda \mid \phi \leq f, \phi simple fets$
Note: the ats Fi can be
complicated atty, not just intervals.

• For a general function
$$f: \mathbb{R}^{n} \to \mathbb{R}$$
 we split the function
into positive and neg. part: $f = f^{+} - f^{-}$
where $f^{+}(x) = \int f(x) = \int$


Mearen (monohoue convergence):
Consider a requesse of functions
$$f_{u}: \mathbb{R}^{n} \rightarrow \mathbb{E}_{0}, \infty\mathbb{E}$$

that is pointwise non-decreasing:
 $\forall x \in \mathbb{R}^{n}: f_{u \in u}(x) \ge f_{u}(x)$.
Assume knot all fu are measurable,
and knot the pointwise limit exists:
 $\forall x: lim f_{u}(x) = : f(x)$

[hen:

$$\int f(x) dx = \lim_{\substack{k \to \infty}} \int f_{k}(x) dx$$

$$\int \lim_{\substack{k \to \infty}} f_{k}(x) dx$$



Functions on R^N
We now cousider punctions
$$f: R^n \rightarrow R$$
.
input space: n-dim output space
 $1-dim$

Generalizing drivations from 1 kon dimensions Two obvious "ideas": (1) Consider the function $f(x) = f\left(\begin{array}{c} x_{n} \\ x_{n} \end{array}\right)$ as hurchion of one coordinate only and here the oker fixed; Hen ypy 1-dim intrition ~ partial duivations

(2) Approximate & brally by something linear shotal derivative As we will see, leads to similar but not identical notions.

Partial disinatives on Rh

$$g: \mathbb{R} \to \mathbb{R}$$
,
 $g(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_{j-1} | x_j | \overline{s}_1, ..., \overline{s}_n)$

 $(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_{j-1} | x_j | \overline{s}_1, ..., \overline{s}_n)$

 $(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_{j-1} | x_j | \overline{s}_1, ..., \overline{s}_n)$

 $(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_{j-1} | x_j | \overline{s}_1, ..., \overline{s}_n)$

 $(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_{j-1} | x_j | \overline{s}_1, ..., \overline{s}_n)$

 $(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_{j-1} | x_j | \overline{s}_1, ..., \overline{s}_n)$

 $(x_j) := f(\overline{s}_1, \overline{s}_2, ..., \overline{s}_j)$



Gradient

If all partial derivations exist, then the vector of all
partial derivations is called the gradient:
$$grad(f)(s) = \nabla f(s) = \begin{pmatrix} \frac{2}{2}f(s) \\ \frac{2}{2}x_1(s) \\ \vdots \\ \frac{2}{2}f(s) \end{pmatrix} \in \mathbb{R}^n$$

Jacobian matrix

If
$$f: \mathbb{R}^{n} \to \mathbb{R}^{m}$$
, we decompose f into itr m
component functions $f=\begin{pmatrix} f_{n} \\ \vdots \\ f_{m} \end{pmatrix}$. We define the
Jacobian matrix $(\text{grad } f_{n})^{t}$
 $\mathbb{D} f(x) = \begin{pmatrix} \frac{\partial f_{n}}{\partial x_{n}} & \cdots & \frac{\partial f_{n}}{\partial x_{n}} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_{m}}{\partial x_{n}} & \cdots & \frac{\partial f_{m}}{\partial x_{n}} \end{pmatrix} \in \mathbb{R}^{m \times n}$

$$F: \mathbb{R}^{2} \rightarrow \mathbb{R}, \quad f(x_{i}y) = \begin{cases} \frac{x \cdot y}{x^{2} + y^{2}} & \text{if } (x_{i}y) \neq (o_{i}o) \\ 0 & \text{if } x = y = 0 \end{cases}$$

One can prove that the two partial dividuations at (2,0) dist and are both d, but the function of it was continuous at O.

Exercise!

Total drivative

Differnhalle fet

$$f(\bar{s}+h) - f(\bar{s}) = L(h) + r(h)$$

with
$$\lim_{h \to 0} \frac{r(h)}{|h|} \to 0$$





Theorem
$$f: \mathbb{R}^{n} \to \mathbb{R}$$
 differentiable at \overline{s} .
(1) Then f is continuous at \overline{s} .
(2) The linear functional L coincides with the gradient :
 $f(\overline{s}, h) - f(\overline{s}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_{i}}(\overline{s}) \cdot h_{j} = \pi r(h)$
 $i = \sqrt{\operatorname{grad}} f(\overline{s}, h) \to \pi r(h)$
Special con: $h = c_{i} = \binom{0}{i}$. Then $\sum_{j=1}^{n} \frac{\partial f}{\partial x_{j}}(\overline{s}) \cdot h_{j} = \frac{\partial f}{\partial x_{j}} = \frac{j-th}{\rhoartial}$

400 :

Directional drivatives

Directional derivatives

$$\begin{aligned}
 \mathcal{D}_{v} f(\overline{s}) &:= \lim_{h \to 0} \frac{f(\overline{s} + h \cdot v) - f(\overline{s})}{h}
 \end{aligned}$$

Observe: partial derivatives are directional derivatives in the direction of the unit vectors.

Theorem:
$$f: \mathbb{R}^{n} \to \mathbb{R}$$
 differentiable in $\overline{3}$. Then all the
directional derivatives exist, and we can compute them by
 $\mathcal{D}_{v} f(\overline{3}) = (\operatorname{grad} f)^{\frac{1}{v}} \cdot v = \sum_{i=1}^{n} v_{i} \cdot \frac{\partial f}{\partial x_{i}} (\overline{3})$
 $(\sum_{i=1}^{n} v_{i} \cdot \frac{\partial f}{\partial x_{i}} (\overline{3}))$

$$\frac{\text{Largert ascell along gradhent}}{Proporition} : f: R^{n} \rightarrow 1R differentially of F. Here the largest value $[D_{V}(f)(f)]$
aurry the dir. derivatives v is allowind in direction of the gradhent,
 $v = \frac{\text{grad} f(f)}{\text{H} \text{grad} f(f) \text{H}}$.

$$\frac{P_{nosf} \text{ intuition}:}{\text{H} \text{grad} f(f) \text{H}}$$

$$\frac{P_{nosf} \text{ intuition}:}{(D_{V} f(f)) \text{H}} = \max_{V_{j} | N | l = 4} | \langle \nabla f(f), v \rangle | \leq (\text{and} g - f \text{dere})$$

$$\max_{V_{j} | N | l = 4} | \nabla f(f) \text{H} \cdot \text{H} \vee \text{H} = \text{H} \nabla f(f) \text{H}.$$

$$\text{In particular}_{j} \text{ for } v = \frac{\nabla f(f)}{\| \nabla f(f) \|} \text{ we have}$$

$$D_{V} (f(f)) = \langle \nabla f(f), \frac{\nabla f(f)}{\| \nabla f(f) \|} = \text{H} \nabla f(f) \text{H}$$$$

so the working is attained for this desice of v and the inequality browned an equality.

Illustration of the gradient



Gradient Vectors Shown at Several Points on the Surface of cos(x) sin(y)

Higher order derivatives

Courier f: Rh -> R arrune it is differentiable,
so all portial derivatives
$$\frac{\partial f}{\partial x_i}$$
: Rh -> R exist. If the partial
derivations are differentiable themselves we can take their derivatives:

$$\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) =: \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Attantion order matters!

In quival, we cannot change the order of derivatives: ∕!∖ $\frac{\partial f^2}{\partial x_i \partial x_j} + \frac{\partial f^2}{\partial x_j \partial x_i}$

$$\begin{array}{l}
 Example \\
 f: \Pi^{2} \supset \Pi_{1} \quad f^{(x_{1}y_{1})} = \frac{x \cdot y^{3}}{x^{2} + y^{2}} \\
 grad f(x_{1}y_{1}) = \left(\frac{y^{3}(y^{2} - x^{2})}{(x^{2} - y^{2})^{2}} , \frac{x y^{2}(3x^{2} + y^{2})}{(x^{2} + y^{2})^{2}} \right)
 \end{array}$$

Have:
$$\frac{\partial e}{\partial x} (0, \gamma) = \gamma$$
 for all γ
 $\frac{\partial}{\partial \gamma} (\frac{\partial f}{\partial x}) = 1$
 $\frac{\partial e}{\partial \gamma} (x, 0) = 0$ \forall all x
 $\frac{\partial}{\partial \gamma} (\frac{\partial f}{\partial \gamma}) = 0$ \forall all x
 $\frac{\partial}{\partial x} (\frac{\partial f}{\partial \gamma}) = 0$ \forall all x

Consequently, the two derivatives do not agree on point (0,0).

Hessian

The f Herrian matrix

$$f: \mathbb{R}^{n} \rightarrow \mathbb{R}$$
, thus we define the Herrian of f at point x by
 $(Hf)_{ij}(x) := \frac{\partial^{2} f}{\partial x_{i} \partial x_{j}}(x)$
 $i_{ij} = \Lambda_{j} \dots N_{ij}$

f: Rⁿ -> Rⁿ function
Vf: Rⁿ -> Rⁿ first derivative : a partial deriv.
Hf: Rⁿ -> R^{n+fi} second derivative :
² "partial derivatives"
$$\frac{2f}{2f}$$

Continuously differentiable fets
Def We say that
$$f: \mathbb{R}^n \to \mathbb{R}$$
 is continuously differentiable,
if all pertial divivative $\frac{\partial f}{\partial x_i}$ exist and are continuous.
We say that f is twice continuously differentiable if
 f is continuously differentiable and all pertial
divivations $\frac{\partial f}{\partial x_i}$ are again continuously differentiable.
Analogously: le times cout. differentiable
Notation: $\mathcal{L}^k(\mathbb{R}^n, \mathbb{R}^m) = \{f:\mathbb{R}^n \to \mathbb{R}^m \mid k times cout.
 $diff. \tilde{g}$
 $\mathcal{L}^{\infty}(\mathbb{R}^n, \mathbb{R}^m) = \{f:\mathbb{R}^n \to \mathbb{R}^m \mid \infty \text{ other cont.}$
 $diff. \tilde{g}$$

Theorem (Schwartz) Assume that
$$f$$
 is twice continuously
differentiable. Then we can exchange the order in which are take
partial obsidentias:
 $\frac{\partial^2 f}{\partial x_i} = \frac{\partial^2 f}{\partial x_j}$
Analogously: le times cout. diff. \Rightarrow can exchange order of first
le partial derivatives.

Kinimum

$$f:\mathbb{R}^n \to \mathbb{R}$$
. f has a local minimum at x_0 if there exists $\mathcal{E} \ge 0$ such that
 $\forall x \in \mathcal{B}_{\mathcal{E}}(x_0) : f(x) \ge f(x_0)$

.

Saddle point



How can we find out which case we have?
Intuition in A-dim rase: second derivatives might help:
Strict local min: strict local way soddle
$$pt$$
:
 $f'(x) = 0$ $f'(x) = 0$ $f'(x) = 0$
 $f''(x) = 0$ $f''(x) = 0$ $f''(x) = 0$
 $f''(x) = 0$ $f''(x) = 0$ $f''(x) = 0$

Critical points and the Herrian
Hisorun f: Rⁿ -> R, f & C²(Rⁿ). Assume that to is
a critical point, i.e.
$$\nabla f(x_0) = 0$$
. Then:
(i) If to is a local minimum (maximum), thus the Hessian
H f (xo) is positive semi-definite (neg. semi-def.)
(ii) If Hf(xo) is positive definite (neg. definite), thus to
is a strict local min (max).
If Hf(xo) is indefinite, thus to is a modelle point.
Negative and positive eigenvilues
) (11), we can have a shirt beal max., yet the herrion is
only semi-definith. Example:
$$f(x) = x^4 at x = 0$$
.
(11) (2), if the Herrion is semi-definite, us statement
can be made.

$$f: \mathbb{R}^{d} \rightarrow \mathbb{R}$$
 that minimizes the least squares error.
In matrix nobalism: $X = \begin{pmatrix} -x_{n} - \\ -x_{2} - \end{pmatrix} \in \mathbb{R}^{n \times d}, \quad Y = \begin{pmatrix} x_{n} \\ \vdots \\ y_{n} \end{pmatrix}$

To ophimize for
$$w$$
, need to have derivative of the objective for b or:
 $\frac{\partial g}{\partial y} \stackrel{!}{=} 0$.
 $\frac{\partial w}{\partial w}$

To compute the gradient by pot is pretty cumbersome:
• Write fet coordinate - with:
$$g\begin{pmatrix}\omega_1\\i\\\omega_n\end{pmatrix} = \sum_{j=1}^{n} (\gamma_j - \sum_{k=1}^{n} \chi_{jk} \omega_k)^2$$

$$\frac{\partial g}{\partial \omega_i} = \sum_{j=1}^{n} (-\kappa_{ji}) \cdot 2 \cdot \left(\kappa_{j} - \sum_{k=1}^{m} \kappa_{jk} \omega_k \right)$$

Observe that we can write usult using matrices

$$\frac{\partial \mathbf{y}}{\partial \omega_{i}} = \sum_{j=n}^{m} \left(-\kappa_{ji}\right) \cdot 2 \left(\gamma_{j} - \sum_{k=n}^{m} \kappa_{jk} \omega_{k}\right)$$

$$-2 \cdot \sum_{j=n}^{n} \kappa_{ji} \cdot \left(\gamma_{j} - \chi_{\omega}\right)_{j}$$

$$\left(\chi^{t} \left(\gamma_{j} - \chi_{\omega}\right)\right) i$$

$$\nabla g (\omega) = -2 \quad \chi^{t} \left(\gamma_{j} - \chi_{\omega}\right)$$

Observe: "Syntax" close to A-dim care:

$$g(\omega) = (\gamma - \kappa \cdot \omega)^2$$

 $g'(\omega) = -\chi(\gamma - \chi \cdot \omega) \cdot 2 = -2\chi(\gamma - \kappa \cdot \omega)$

The matrix cool book

Examples for functions of vectors:
$$f: \mathbb{R}^{n} \to \mathbb{R}$$

• $f(x) = a^{t}x$ ($a \in \mathbb{R}^{n}$) linear for
 $= \langle a, x \rangle$
 $\frac{\partial f}{\partial x} = a \in \mathbb{R}^{n}$

$$f(x) = x^{t}A \times quadrahic fct$$

$$=) \frac{\partial f}{\partial x} = (A + A^{t}) \times \in \mathbb{R}^{n}$$

$$f(x) = a^{t} X b \mu r a \in \mathbb{R}^{n}, b \in \mathbb{R}^{m}$$

$$\frac{\partial f}{\partial X} = a \cdot b^{t} \in \mathbb{R}^{n \times m}$$

$$\frac{\partial f}{\partial X} = 1$$

•
$$f(x) = a^{t} X^{t} C X b$$
 for $a \in \mathbb{R}^{m}$, $b \in \mathbb{R}^{m}$, $C \in \mathbb{R}^{nm}$
 $X \in u \neq n$ $u \neq n$ $u \neq n$ $u \neq n$
 $\frac{\partial f}{\partial X} = C X a b + C X b a^{t}$

Examples for functions of matrices:
$$f: \mathbb{R}^{n \times n} \to \mathbb{R}$$

• $f(x) = tr(X) \Rightarrow \frac{\partial f}{\partial x} = I \in \mathbb{R}^{n \times n}$ brace
• $f(x) = tr(A \cdot x) \Rightarrow \frac{\partial f}{\partial x} = A$
 $f(x) = tr(X^{\dagger}A \cdot x) \Rightarrow \frac{\partial f}{\partial x} = (A + A^{9}x)$

•
$$f(x) = def(x)$$
 Determinant
 $\frac{\partial f}{\partial x} = def(x) (x^{b})^{-1}$
 $\frac{\partial def}{\partial x} = def(A) \cdot (A^{-1})$
 $\frac{\partial}{\partial a_{sr}}$

•
$$f(A) > A^{-1}$$
, $f_{ij} := (A^{-1})_{ij}$ luverse

$$\frac{\partial f_{ij}}{\partial a_{ur}} = -(a_{iu})^{-1}(a_{rj})^{-1}$$