Quick Math Cheat Sheet

12Dec21

February 5, 2023

Differentiation

- 1D derivative: $\frac{1}{h}\lim_{h\to 0}(f(x+h)-f(x)-mh)$ (i.e. $f'(x)=m=\lim_{h\to 0}(f(x+h)-f(x))$)
- ND derivative: $\frac{1}{|h|}\lim_{h\to 0}(f(x+h)-f(x)-[Df(x)]h)$
- $x=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\in\mathbb{R}^n, f(x)=\begin{bmatrix}f_1(x)\\ \vdots\\ f_m(x)\end{bmatrix}\in\mathbb{R}^m,$

  $[Df(x)]=\nabla_x\begin{bmatrix}f(x_1) & \dots & f(x_n)\end{bmatrix}=\begin{bmatrix}\frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_n}\\ \vdots & & \\ \frac{\partial f_m(x)}{\partial x_1} & \dots & \frac{\partial f_m(x)}{\partial x_n}\end{bmatrix}, [Df(x)]\in\mathbb{R}^{m\times n}$
- $\frac{\partial f}{\partial x_i}=\frac{1}{h}\lim_{h\to 0}f(\begin{bmatrix}x_1\\ \vdots\\ x_i+h\\ \vdots\\ x_n\end{bmatrix})-f(\begin{bmatrix}x_1\\ \vdots\\ x_i\\ \vdots\\ x_n\end{bmatrix})$, more precisely $\frac{\partial f_j}{\partial x_i}=\frac{1}{h}\lim_{h\to 0}f_j(\begin{bmatrix}x_1\\ \vdots\\ x_i+h\\ \vdots\\ x_n\end{bmatrix})-f_j(\begin{bmatrix}x_1\\ \vdots\\ x_i\\ \vdots\\ x_n\end{bmatrix})$)

Vector shit

- Norm: $\|\mathbf{x}\|:V\to\mathbb{R},\mathbf{x}\mapsto\|\mathbf{x}\|$
- Norm induced by inner product: $\|\mathbf{x}\|:=(\langle\mathbf{x},\mathbf{x}\rangle)^{\frac{1}{2}}$
- A hyperplane in $\mathbb{R}^n$ is a set of $\mathbf{x}\in\mathbb{R}^n$: $H=\{x:\mathbf{a}^T\mathbf{x}=c\}$ for $\mathbf{x},\mathbf{a}\in\mathbb{R}^n$, $c\in\mathbb{R}$. If $c=0$, H is the set of vectors orthogonal to $\mathbf{a}$; if $c\neq 0$, H is the set of vectors orthogonal to $\mathbf{a}$ with translational offset.

Convolution/other operations?

- 1D convolution (discrete); g kernel, f target fct: $h[n]=f[n]*g[n]=\sum_{x=-\infty}^{\infty}f(x)g(n-x)$

- 1D conv (continuous): $h[n] = f[n] * g[n] = \int_{-\infty}^{\infty} f(x)g(n-x)dx$
- Intuition: easier to visualize on discrete conv: iteratively convolving/sliding relfected, translated fct $g(x_k - x)$ over desired domain of f, computing inner product (elementwise product) of f, $g(x_k - x)$ at given location x_k, for all $x \in X$.
- $g(x) \to g(-x)$ is a reflection about y-axis ($g(x) \to -g(x)$ is a reflection about x-axis); $g(x) \to g(x + x_k)$ is a translation; $g(x_k - x)$ is a translation $x_k - x$ that b/c of its negative sign, functions as a vertical reflection. (*This latter is the usual conv notation: $x_k \mapsto x_k - x$ for a free variable x.*
- *It would be interesting to think about symmetry operations on multivariate functions/vector fields. Also, thinking about this a bit deeper, this translational symmetry $x \mapsto x$, $x \mapsto -x$ is just the "nature" of addition operation. Would like cooler way of thinking of this. Lastly, the only remaining thing in 1D conv is gaining a better feel for this operation $\langle (g(X; x_k), f(X) \rangle$*
- Illustration:

  Let
  $$g(x) = \begin{cases} 0 & x < 0 \\ 1 - x & 0 \le x \le 1 \\ 0 & x > 1 \end{cases}$$
  Construct the discrete sequence of $g(x_k - x)$ for a given $x_k \in X$. eg. let $X' \subseteq X$, eg. $X' = [-10, -5]$. For g[-8] we have $(g(-8 - (-10)), g(-8 - (-9)), \ldots, g(-8 - (-5))) = (g(2), g(1), \ldots, g(-3))$ at these chosen sampling points.

  Thus the operation $h[x_k] = f[x_k] * g[x_k] = \langle (g(X; x_k), f(X) \rangle$

Probability

- Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$; example, rolling dice: let $A =$ odd rolls $\{1, 3, 5\}$, $B =$ even rolls $\{2, 4, 6\}$. $P(X = 2|B) = 1/3$. This is immediately visible in $\{2, 4, 6\}$; otherwise, $\frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$
- Joint probability $P(A, B) := P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- Marginalization from joint distribution: $P_B(b) = \sum_a P_{A,B}(a, b) = \int_A f_{A,B}(a, b)da$
- Law of total probability: relating marginal probabilities to conditional probabilities: $P(A) = \sum_n P(A, B_n) = \sum_n P(A|B_n)P(B_n)$ if $\{B_n : n = 1, 2, 3, \ldots\}$ is a finite/countably infinte partition of the sample space.
- *Another statement of marginalization, if like better:* Law of total probability/marginalization on continuous probability spaces: For continuous

sample space: consider a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X with distribution function F_X, and an event A on $(\Omega, \mathcal{F}, P)$.

  - $P(A) = \displaystyle\int_{-\infty}^{\infty} P(A|X = x) dF_X(x)$
  - If X admits a density function f_X (why wouldn't it?): $P(A) = \displaystyle\int_{-\infty}^{\infty} P(A|X = x) f_X(x) dx$

- eg. Prove "law of total expectation": If X with $\mathbb{E}[X]$ defined, and Y is an RV on the same probability space, then: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$
  - (Insert proof)
- Products of independent variables:
  - $f_{X_1,\ldots,X_N}(x_1, \ldots, X_N) = \prod_i f_{X_i}(x_i)$
  - $\mathbb{E}[\prod_i X_i] = \prod_i \mathbb{E}[X_i]$