Machine Learning

Daniel Drake

May 12, 2020

1 Gradients, Jacobian, Ferchet Drivative, and Sub-Gradients

Theorem 1.1. ***Gradient***
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Differentiable function.
Then $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ where:

$$\nabla f(x) := \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \forall x \in \mathbb{R}^n$$

is called the Gradient of f.

Theorem 1.2. ***Jacobian***
Let $f : \mathbb{R}^n \to \mathbb{R}^m$ where:

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \forall x \in \mathbb{R}^n \text{ and } (\forall j \in \mathbb{N}_m)(f_j : \mathbb{R}^n \to \mathbb{R})$$

Then $J_f : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ where:

$$J_f(x) := \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x) & \cdots & \frac{\partial}{\partial x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x) & \cdots & \frac{\partial}{\partial x_n} f_m(x) \end{bmatrix}$$

Theorem 1.3. ***When the Jacobian is the Gradient***
Let $f : \mathbb{R}^n \to \mathbb{R}$
Then $(\nabla f(x) = (J_f(x))^T)(\forall x \in \mathbb{R}^n)$

Theorem 1.4. ***Chain Rule*** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ be differentiable functions.*
Then: $J_{g \circ f}(x) = J_g(f(x)) J_f(x)$

2 Finite Composition Operator

Definition 2.1. ***Finite Composition Operator***
Let the collection $X = \{X_j\}_{j=0}^n$ be a finite sequence of sets.

Further let $\{T_j\}_{j=0}^{n-1}$ *be a finite sequence of operators such that* $(\forall j \in \mathbb{N}_{n-1})(T_j : X_j \to X_{j+1})$
Then $T : X_0 \to X_n$ *defined by:*

$$T := \bigcirc_{j=0}^{n-1} T_j := ...$$

is called the ***Finite Composition Operator defined on X****.*

Theorem 2.1. ***Finite Composition Jacobian***
Let T *be defined as above.*
Then:

$$J_T(x) = ...$$

And so to calculate J_T *we need to calculate each* J_{T_j} *for each* j *only once.*
Proof on the next page:

Proof:
Case: $n = 1$

$$X = \{X_0, X_1\}$$

$$T_0 : X_0 \to X_1$$

$$T^1(x) = T_0(x)$$

$$J_{T^1}(x) = J_{T_0}(x)$$

Case: $n = 2$

$$X = \{X_0, X_1, X_2\}$$

$$T_0 : X_0 \to X_1$$

$$T_1 : X_1 \to X_2$$

$$T^2(x) = (T_1 \circ T_0)(x)$$

$$J_{T^2}(x) = J_{T_1 \circ T_0}(x) = J_{T_1}(T_0(x))J_{T_0}(x) = (J_{T_1} \circ T_0)(x) * J_{T_0}(x)$$

Case: $n = 3$

$$X = \{X_0, X_1, X_2, X_3\}$$

$$T_0 : X_0 \to X_1$$

$$T_1 : X_1 \to X_2$$

$$T_2 : X_2 \to X_3$$

$$T^3(x) = (T_2 \circ T_1 \circ T_0)(x) = (T_2 \circ T^2)(x)$$

$$J_{T^3}(x) = J_{T_2 \circ T^2}(x) = J_{T_2}(T^2(x))J_{T^2}(x) = (J_{T_2} \circ T_1 \circ T_0)(x) * J_{T^2}(x) = (J_{T_2} \circ T_1 \circ T_0)(x) * ((J_{T_1} \circ T_0)(x) * J_{T_0}(x))$$

Case: $n = k$

$$X = \{X_0, ..., X_k\}$$

$$T_0 : X_0 \to X_1$$

$$\vdots$$

$$T_{k-1} : X_{k-1} \to X_k$$

$$T^k(x) = (\bigcirc_{j=0}^{k-1} T_j)(x) = (T_{k-1} \circ T^{k-1})(x)$$

$$J_{T^k}(x) = J_{T_{k-1} \circ T^{k-1}}(x) = J_{T_{k-1}}(T^{k-1}(x))J_{T^{k-1}}(x)$$

Theorem 2.2. ***Examples:***
Let $T : \mathbb{R}^n \to \mathbb{R}^n$ such that:

$$T(x) := \begin{bmatrix} \arctan(\pi_1(x)) \\ \vdots \\ \arctan(\pi_n(x)) \end{bmatrix} = \begin{bmatrix} \arctan(x_1) \\ \vdots \\ \arctan(x_n) \end{bmatrix}$$

Then

$$J_T(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} T_1(x) & \cdots & \frac{\partial}{\partial x_n} T_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} T_n(x) & \cdots & \frac{\partial}{\partial x_n} T_n(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} \arctan(x_1) & \cdots & \frac{\partial}{\partial x_n} \arctan(x_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} \arctan(x_n) & \cdots & \frac{\partial}{\partial x_n} \arctan(x_n) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \arctan(x_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial}{\partial x_n} \arctan(x_n) \end{bmatrix} = \begin{bmatrix} \frac{1}{x_1^2+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{x_n^2+1} \end{bmatrix} = I \begin{bmatrix} \frac{1}{x_1^2+1} \\ \vdots \\ \frac{1}{x_n^2+1} \end{bmatrix}$$

Theorem 2.3. ***Examples:***
Let $T_b : \mathbb{R}^n \to \mathbb{R}^n$ such that:

$$T_b(x) := \begin{bmatrix} \pi_1(x) + b_1 \\ \vdots \\ \pi_n(x) + b_n \end{bmatrix} = \begin{bmatrix} x_1 + b_1 \\ \vdots \\ x_n + b_n \end{bmatrix}$$

Then

$$J_{T_b}(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} [x_1 + b_1] & \cdots & \frac{\partial}{\partial x_n} [x_1 + b_1] \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} [x_1 + b_1] & \cdots & \frac{\partial}{\partial x_n} [x_n + b_n] \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{bmatrix} = Ix$$

Theorem 2.4. ***Examples:***
Let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ such that:

$$T_A(x) := \begin{bmatrix} \langle a_1, x \rangle \\ \vdots \\ \langle a_m, x \rangle \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1,j} x_j \\ \vdots \\ \sum_{j=1}^n a_{m,j} x_j \end{bmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = Ax$$

Then:

$$J_{T_A}(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} \left[\sum_{j=1}^n a_{1,j} x_j \right] & \cdots & \frac{\partial}{\partial x_n} \left[\sum_{j=1}^n a_{1,j} x_j \right] \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} \left[\sum_{j=1}^n a_{m,j} x_j \right] & \cdots & \frac{\partial}{\partial x_n} \left[\sum_{j=1}^n a_{m,j} x_j \right] \end{bmatrix} = \begin{bmatrix} a_{1,1} x_1 & \cdots & a_{1,n} x_n \\ \vdots & \ddots & \vdots \\ a_{m,1} x_1 & \cdots & a_{m,n} x_n \end{bmatrix}$$

3 Surjective Continuous Non-decreasing Bounded Functionals

Let $B = \{f : \mathbb{R} \to [0,1] | f \text{ is surjective, continuous, and non-decreasing.}\}$

Theorem 3.1. ***B is convex.***
Let $f, g \in B$ and $h(x) := \lambda f(x) + (1-\lambda)g(x)$ where $\lambda \in [0,1]$
Then h is still continuous since the linear combination of continuous functions is continuous.

Since both f and g are surjective and non-decreasing, then there exists x_0, y_0, x_1, y_1 in $\mathbb{R}$ such that:
$f(x_0) = 0 = g(y_0)$ and $f(x_1) = 1 = g(y_1)$
Suppose WLOG that $x_0 \leq y_0$ and $x_1 \leq y_1$
Then we know that:

$$h(x_0) = \lambda f(x_0) + (1-\lambda)g(x_0) = \lambda 0 + (1-\lambda)0 = 0$$

and

$$h(y_1) = \lambda f(y_1) + (1-\lambda)g(y_1) = \lambda 1 + (1-\lambda)1 = 1$$

Now if we pick $\alpha \in [0,1]$ by the intermediate value theorem, we know that there exists an $x_\alpha \in [x_0, y_1]$ such that:

$$h(x_\alpha) = \alpha$$

Since α was arbitrary element, I have shown that h is surjective.

Finally, let $x_0 < x_1$ be elements in $\mathbb{R}$
Then we know that $f(x_0) \leq f(x_1)$ and $g(x_0) \leq g(x_1)$
$\Rightarrow \lambda f(x_0) \leq \lambda f(x_1)$ *and* $(1-\lambda)g(x_0) \leq (1-\lambda)g(x_1)$
$\Rightarrow \lambda f(x_0) + (1-\lambda)g(x_0) \leq \lambda f(x_1) + (1-\lambda)g(x_1)$
$\Rightarrow h(x_0) \leq h(x_1)$
Thus h is non-decreasing.
Since h is surjective, continuous, and non-decreasing, then $h \in B$
Thus B is convex.

Theorem 3.2. ***B is translation invariant.***
Let $f \in B$ and $g(x) := f(x+c)$ where $c \in \mathbb{R}$
f is continuous and so is the addition operator so g is continuous.
Let $\alpha \in [0,1]$ since f is surjective then $\exists x \in \mathbb{R} \pitchfork f(x) = \alpha$
Then $g(x-c) = f(x+c-c) = f(x) = \alpha$ and so g is surjective.
Let $x < y$ be elements in $\mathbb{R}$
Then $f(x) \leq f(y) \Rightarrow f(x+c) \leq f(y+c)$
$\Rightarrow g(x) \leq f(y)$ *and so g is non-decreasing.*
Thus $g \in B$ and B is therefore translation invariant.

Theorem 3.3. ***B is not complete.***

Theorem 3.4. ***Every element in B can be decomposed as a finite non-trivial convex combination from B***