max planck institut informatik

High Level Computer Vision
Deep Neural Networks and Backpropagation
Exercise Introduction

Seong Joon Oh

Proving minimality of a function

Prove that function f(u) defined on S attains minimum at $u=u^*$.

(1) Using the first-order condition

- If f is differentiable and S is closed & bounded, then the minimum is attained on the set $\{u \mid f'(u)=0\} \cup \mathrm{bdy}(S)$.

(2) Lower bounding f

- If f is bounded from below: $f(u) >= V$ for all u in S, and if $f(u^*)=V$ for some u^* in S, then the statement above holds.

Smoothness of composite functions

If *f* is smooth and *g* is smooth, then *g* o *f* is also smooth.

"Smooth":

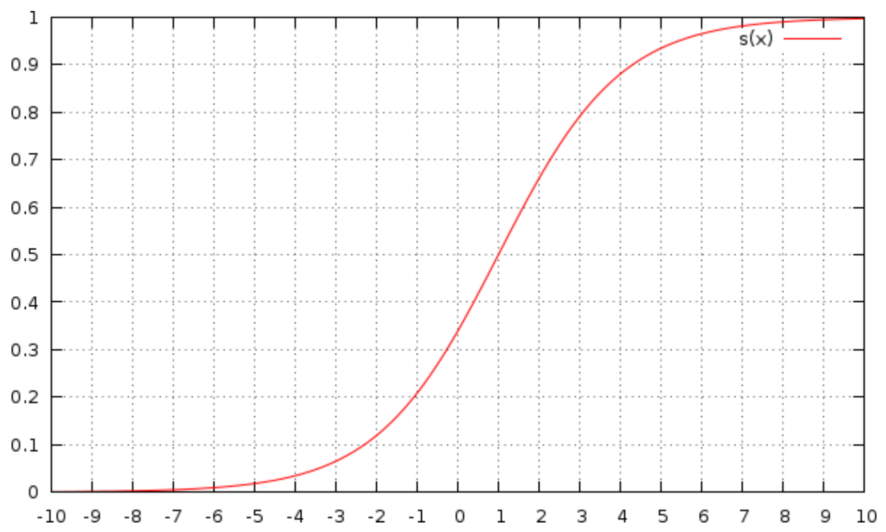
- differentiable, twice differentiable, …, infinitely differentiable (C^∞).
- continuously differentiable (C^1), twice continuously differentiable (C^2), …, infinitely differentiable (C^∞).

Our neural network is C^∞.

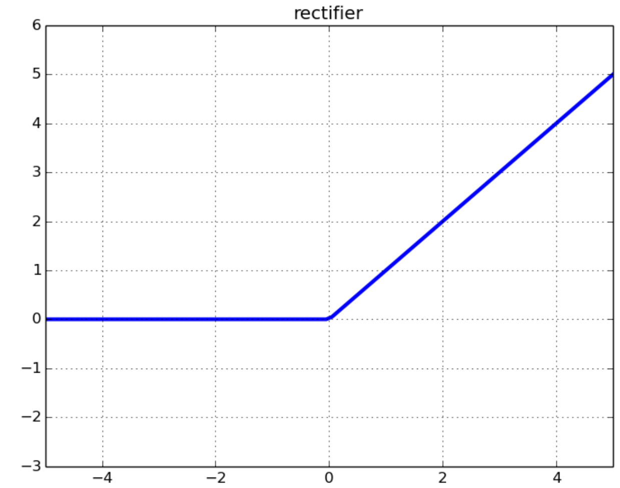
Sub-differentiability

For gradient descent algorithm, not even differentiability is needed.

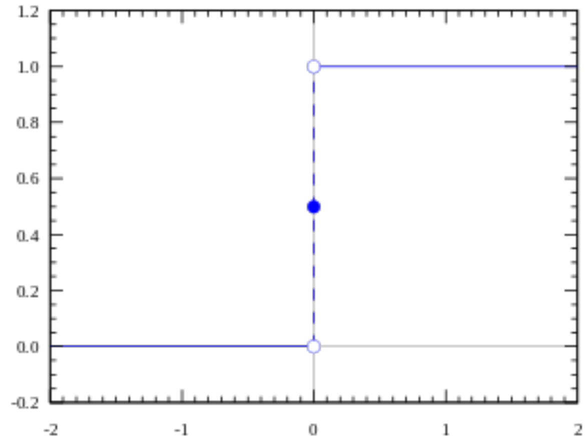
Only needs to be sub-differentiable: derivative exists almost everywhere. (covers many functions with practical usage)



Infinitely differentiable



Sub-differentiable



Sub-differentiable (but grad = 0 a.e.)

Multivariate chain rule

$$f : \mathbb{R}^M \to \mathbb{R}^N$$
$$g : \mathbb{R}^L \to \mathbb{R}^M$$

$$\left.\frac{\partial \left(f \circ g\right)_i (x)}{\partial x_k}\right|_{x=u} = \sum_{j=1}^{M} \left.\frac{\partial f_i(y)}{\partial y_j}\right|_{y=g(u)} \left.\frac{\partial g_j(x)}{\partial x_k}\right|_{x=u}$$

Example

$$f : \mathbb{R}^3 \to \mathbb{R}$$
$$g : \mathbb{R}^2 \to \mathbb{R}^3$$

$$f(y) = \sum_{p=1}^{3} y_p^2$$

$$g_p(x) = \sum_{q=1}^{2} w_{pq} x_q^2$$

Example, continued

$$
\begin{aligned}
\frac{\partial f(y)}{\partial y_j} &= \frac{\partial}{\partial y_j} \sum_{p=1}^{3} y_p^2 \\
&= \sum_{p=1}^{3} \frac{\partial}{\partial y_j} y_p^2 \\
&= \sum_{p=1}^{3} 2 y_p \delta_{jp} \\
&= 2 y_j
\end{aligned}
$$

Example, continued

$$\nabla f(y) = \qquad\qquad\qquad\qquad 2y$$

Example, continued

$$\begin{aligned}
\frac{\partial g_j(x)}{\partial x_k} &= \frac{\partial}{\partial x_k} \sum_{q=1}^{2} w_{jq} x_q^2 \\
&= \sum_{q=1}^{2} \frac{\partial}{\partial x_k} \left(w_{jq} x_q^2 \right) \\
&= \sum_{q=1}^{2} (2 w_{jq} x_q \delta_{kq}) \\
&= 2 w_{jk} x_k
\end{aligned}$$

Example, continued

$$\left.\frac{\partial\left(f\circ g\right)(x)}{\partial x_k}\right|_{x=u} = \sum_{j=1}^{3}\left.\frac{\partial f(y)}{\partial y_j}\right|_{y=g(u)}\left.\frac{\partial g_j(x)}{\partial x_k}\right|_{x=u}$$

$$= \sum_{j=1}^{3} 2g_j(u)\left(2w_{jk}u_k\right)$$

$$= 4u_k\sum_{j=1}^{3} w_{jk}g_j(u)$$

$$= 4u_k\sum_{j=1}^{3} w_{jk}\sum_{q=1}^{2} w_{jq}u_q^2$$

Example, continued

$$\nabla(f \circ g)(x) = \qquad 4u \bullet \left(W^T W \, (u \bullet u)\right)$$