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\} U bdy(S).

(2) **Lower bounding $f$**
   - If $f$ is bounded from below: $f(u) \geq V$ 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 $g \circ f$ is also smooth.

“Smooth”: 
- differentiable, twice differentiable, ..., infinitely differentiable ($C^\infty$).
- continuously differentiable ($C^1$), twice continuously differentiable ($C^2$), ..., infinitely differentiable ($C^\infty$).

Our neural network is $C^\infty$. 
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

\[ f : \mathbb{R}^M \rightarrow \mathbb{R}^N \]
\[ g : \mathbb{R}^L \rightarrow \mathbb{R}^M \]

\[ \frac{\partial (f \circ g)_i(x)}{\partial x_k} \bigg|_{x=u} = \sum_{j=1}^{M} \frac{\partial f_i(y)}{\partial y_j} \bigg|_{y=g(u)} \frac{\partial g_j(x)}{\partial x_k} \bigg|_{x=u} \]
Example

\[ f : \mathbb{R}^3 \rightarrow \mathbb{R} \]
\[ g : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \]

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

\[ g_p(x) = \sum_{q=1}^{2} w_{pq} x_q^2 \]
Example, continued

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

$$= \sum_{p=1}^{3} \frac{\partial}{\partial y_j} y_p^2$$

$$= \sum_{p=1}^{3} 2y_p \delta_{jp}$$

$$= 2y_j$$
Example, continued

\[ \nabla f(y) = 2y \]
Example, continued

\[ \frac{\partial g_j(x)}{\partial x_k} = \] 

\[ = \sum_{q=1}^{2} \frac{\partial}{\partial x_k} \left( w_{jq} x_q^2 \right) \]

\[ = \sum_{q=1}^{2} \left( 2 w_{jq} x_q \delta_{kq} \right) \]

\[ = 2 w_{jk} x_k \]

\[ \sum_{q=1}^{2} w_{jq} x_q \]
Example, continued

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

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

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

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

\[ \nabla (f \circ g)(x) = 4u \cdot (W^T W (u \cdot u)) \]