

通信理論に特化した深層学習

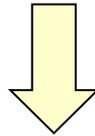
第6回ゼミ資料

誤差逆伝播法

豊橋技術科学大学
電気・電子情報工学系
准教授 竹内啓悟

勾配の計算

損失関数を局所最小化するためには、順伝播型ネットワークに含まれるすべてのパラメータに関する損失関数の偏微分を効率的に計算する必要がある。



「深層」合成関数のパラメータに関する偏微分を効率的に計算する問題

「深層」合成関数

$g^t(\cdot, \theta^t): \mathbb{R}^{N_{t-1}} \rightarrow \mathbb{R}^{N_t}$ をパラメータ θ^t を持つ微分可能なベクトル値関数とし、

$$G_{t_1}^{t_2}(\mathbf{z}, \Theta_{t_1}^{t_2}) = g^{t_2} \circ g^{t_2-1} \circ \dots \circ g^{t_1}(\mathbf{z}, \theta^{t_1}), \quad \Theta_{t_1}^{t_2} = \{\theta^{t_1}, \dots, \theta^{t_2}\}.$$

注意

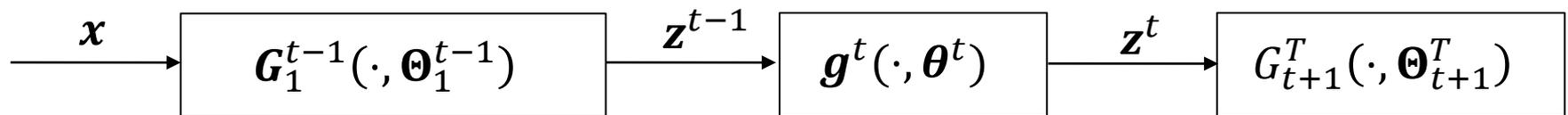
$N_T = 1$ として、スカラー関数 $g^T(\cdot, \theta^T): \mathbb{R}^{N_{T-1}} \rightarrow \mathbb{R}$ に損失関数を含める。

パラメータ Θ_1^T に関する $G_1^T(x, \Theta_1^T)$ のすべての偏微分を効率的に計算したい。

偏微分の計算

偏微分の連鎖測

t 番目の関数のパラメータ θ^t に注目する。



$$G_1^T(x, \Theta_1^T) = G_{t+1}^T(g^t(z^{t-1}, \theta^t), \Theta_{t+1}^T), \quad z^t = G_1^t(x, \Theta_1^t).$$

上記の表現の両辺を $\theta_i^t = [\theta^t]_i$ に関して偏微分すると、

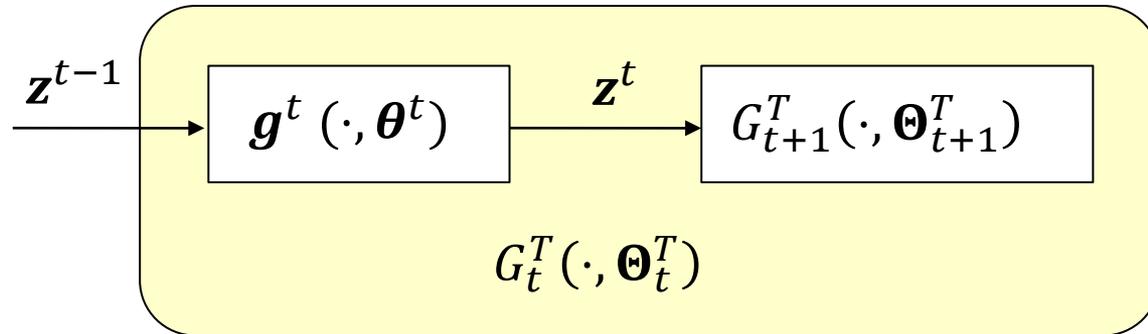
$$\frac{\partial}{\partial \theta_i^t} G_1^T(x, \Theta_1^T) = \sum_{n=1}^{N_t} \partial_n G_{t+1}^T(z^t, \Theta_{t+1}^T) \frac{\partial g_n^t}{\partial \theta_i^t}(z^{t-1}, \theta^t), \quad g_n^t = [g^t]_n.$$

後者の因子は順伝播時に計算可能

逆伝播

表記法

関数 f の n 番目の変数に関する偏微分を $\partial_n f$ と書く。



$$G_t^T(\mathbf{z}, \Theta_t^T) = G_{t+1}^T(g^t(\mathbf{z}, \theta^t), \Theta_{t+1}^T)$$

上記の表現の両辺を $z_n = [\mathbf{z}]_n$ に関して点 $\mathbf{z} = \mathbf{z}^{t-1}$ で偏微分すると、

$$\partial_n G_t^T(\mathbf{z}^{t-1}, \Theta_t^T) = \sum_{n'=1}^{N_t} \partial_n g_{n'}^t(\mathbf{z}^{t-1}, \theta^t) \partial_{n'} G_{t+1}^T(\mathbf{z}^t, \Theta_{t+1}^T)$$

青色部分は逆伝播時に逐次的に計算可能

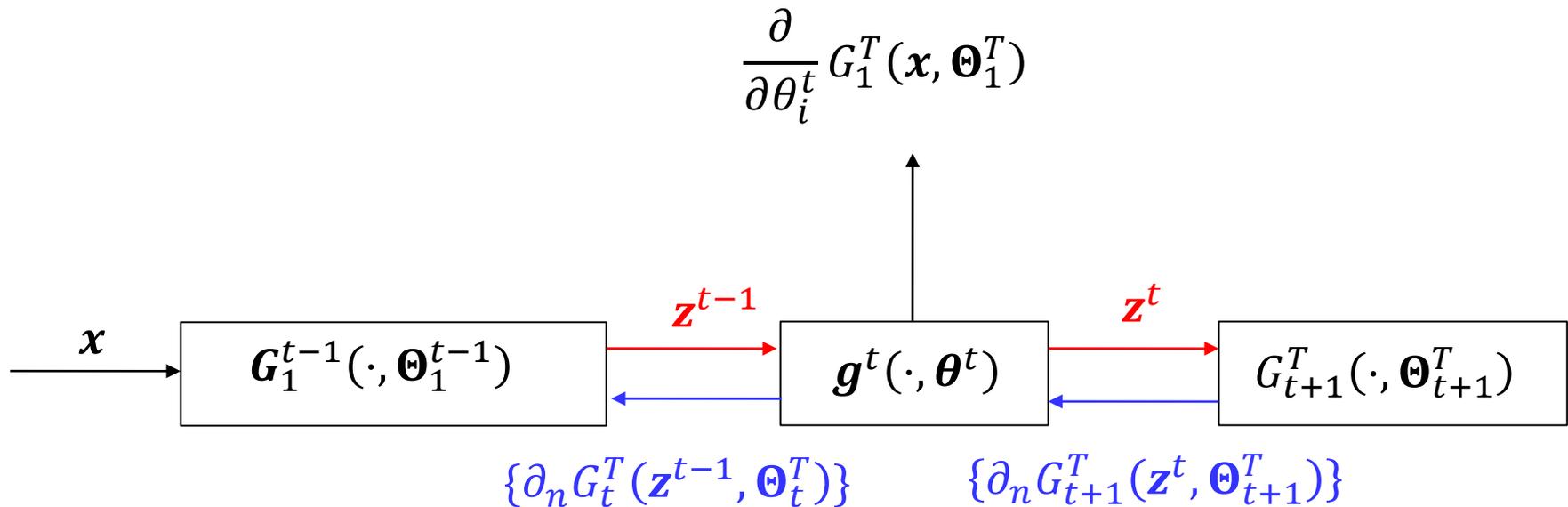
誤差逆伝播法 (Back propagation)

手順1

順伝播により、 $\{\mathbf{z}^t, \partial \mathbf{g}^t(\mathbf{z}^{t-1}, \boldsymbol{\theta}^t) / \partial \theta_i^t, \partial_n \mathbf{g}^t(\mathbf{z}^{t-1}, \boldsymbol{\theta}^t)\}$ を計算する。

手順2

逆伝播により、出力側から勾配を順に計算する。

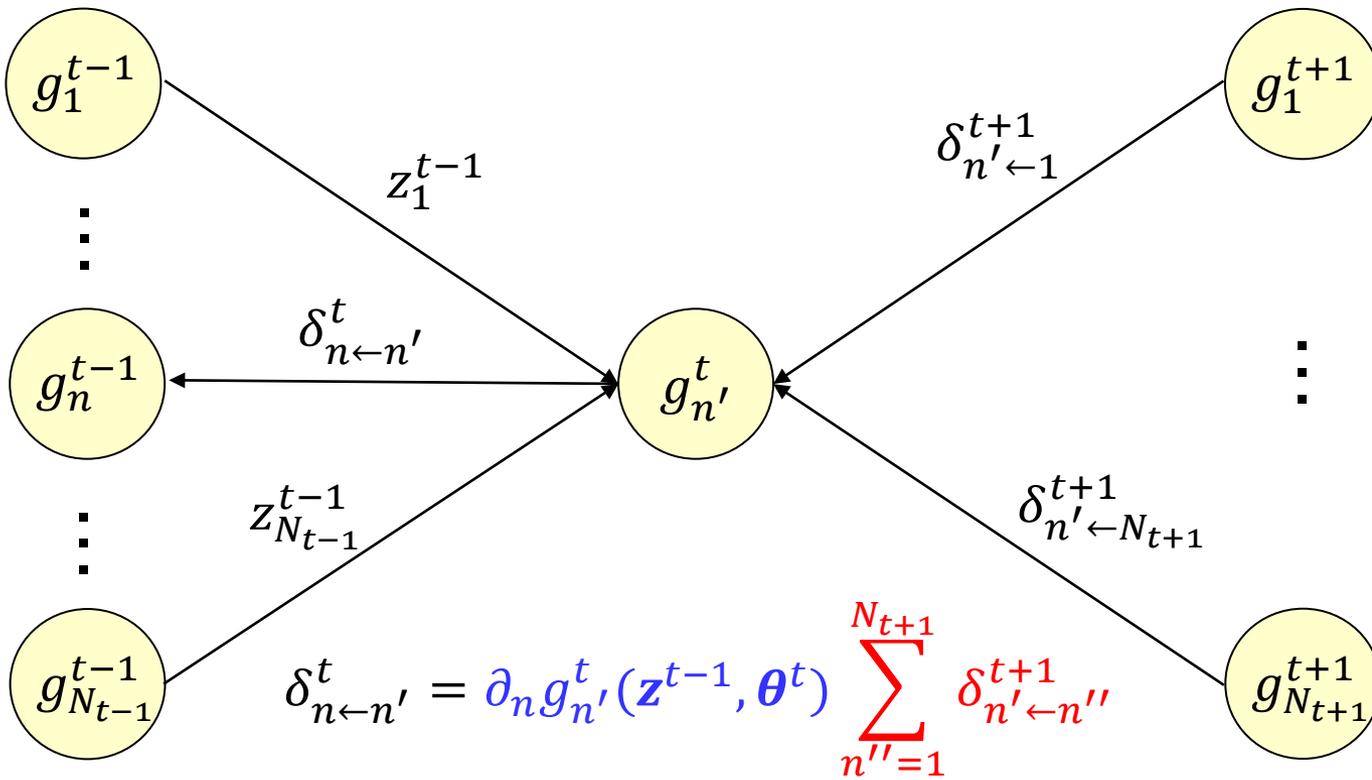


計算グラフ

計算結果が誤差逆伝播法による計算結果と等価になるように、ノード間でやり取りされるメッセージを定義する。

関数 g_n^t を t 層 n 番目のノードに対応付ける。

$\delta_{n \leftarrow n'}^t$: t 層 n' 番目のノードから $t - 1$ 層 n 番目のノードに送られるメッセージ



入力メッセージの和を取り、宛先ノードに関する偏微分を乗算する。

計算結果の等価性

誤差逆伝播法

$$\partial_n G_t^T(\mathbf{z}^{t-1}, \Theta_t^T) = \sum_{n'=1}^{N_t} \partial_n g_{n'}^t(\mathbf{z}^{t-1}, \boldsymbol{\theta}^t) \partial_{n'} G_{t+1}^T(\mathbf{z}^t, \Theta_{t+1}^T).$$

計算グラフ

$$\delta_{n \leftarrow n'}^t = \partial_n g_{n'}^t(\mathbf{z}^{t-1}, \boldsymbol{\theta}^t) \partial_{n'} G_{t+1}^T(\mathbf{z}^t, \Theta_{t+1}^T).$$

帰納法による証明のスケッチ

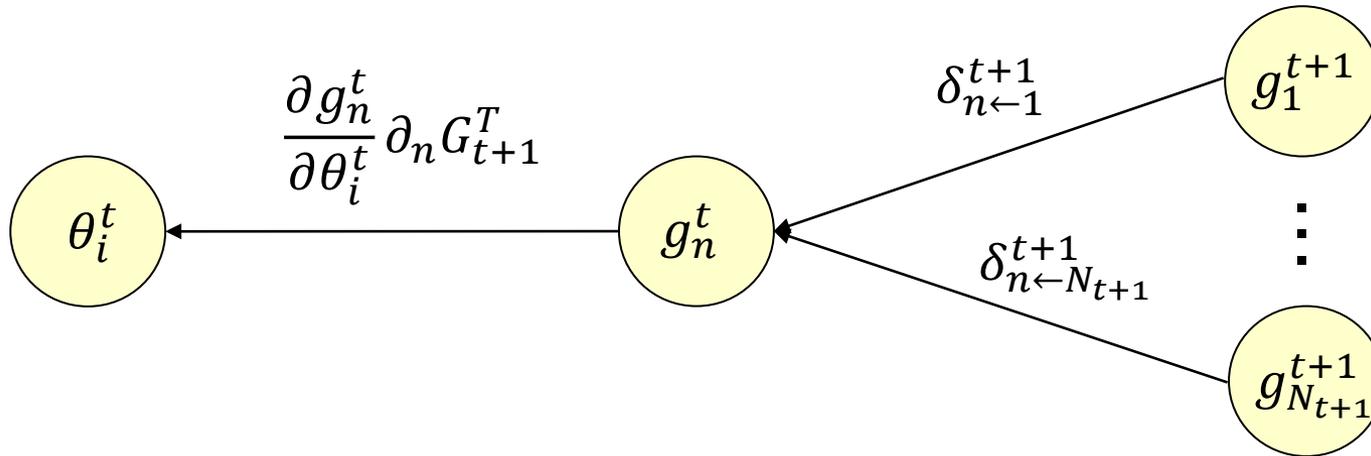
$$\begin{aligned} \delta_{n \leftarrow n'}^{t-1} &= \partial_n g_{n'}^{t-1}(\mathbf{z}^{t-2}, \boldsymbol{\theta}^{t-1}) \sum_{n''=1}^{N_t} \delta_{n' \leftarrow n''}^t \\ &= \partial_n g_{n'}^{t-1}(\mathbf{z}^{t-2}, \boldsymbol{\theta}^{t-1}) \partial_{n'} G_t^T(\mathbf{z}^{t-1}, \Theta_t^T). \end{aligned}$$

最後の等号は、帰納法の仮定と逆伝播の定義式から従う。 ■

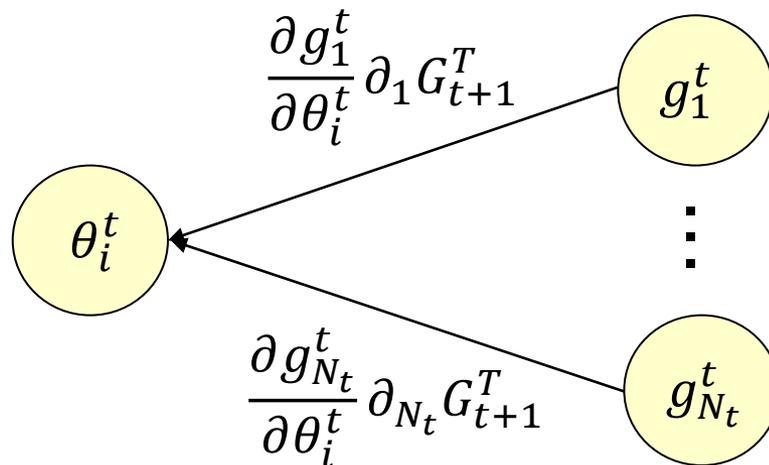
三つの添え字を持つメッセージ $\delta_{n \leftarrow n'}^t$ (3階のテンソル) が計算グラフ上を流れる。

TensorFlowの名前の由来

計算グラフによる勾配計算



入力メッセージの和を取り、宛先ノードに関する偏微分をかける。



入力メッセージの和を取る。

$$\frac{\partial}{\partial \theta_i^t} G_1^T(\mathbf{x}, \Theta_1^T) = \sum_{n=1}^{N_t} \partial_n G_{t+1}^T(\mathbf{z}^t, \Theta_{t+1}^T) \frac{\partial g_n^t}{\partial \theta_i^t}(\mathbf{z}^{t-1}, \theta^t).$$