通信理論に特化した深層学習
第1回ゼミ資料
はじめに

豊橋技術科学大学
電気・電子情報工学系
准教授　竹内啓悟

目標

通信の復調方式として利用されているメッセージ伝播法に基づく
深層ニューラルネットワークの最先端の学習方法を学ぶ。

原理に関する方針

- 核心部分をおおまかに理解することを目指す。

Q 深層学習の本質は何か？

A 大自由度な損失関数の確率的局所最小化である。

ポイント：

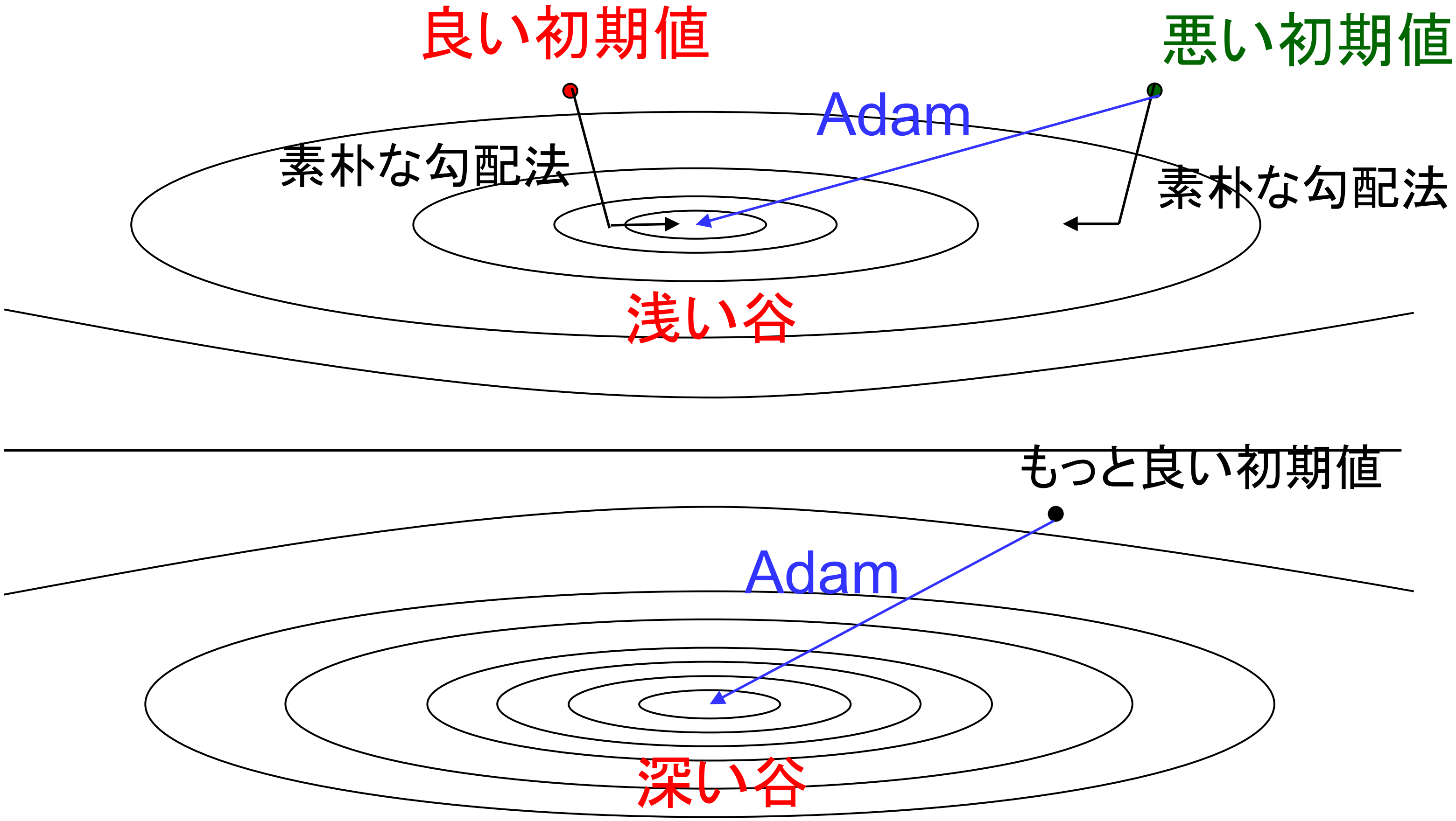
・良い解を見つけやすい関数の構造（畳み込みニューラルネットワーク）

・良い解を見つけやすい初期値（メッセージ伝播法）

・良い解を見つけやすいアルゴリズム（Adam）

このような簡潔な回答をできるようになるのが目標

学習のイメージ図（等高線図）



関数の構造は地形（谷の個数）に影響を与える。

プログラミングに関する方針

TensorFlow 2を利用したPython 3によるコードを作成する。

方針

Pythonの分厚い教科書を読むところからは始めない。

- 何を作りたいのかをまず明確にする。
- 作りたいものを作るのに必要最低限の知識を明確にする。
- 必要になって初めて知らないことを調べる。

対象外の内容

- 実データを使った学習
- 畳み込みニューラルネットワーク
- Python 3の基礎

最初からTensorFlow 2のコードを書いて、動かして、理解する。

TensorFlow 2について

Googleが開発した（主に）Pythonの機械学習用のライブラリ

できること

- 任意の計算グラフの作成
- 学習する変数と学習しない変数の設定
- 誤差逆伝播法による学習
- Kerasという高水準Application Programming Interface (API)を使った簡潔なソートコード記述

TensorFlow 2.0

Define-by-run方式による数学的に自然なソースコード記述

以前のバージョンは、Define-and-runという評判の悪い方式であった。

Webを参考にする際は、バージョンをよく確認すること。

インストールと実行方法

Python 3全般

Debianにログイン後、root権限で以下のコマンドを実行する。

```
apt install python3-pip
```

TensorFlow 2.0

仮想環境を使わないオプション

```
pip3 install --upgrade pip --break-system-packages
```

pipの最新版をインストール

```
pip install tensorflow --break-system-packages
```

インストールしたpipを使用

実行方法

後述するソースコード***.pyを作成した上で

```
python3 ***.py
```

デフォルトがPython 2なので、「3」を忘れないこと。

加法的白色ガウス雑音通信路

$$y = x + w, \qquad w \sim \mathcal{N}(0, \sigma^2).$$

二位相偏移変調（Binary phase-shift keying, BPSK）

$$\mathbb{P}(x = \pm 1) = \frac{1}{2}.$$

加法的白色ガウス雑音（Additive white Gaussian noise, AWGN）

wはxと独立な平均0分散σ^2のガウス分布に従う。

平均$\mu = \mathbb{E}[z]$分散$\sigma^2 = \mathbb{E}[(z-\mu)^2]$のガウス確率変数$z$の確率密度関数

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}, \qquad z \in \mathbb{R}.$$

推定量と性能の評価基準

推定量（Estimator）

受信信号$y \in \mathbb{R}$の任意の関数$\hat{x}(y)$

平均二乗誤差（Mean square error, MSE）

$$\mathbb{E}[\{x-\hat{x}(y)\}^2] = \mathbb{E}_x\left[\int \{x-\hat{x}(y)\}^2 p(y|x)dy\right].$$

条件付き確率密度関数

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}}.$$

∵　ガウス性はノイズのガウス性から従う。条件付き平均と分散を計算すると、

定義$y = x + w$、期待値の線形性、xとwの独立性、wの定義から、

$$\mathbb{E}[y|x] = \mathbb{E}[x+w|x] = x + \mathbb{E}[w|x] = x + \mathbb{E}[w] = x,$$

$$\mathbb{E}[(y-\mathbb{E}[y|x])^2|x] = \mathbb{E}[w^2|x] = \mathbb{E}[w^2] = \sigma^2.$$

最適な推定量

事後平均推定量（Posterior mean estimator, PME）

$$\hat{x}_{\mathrm{PME}}(y) = \mathbb{E}[x|y].$$

定理1.1：　事後平均推定量は平均二乗誤差を最小にする。

証明：　$\hat{x}$をxの任意の推定量とする。

$$\mathbb{E}[(x-\hat{x})^2] = \mathbb{E}_y\left\{\mathbb{E}_{x|y}[\{(x-\hat{x}_{\mathrm{PME}}) + (\hat{x}_{\mathrm{PME}} - \hat{x})\}^2]\right\}$$

推定量$\hat{x}$はyの関数なので、条件付き平均$\mathbb{E}_{x|y}$の計算では定数とみなせる。

$\hat{x}_{\mathrm{PME}} = \mathbb{E}[x\,|y]$なので、平方を展開すると、$2(x-\hat{x}_{\mathrm{PME}})(\hat{x}_{\mathrm{PME}}-\hat{x})$は消える。

$$\mathbb{E}[(x-\hat{x})^2] = \mathbb{E}[(x-\hat{x}_{\mathrm{PME}})^2] + \mathbb{E}[(\hat{x}_{\mathrm{PME}}-\hat{x})^2] \geq \mathbb{E}[(x-\hat{x}_{\mathrm{PME}})^2]$$

等号は$\hat{x} = \hat{x}_{\mathrm{PME}}$のときに限る。　■

事後平均推定量の計算

ベイズの公式

$$\mathbb{P}(x|y) = \frac{p(y|x)\mathbb{P}(x)}{p(y)}$$

$$\because p(x,y) = p(y|x)\mathbb{P}(x) = \mathbb{P}(x|y)p(y).$$

$p(y)$の定義

$$p(y) = \sum_{x=\pm 1} p(y|x)\mathbb{P}(x).$$

事後平均推定量の計算

$$\mathbb{E}[x|y] = \sum_{x=\pm 1} x\mathbb{P}(x|y) = \frac{\sum_{x=\pm 1} xp(y|x)\mathbb{P}(x)}{\sum_{x=\pm 1} p(y|x)\mathbb{P}(x)}$$

$$= \frac{\sum_{x=\pm 1} x\exp\left\{-\frac{(y-x)^2}{2\sigma^2}\right\}}{\sum_{x=\pm 1} \exp\left\{-\frac{(y-x)^2}{2\sigma^2}\right\}} = \frac{e^{\frac{y}{\sigma^2}} - e^{-\frac{y}{\sigma^2}}}{e^{\frac{y}{\sigma^2}} + e^{-\frac{y}{\sigma^2}}} = \tanh\left(\frac{y}{\sigma^2}\right).$$