

ICT リテラシー (情報技術論) B

-- 第 10 回 : 深層学習の基本 --

栗野 俊一

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く
禁じます

2023/11/14 ICT リテラシー (情報技術論) B

伝言

私語は慎むように !!

□ 席は自由です

- できるだけ前に詰めよう
- コロナ対策のために、ソーシャルディスタンスをたもとう

□ 色々なお知らせについて

- 栗野の Web Page に注意する事

<http://edu-gw2.math.cst.nihon-u.ac.jp/~kurino>

- google で「kurino」で検索

前回 (第 09 回) の復習

ICT リテラシー (情報技術論) B

前回 (第 09 回) の復習

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

前回 (第 09 回) の復習

□ 前回 (第 09 回) の復習

○ 講義：ニューラルネットワーク

- ▶ 人間の脳：ニューロン(神経細胞)のネットワーク (シナプスによる信号制御)
- ▶ (コンピュータの)ニューロン：多数の入力の総和に、しきい値をこえたかどうかを判定
- ▶ パーセプトロン：単一ニューロンに重み付け (超平面による分割)
- ▶ ニューラルネットワーク：人間の脳のモデル化(重みで信号制御)

今週 (第 10 回) の概要

ICT リテラシー (情報技術論) B

今週 (第 10 回) の概要

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

今週 (第 10 回) の予定

□ 今週 (第 10 回) の予定

○ 講義：深層学習の基本 (Text p.93, 7.4 節)

- ▶ ニューラルネットワークにおける学習
- ▶ 深層学習 (バックプロパゲーション：誤差逆伝播法)
- ▶ 学習教師なし学習

今週 (第 10 回) の目標

□ 今週 (第 10 回) の目標

- 深層学習がどのように行われるかを理解する

今週 (第 10 回) の課題

- 前回 (第 09 回) の課題
 - Web Class「小テスト-09」
- 今週 (第 10 回) の課題
 - Web Class「小テスト-10」

深層学習の基本

ICT リテラシー (情報技術論) B

深層学習の基本

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

ニューラルネットワークにおける学習

□ ニューラルネットワークの構成要素

○ ネットワークの構造 (モデル) : 隣りの層とは全結合/情報は一方向

▶ 幾つの層からなるか

▶ 個々の層のニューロンの個数は幾つか

▶ 個々の層のニューロンの活性化関数は何か

○ 重み

▶ 個々のニューロン間を結ぶリンクに付けられた重みは幾つか

□ ニューラルネットワークにおける学習

○ 内部状態は、「重み」だけ (モデルは変化しない)

▶ 望みの形になる(学習データに適合する)ように、「重み」を変更すればよい

ニューラルネットワークの教師あり学習

□ ニューラルネットワークの教師あり学習

○ 学習データには、課題と正解ラベルがある

- ▶ 現在の状態で、課題を解き(予想し)、正解ラベルと比較
- ▶ より良い予想ができるように、重みを修正する

○ ニューロンの重みの修正

- ▶ 出力の誤差を小さくするように、重みを調整する

勾配降下法

□ 勾配降下法：微分可能な関数 $y=f(x)$ の極小値を求める方法

○ 微分係数： $f'(a) = \lim (f(a+h)-f(a))/h = \lim (f(a+h)-f(a))/((a+h)-a)$

▷ x を a から $a+h$ に変化させた時の f の変化 $f(a+h)-f(a)$ は、近似的に $f'(a)$ に比例する

▷ $f'(a)$ を利用して、 y の誤差から x の誤差が計算できる(逆伝播)

○ 解の候補 a を利用して、次の候補 a' を $a' = a - h f'(a)$ で改良 (h は学習率 / f' は f の微分)

○ $f'(x)$ が x に対する y の変化の割合

▷ $f'(x)$ が 0 になる方向に x を修正する

▷ 特に、 $f'(x) > 0$ なら減らし、 $f'(x) < 0$ なら増やす ($f'(x) = 0$ は答え)

○ 欠点

▷ 最小値ではなく、極小値になる可能性がある(Local Minimum 問題)

▷ 学習率が大きいと発散し、小さいと収束が遅い

□ 確率的勾配降下法

○ 学習率を調整して、発散せずに、収束を速くする工夫

ニューロンの学習

□ニューロンの学習

○ニューロンの振舞い (入力 x_1, x_2, \dots, x_k / 重み w_1, w_2, \dots, w_k / 出力

▷ $y = f(x_1, x_2, \dots, x_k) = F(\sum w_i x_i)$ [F は活性化関数]

▷ y は、 x_i の関数と定義されるが、 w_i の関数と見なす事ができる

○ニューロンへの勾配降下法の適用

▷ y の誤差を減らすには、 w_i をどう変化させれば良いか？

▷ f を w_i で偏微分 (w_i 以外は定数と考えると、 f は w_i の一変数関数なので、それを w_i で微分)

○微分の連鎖律の利用

▷ $y' = f'(w_i) = F'(w_i) * (\sum w_i x_i') = F'(w_i) * x_i$

▷ 更に $F(x) = \text{ReLU}(x)$ ならば、 $F'(x) = 0 (x < 0)$ または $1 (x \geq 0)$

バックプロパゲーション (誤差逆伝播法)

□ 発想

○ ニューロンの学習 : ニューロンの出力の誤差が分れば、重みの誤差が解る

▶ ニューロンの出力の誤差が分れば、ニューロンの入力の誤差が解る

▶ (出力層の)誤差を(深い層から浅い層に..)逆伝播させればよい

○ 微分の連鎖律の利用

▶ $y' = f'(xi) = F'(xi) * (\sum wi xi)' = F'(xi) * wi$

□ バックプロパゲーション (誤差逆伝播法)

○ 出力層の誤差 (ニューラルネットワークの予想と教師ラベルの差) を逆伝播させる

▶ 誤差の逆伝播と、重みの調整を並行して行う

教師なし Deep Learning

□ Deep Learning の制約

- Deep Learning は、基本、教師あり

- ▶ バックプロパゲーションが、出力の誤差 (予想と正解ラベルの比較) を利用するから

□ 教師なし Deep Learning

- 課題から、教師データに相当するものを作り出す (例 : k-means)

- ▶ 「最小化したい目標」があれば、誤差計算はできる

- ▶ 目標関数が微分可能なら、正解(最小値)が解らなくても、誤差[修正の方針]は(微分係数を用いて近似)計算できる

□ オートエンコーダ (自己符号化器)

- 出力と同じ結果を正解とする : 途中の層が狭いならば、「次元圧縮」ができる

- ▶ 情報の生成にも利用可能

おしまい

ICT リテラシー (情報技術論) B

おしまい

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます