

ニューラルネットワークの学習

キーワード

損失関数

局所最適解、大域最適解

勾配降下法、確率的勾配降下法

ミニバッチ、バッチサイズ

ミニバッチ学習、オンライン学習、バッチ学習

学習率、エポック数、早期終了

損失関数

- **損失関数**はニューラルネットの出力と正解の間の誤差を表す
- **損失関数を最小化**するように、ネットワークのパラメータを学習する
- 関数の最小化問題に関わる概念は「微分」
- **損失関数が最小** = 損失関数を微分した値 (=勾配) がゼロとなる

∇L : 損失関数 L を偏微分した勾配値

$$\nabla L = \frac{\partial L}{\partial w}$$

w : パラメータ (重み)
各層において、 ∇L が小さくなるよう w の
組み合わせを求めれて、値を更新する

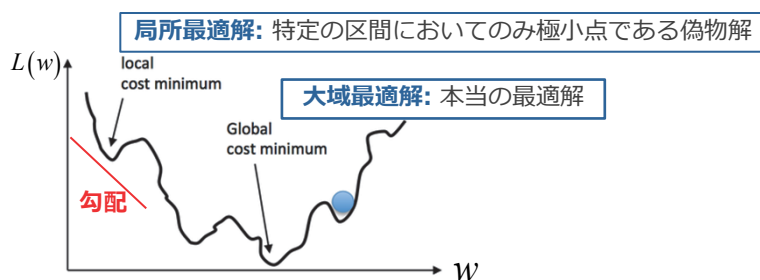
$$w \leftarrow w - \alpha \cdot \nabla L$$

- ニューラルネットの学習データが高次元でパラメータ数も膨大なので、解析的に最適解が求められないこともある
→ 代わりに、**勾配降下法**というアルゴリズムを利用して最適解を見つける

本講座の教材、画像、音声 等の無断使用を固く禁じます

勾配降下法①

- ニューラルネットワークの学習法の基本は**勾配降下法**
 - 損失関数の勾配 (接線の傾き) に沿って、最小点 (予測値と正解の誤差がゼロとなる点) の方向を目指して、パラメータ値を探索し、更新する

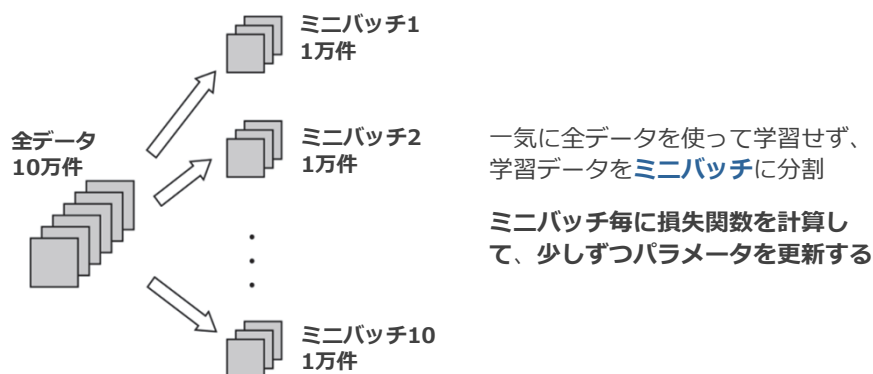


- 勾配降下法の難点:
 - **局所最適解にトラップ**され、**大域最適解にたどり着けない**ことがある
 - 勾配がゼロに近い平坦領域では、学習が停滞する**プラトー**に陥りやすい
- 上記を解消するために ① **学習データの渡し方** ② **学習率の設定**、を工夫することができる (詳細は後述)

本講座の教材、画像、音声 等の無断使用を固く禁じます

勾配降下法②

- ・ 局所最適解に陥る確率や異常値の影響を抑えるために、勾配降下法の一種である**確率的勾配降下法** (SGD, Stochastic Gradient Descent) がよく使われる



- ・ 損失関数は本来全サンプルに対する損失の平均
- ・ データ数が多い場合、ミニバッチに分けて学習する応報は、損失関数の計算時間の面でも利点はある

本講座の教材、画像、音声等の無断使用を固く禁じます

確率的勾配降下法に関連する用語

・ バッチサイズ

- 1つのミニバッチ内のデータ数、つまりモデルに一度に入力するデータ数
- 学習の際に設定するハイパーパラメータの1つ

・ オンライン学習

- 毎回サンプルを1つ取り出し、それを使ってパラメータを更新（バッチサイズ1）
- 局所解に陥りにくい
- 一つ一つのデータに対して更新を行うため、外れ値に敏感に反応し、結果が不安定になりやすいという欠点がある

・ ミニバッチ学習

- ニューラルネットワークでよく用いられる学習手法
- 毎回一部のサンプル（ミニバッチ）を取り出して、それを使ってパラメータを更新
- 一度に使うデータ数が比較的少ないため、パラメータの変化に対して損失関数が敏感に反応し、そのため、傾きが0となる地点が少なく学習が停滞しづらい
- ニューラルネットワークでよく用いられる学習手法

・ バッチ学習(最急降下法)

- 学習データの総数が少ないときに有効な学習手法
- 全データ数を一気に学習に使うため、学習結果が安定しやすい
- 一方で、新たな学習データを追加する度に、全データを用いて再度計算を行わなければならないので、計算コストが高い

本講座の教材、画像、音声等の無断使用を固く禁じます