

機械学習勉強メモ --- ADALINE編

ADALINEとは

- ADALINEはパーセプトロンアルゴリズムの改良である。
- パーセプトロンとの主な違いは重みの更新方法である。パーセプトロンは単位ステップ関数で重みを更新していたが、ADALINEでは線形活性化関数で重みを更新する。
- 線形活性化関数 $\phi(z)$ は単に層入力 z の高等関数である。

$$z = w^T x$$

$$\phi(z) = z$$

重みの学習方法

- ADALINE是最急勾配降下法を用いて学習が行われる。
- 勾配降下法とは、コスト関数が最小となるような重みを見つけることである。
- ADALINEで重みの学習に用いられるコスト関数 J は以下のように定義できる。

$$J(w) = \frac{1}{2} \sum_i (y^{(i)} - \phi(z^{(i)}))^2$$

- コストが最小となる重みの探し方は以下のように行われる。

$$w := w + \Delta w$$

- 重みの変化である Δw は以下の式で定義される。 η が学習率で、 $\nabla J(w)$ が勾配である。

$$\Delta w = -\eta \nabla J(w)$$

- 勾配の求め方は、各重み w_j ごとにコスト関数を偏微分することで求められる。

$$\frac{\partial J}{\partial w_j} = - \sum_i (y^{(i)} - \phi(z^{(i)})) x_j^{(i)}$$

勾配降下法の改善

- 勾配降下法には、以下のデメリットが存在する。
 - 重みを全て更新するため、大規模な学習では、計算コストが掛かる
- そのため、これらの改善として、特徴量のスケーリングや確率的勾配降下法、ミニバッチ降下法などが存在する。

特徴量のスケーリング

- スケーリングには、標準化と正規化が存在する。

標準化

- データの平均を0、標準偏差が1となるようにする正規化法である。

- σ :標準偏差, μ :平均である.

$$x_{new}^i = \frac{x_i - \mu}{\sigma}$$

正規化

- データの最大値が1, 最小値が0になるように行う正規化法である.

$$x_{new}^i = \frac{x^i - x_{min}}{x_{max} - x_{min}}$$

確率的勾配降下法

- 確率的勾配降下法は最急降下法のように全てのデータの重みを更新するのではなく, ランダムにデータを選んで重みを更新する方法である.

ミニバッチ勾配降下法

- ミニバッチ勾配降下法では, 全てのデータを更新するのではなく, 数個から数百個に区切られたバッチと言われるデータの塊をランダムに選択し, 重みを更新する方法である.

問題点

- 学習率が十分小さくないと収束しない
- 線形分離可能でないと収束しない