勾配降下法
+ Day4 カリキュラム

ペパボ研究所 三宅悠介 / GMO PEPABO inc.

2021.07.20　機械学習研修 Day4

カリキュラム概要と目標

- **Day3: 機械学習を自分の言葉で説明できる**
  - データから知識を学習させるための方法について、初歩的な知識を導入する
  - レッスン＆ハンズオンを通して、手を動かしながら知識の解釈を高める
- **Day4: 機械学習を使って課題を解決する流れを思い浮かべることができる**
  - 機械学習で課題を解決するために考慮すべき点について、基礎的な知識を導入する
  - レッスン＆ハンズオンを通して、手を動かしながら知識の解釈を高める
- **Day5: 機械学習をサービスに導入するまでの流れを思い浮かべることができる**
  - 配属先のサービスで機械学習を使った施策を実現するために、知っておくべき知識を導入する
  - Day1~4までの学びを用いて、擬似的なサービス導入準備をハンズオンで体験する

GMOペパボ

Day4: 機械学習を使って課題を解決する流れを思い浮かべることができる

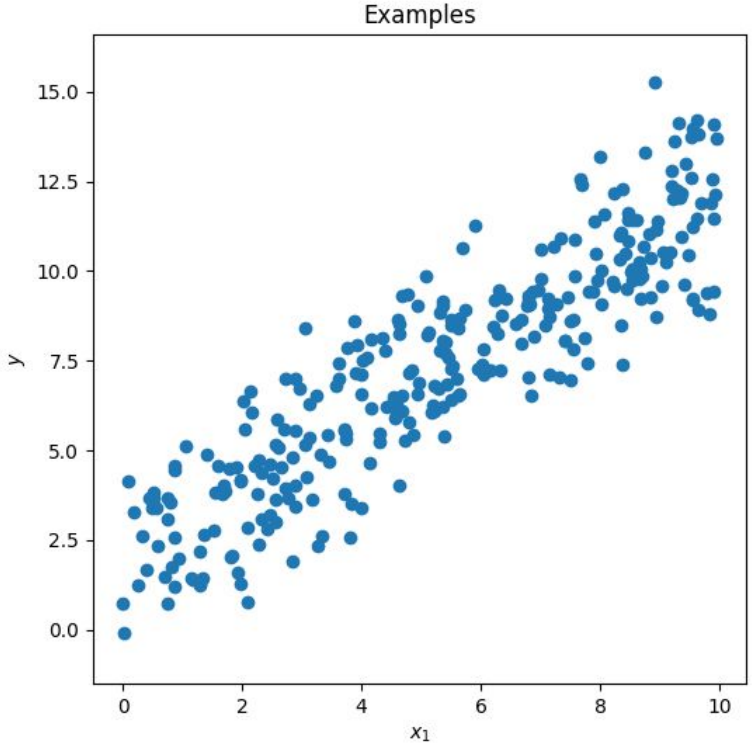
- 座学（勾配降下法）(15min）
- レッスン＆ハンズオン(160min）

勾配降下法

GMOペパボ

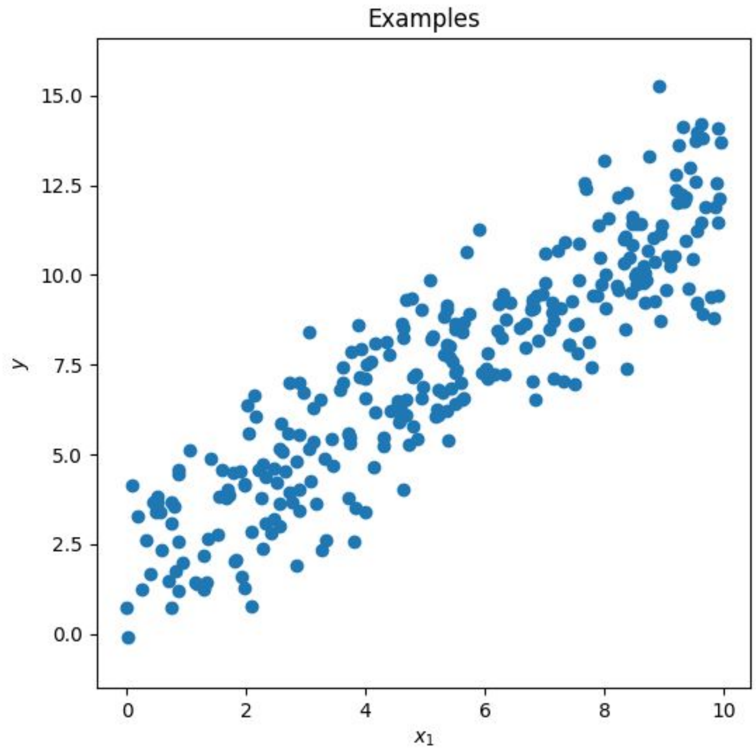
“はじめに”

やりたいことと、やりたくないこと



入力から出力を推測したい

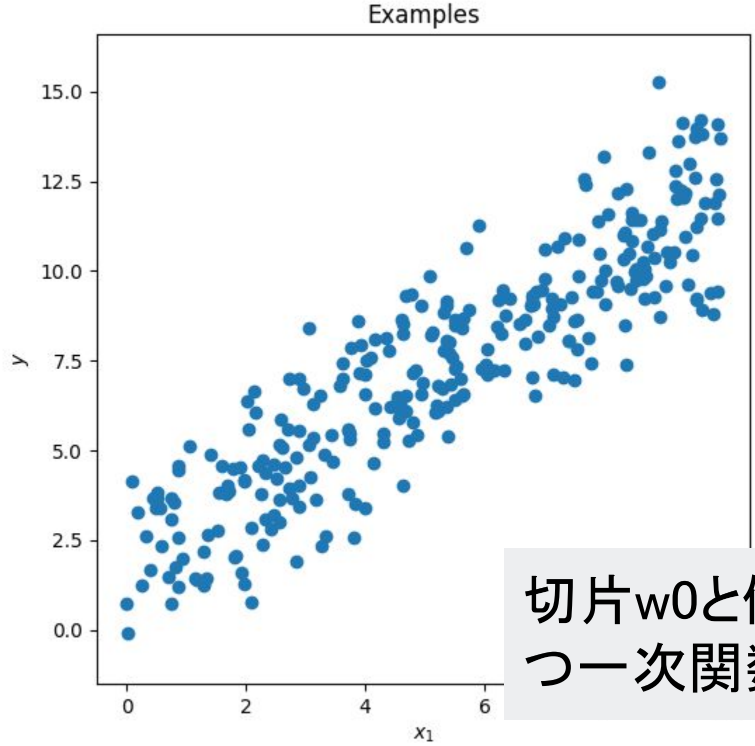
$$\boldsymbol{x} = \begin{pmatrix} x_0 = 1 \\ x_1 \end{pmatrix} \longrightarrow y'$$



入力から出力を推測したい

$$\boldsymbol{x} \longrightarrow f(\boldsymbol{x}) \longrightarrow y'$$

入力から出力を予測する関数を考える



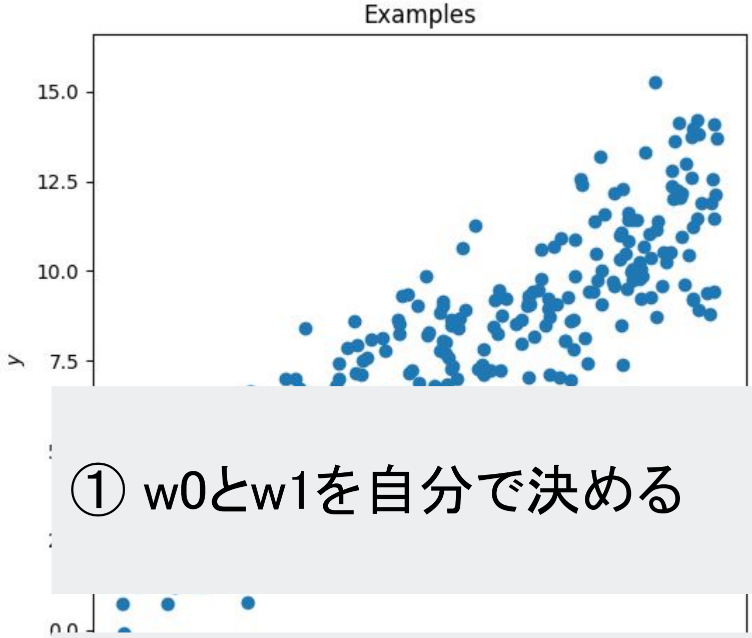
入力から出力を推測したい

$$\boldsymbol{x} \longrightarrow f_{\boldsymbol{w}}(\boldsymbol{x}) \longrightarrow y'$$

$$= \boldsymbol{w}^\top \boldsymbol{x}$$

$$= w_0 x_0 + w_1 x_1$$

切片w0と傾きw1を持つ一次関数で表す



① w0とw1を自分で決める

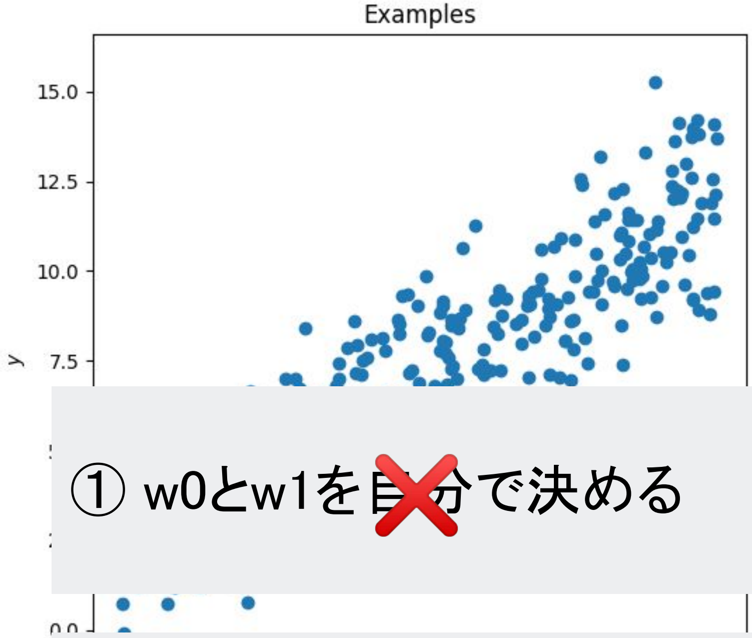
② w0とw1を発見させるプログラムを書く

入力から出力を予測したい

$$\boldsymbol{x} \longrightarrow f_{\boldsymbol{w}}(\boldsymbol{x}) \longrightarrow y'$$

$$= \boldsymbol{w}^{\top}\boldsymbol{x}$$

$$= w_0 x_0 + w_1 x_1$$



① w0とw1を自分で決める

② w0とw1を発見させるプログラムを書く

入力から出力を予測したい

$$\boldsymbol{x} \longrightarrow f_{\boldsymbol{w}}(\boldsymbol{x}) \longrightarrow y'$$

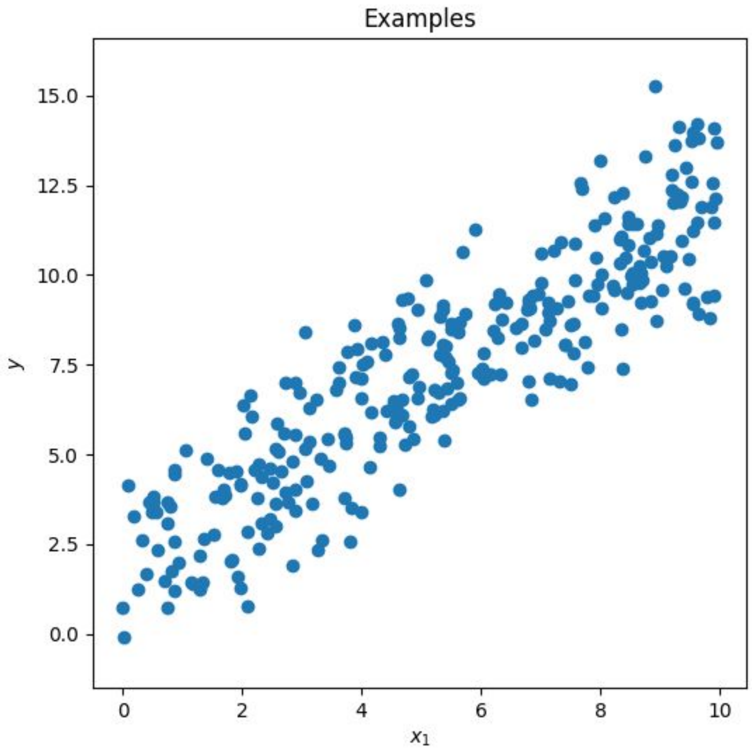
$$= \boldsymbol{w}^{\mathsf{T}}\boldsymbol{x}$$

$$= w_0 x_0 + w_1 x_1$$

“誤差関数”

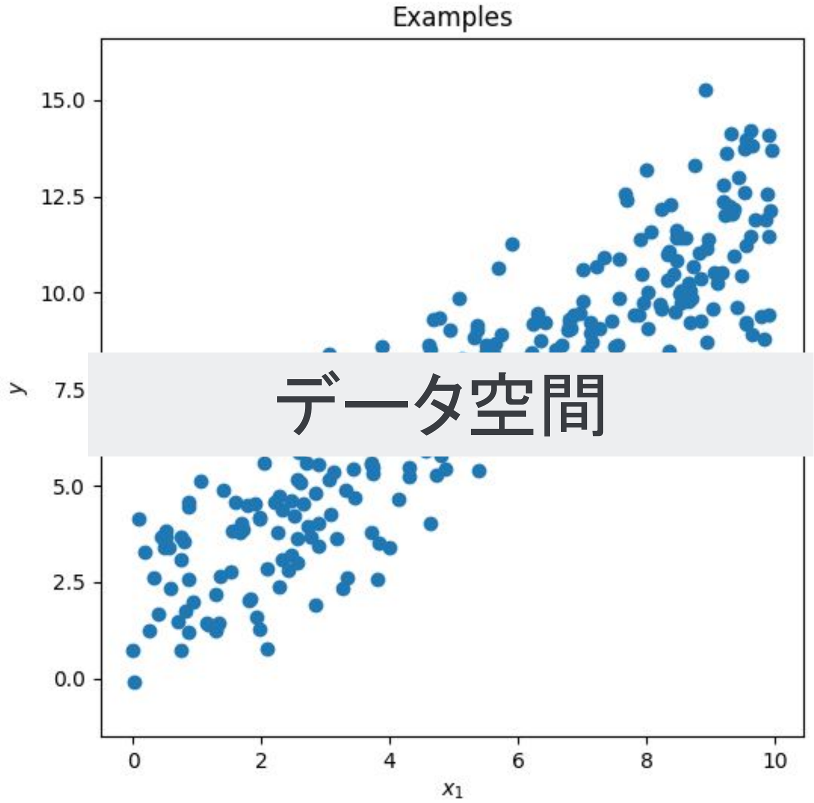
機械が学習するとは？

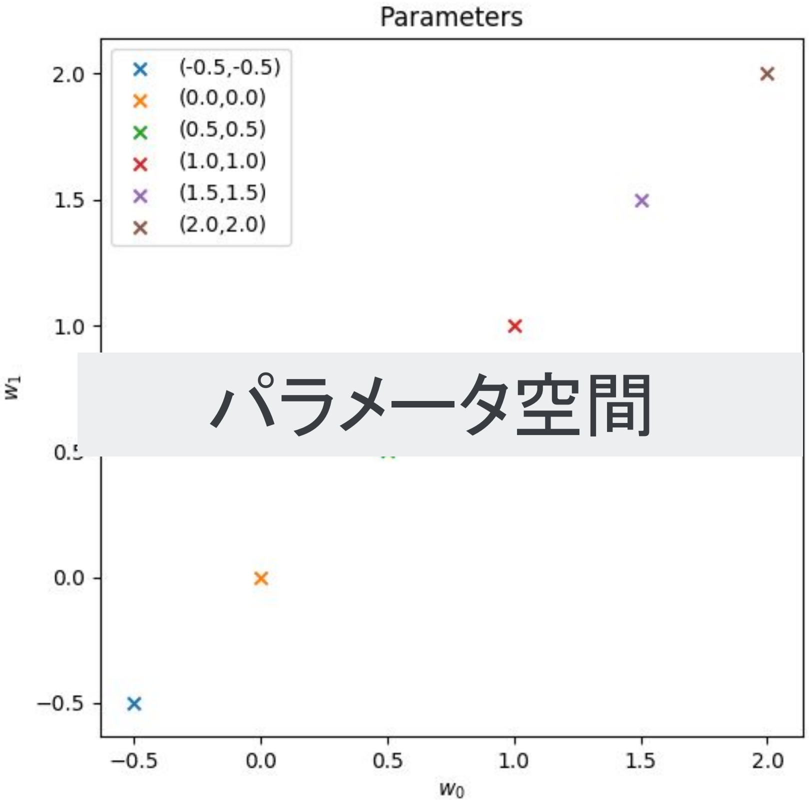
GMOペパボ



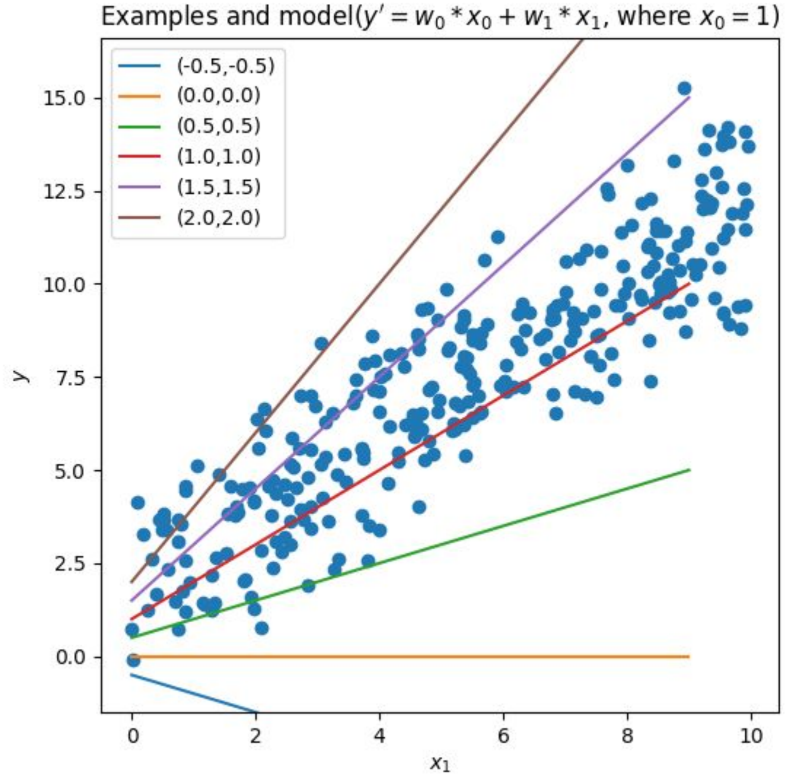
適切なw0とw1を発見したい

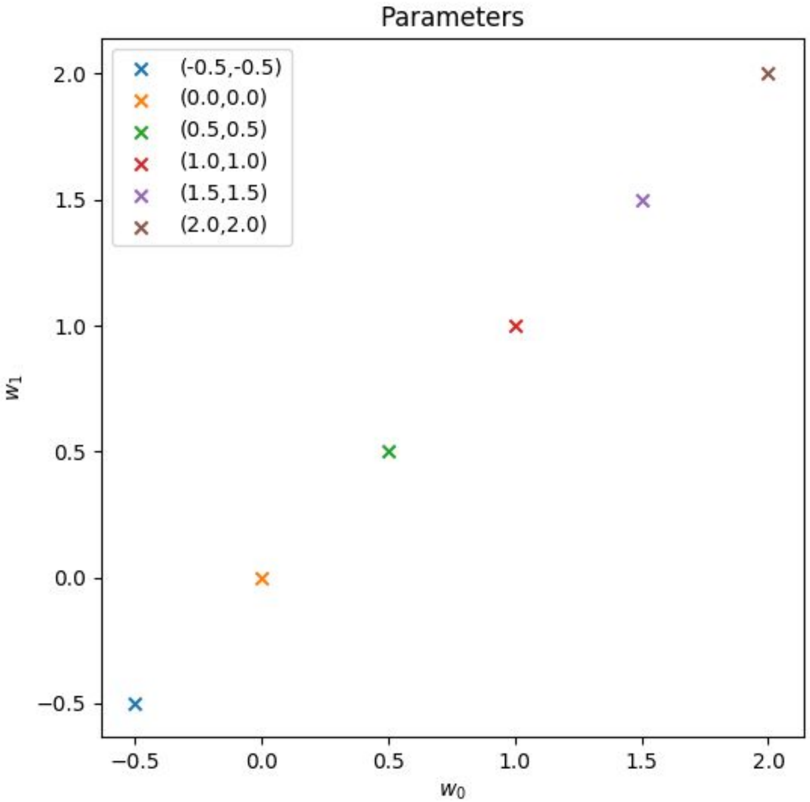
パラメータを探す


Examples
15.0
12.5
10.0
7.5
5.0
2.5
0.0
y
0
2
4
6
8
10
x_1
データ空間

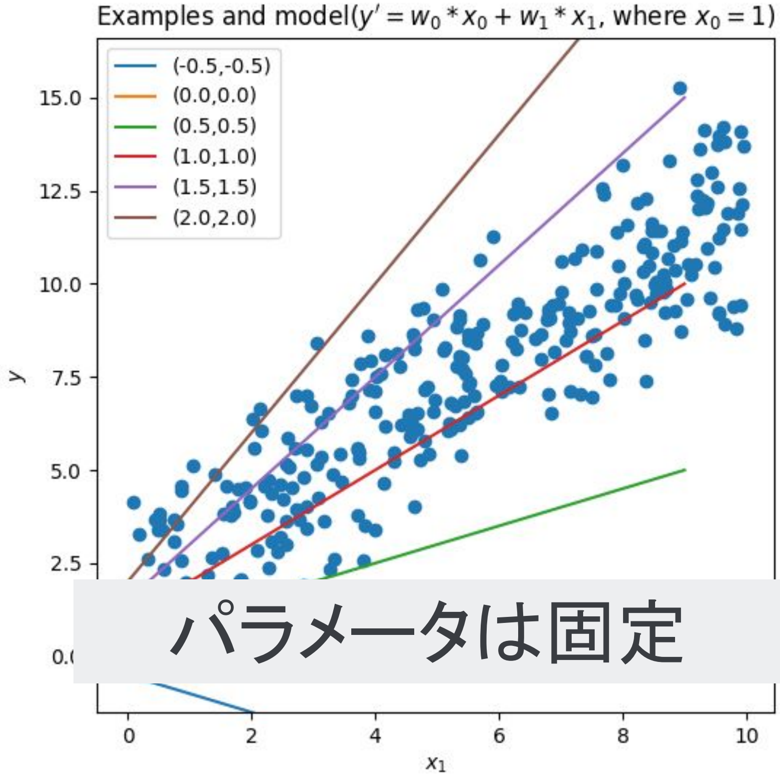

Parameters
(-0.5,-0.5)
(0.0,0.0)
(0.5,0.5)
(1.0,1.0)
(1.5,1.5)
(2.0,2.0)
2.0
1.5
1.0
0.0
−0.5
w_1
−0.5
0.0
0.5
1.0
1.5
2.0
w_0
パラメータ空間

データ空間とパラメータ空間の関係

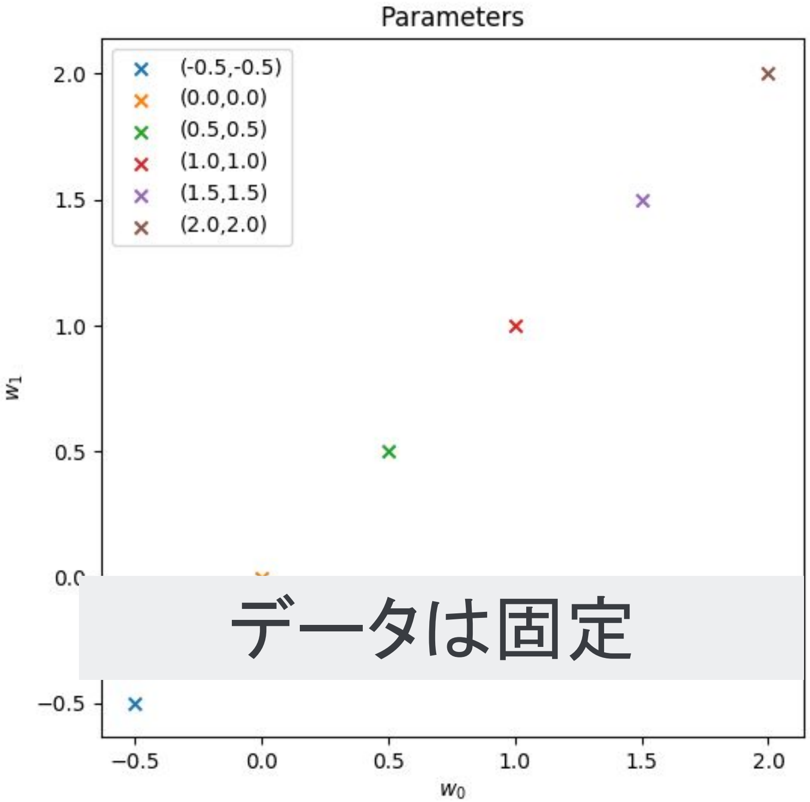




データ空間とパラメータ空間の関係



パラメータは固定

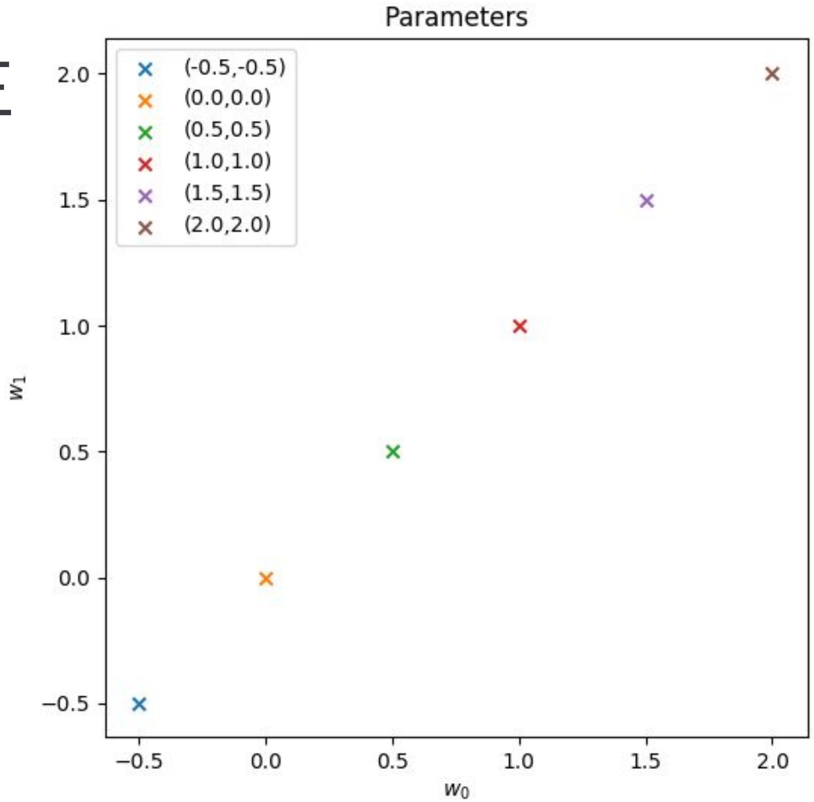


データは固定

あるパラメータを用いた時の正解と推定した値の差の合計が最も小さくなる点

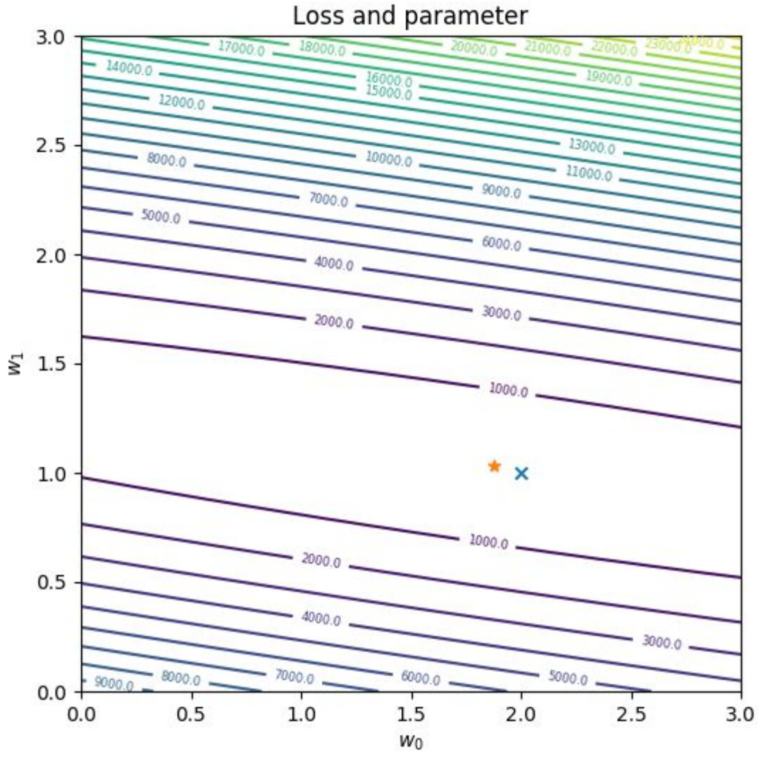
誤差関数

$$L(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - y_i')^2$$

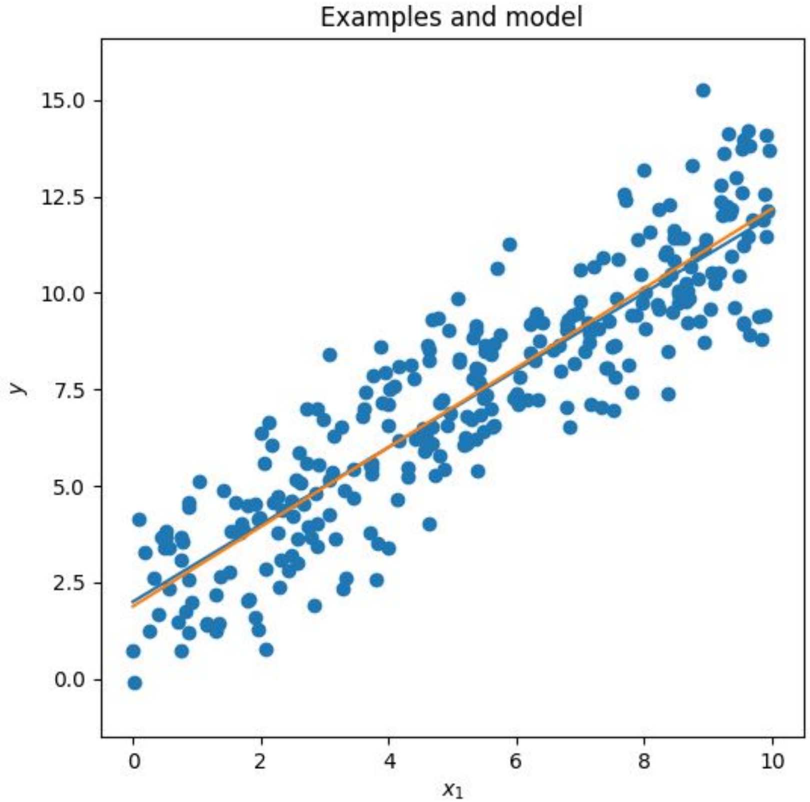


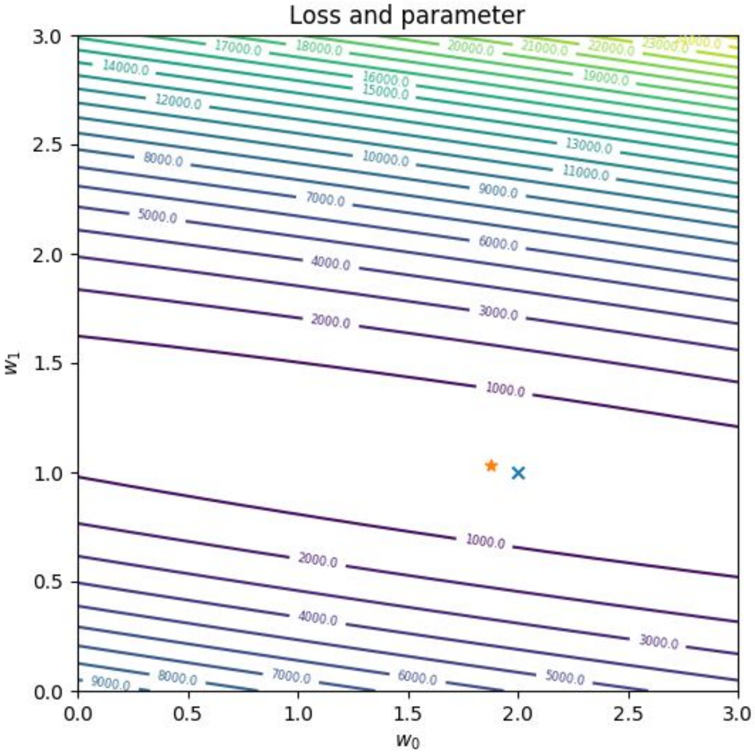
全ての組み合わせ*を試して探してみる

(*) w0とw1について0~3の範囲をそれぞれ100分割した組み合わせ



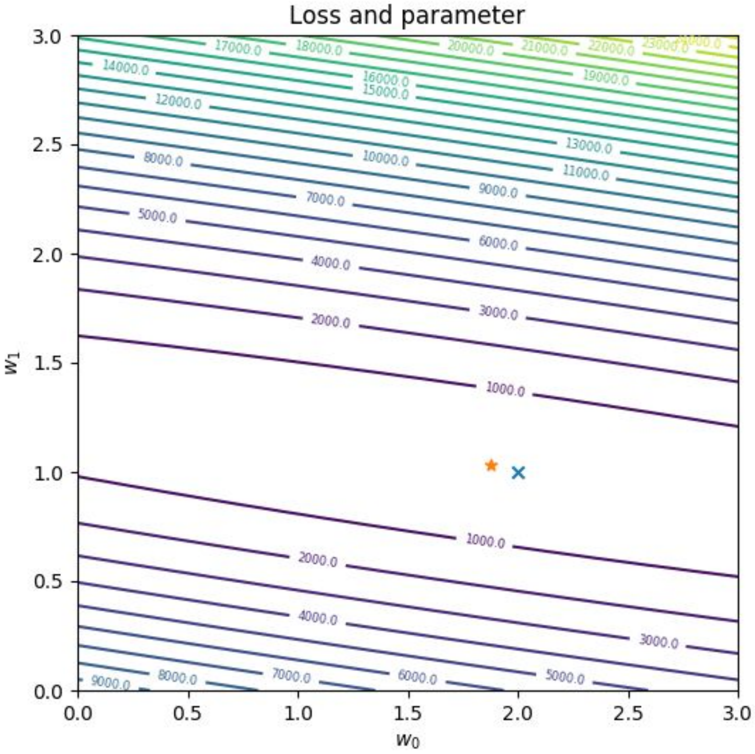
パラメータ探索における目標地点を発見する





全ての組み合わせ*を試して探してみる

より精度良く効率の良い方法はないか？
- 交点数 * データ数 * パラメータ数
- 範囲は未知
- 分割数は未知



“勾配”

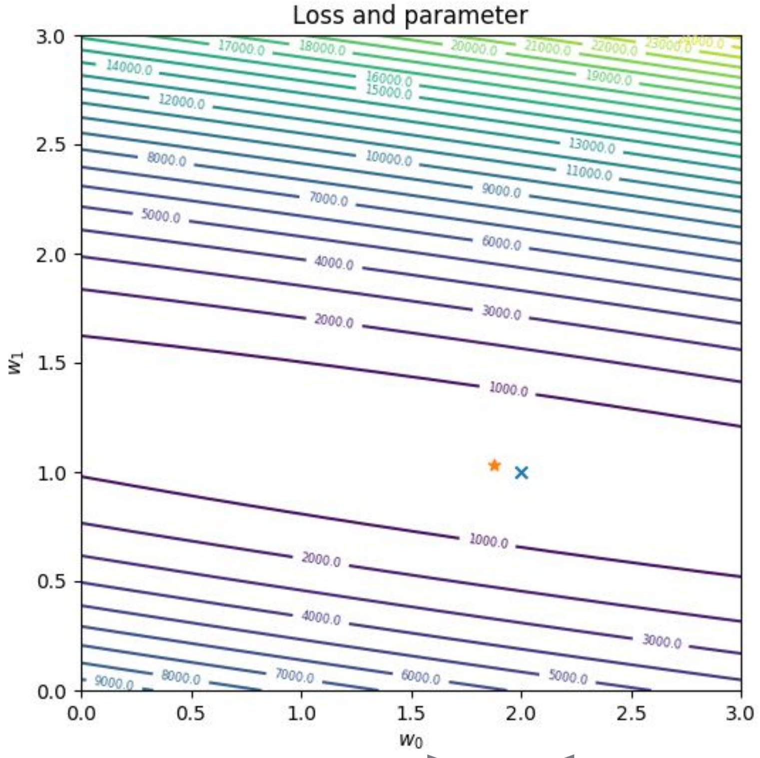
ナブラ！

GMOペパボ

誤差が小さくなる方向へパラメータを調整する

誤差関数の微分係数

点では方向はわからないが、連続するならば周辺との変化率から判断できる



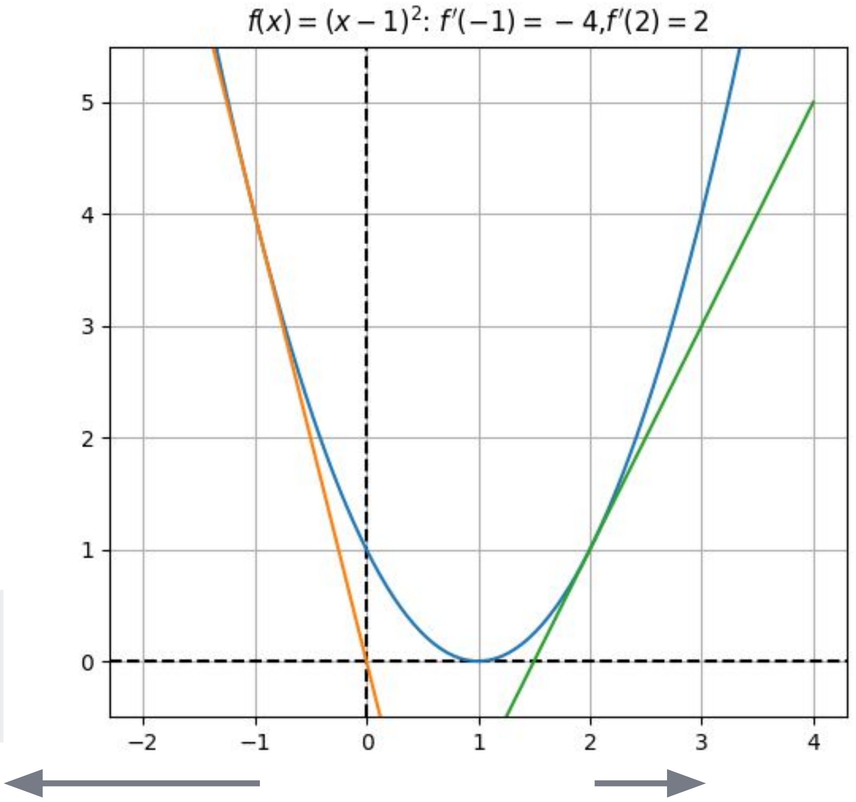
微分係数 $f'(a)$

- 関数fの点aにおける平均変化率
- 導関数から求める

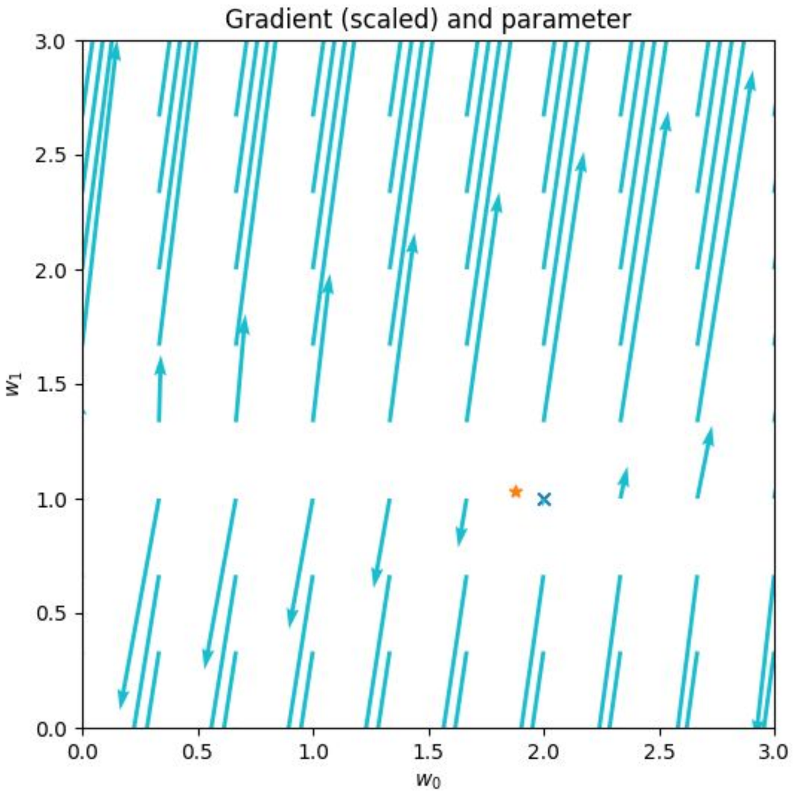
導関数 $f'(x)$

- 関数fの変数xにおける微分係数を求めるための関数
- 関数fを微分して求める

微分係数の値は、関数fの結果を大きくする方向と大きさを示す



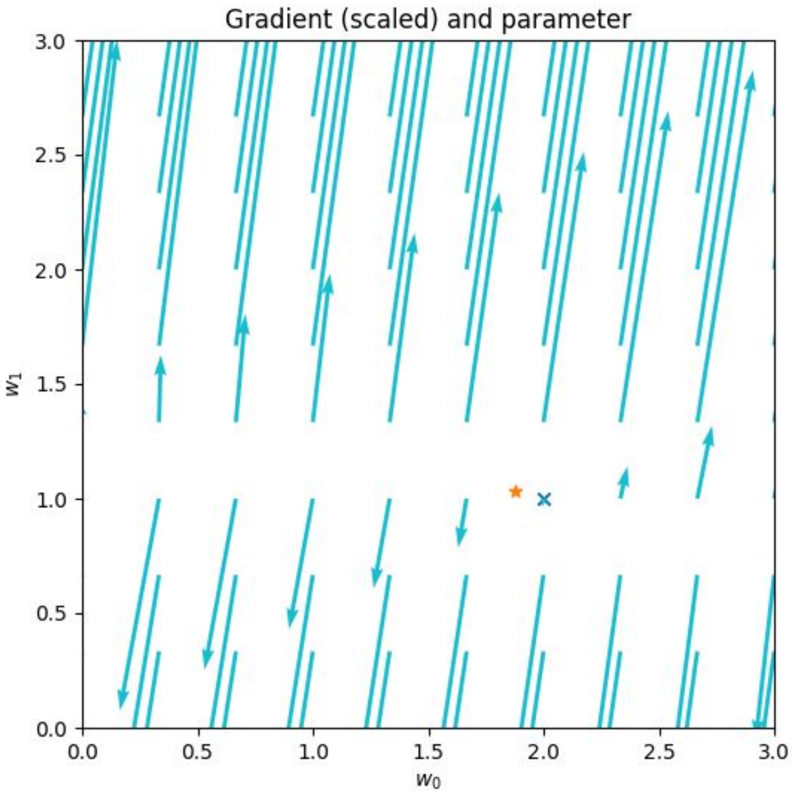
誤差関数の微分係数をいくつかの点で求めてみる



誤差関数の微分係数をいくつかの点で求めてみる

偏微分

パラメータが複数ある場合は、一つのパラメータ以外は固定して微分係数を求める。w0の軸での方向、w1の軸での方向をそれぞれ見つける。

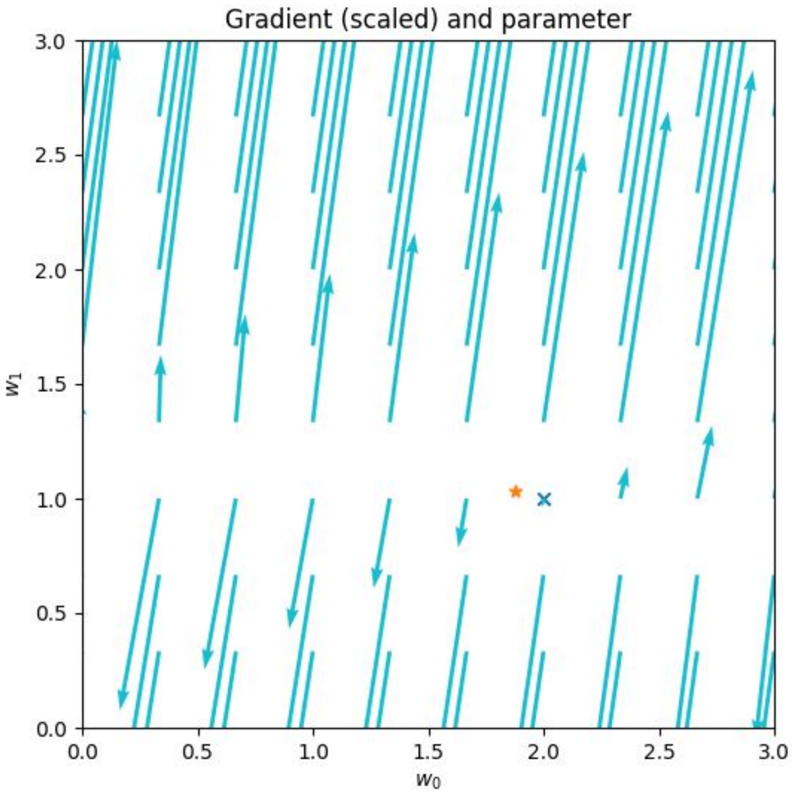


誤差関数の微分係数をいくつかの点で求めてみる

勾配

パラメータごとに偏微分した結果。

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial L(\boldsymbol{w})}{\partial w_0}, \frac{\partial L(\boldsymbol{w})}{\partial w_1} \right)$$



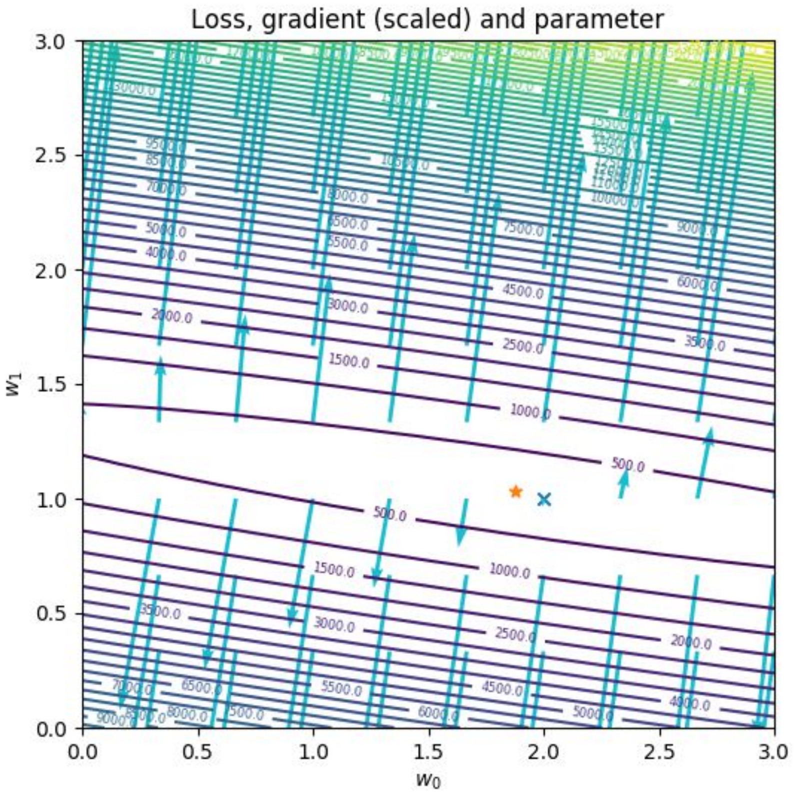
GMOペパボ

誤差関数の微分係数をいくつかの点で求めてみる

勾配

パラメータごとに偏微分した結果。

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial L(\boldsymbol{w})}{\partial w_0}, \frac{\partial L(\boldsymbol{w})}{\partial w_1}\right)$$



“勾配降下法”

勾配の逆

GMOペパボ

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial L(\boldsymbol{w})}{\partial w_0}, \frac{\partial L(\boldsymbol{w})}{\partial w_1}\right)$$

パラメータごとの偏微分

$$\frac{\partial L(\boldsymbol{w})}{\partial w_j} = \frac{\partial L(\boldsymbol{w})}{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)} \frac{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)}{\partial w_j}$$

$$= \sum_{i=1}^{n} (f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i) x_{ij}$$

合成関数の微分

$$f'(g(x)) = f'(x)g'(x)$$

f(g(x))のような合成関数の微分はそれぞれの関数の微分の積と等しい。
簡単な微分に分解することで、難しい微分も解くことができる。

GMOペパボ

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial L(\boldsymbol{w})}{\partial w_0}, \frac{\partial L(\boldsymbol{w})}{\partial w_1} \right)$$

パラメータごとの偏微分

合成関数の微分

$$\frac{\partial L(\boldsymbol{w})}{\partial w_j} = \frac{\partial L(\boldsymbol{w})}{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)} \frac{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)}{\partial w_j}$$
$$= \sum_{i=1}^{n} (f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i) x_{ij}$$

$$L(\boldsymbol{w}) = \frac{1}{2} \sum_{i=1}^{n} (y_i - y'_i)^2$$
$$= \frac{1}{2} \sum_{i=1}^{n} (y_i - f_{\boldsymbol{w}}(\boldsymbol{x}_i))^2$$
$$= \frac{1}{2} \sum_{i=1}^{n} (y_i - (w_0 x_{i0} + w_1 x_{i1}))^2$$

L(w)はfwからなり、fw(x)はwxからなる合成関数

GMOペパボ

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial L(\boldsymbol{w})}{\partial w_0}, \frac{\partial L(\boldsymbol{w})}{\partial w_1} \right)$$

パラメータごとの偏微分

$$\frac{\partial L(\boldsymbol{w})}{\partial w_j} = \frac{\partial L(\boldsymbol{w})}{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)} \frac{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)}{\partial w_j}$$

$$= \sum_{i=1}^{n} (f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i) x_{ij}$$

$$\begin{aligned} \frac{\partial L(\boldsymbol{w})}{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)} &= \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)} (y_i - f_{\boldsymbol{w}}(\boldsymbol{x}_i))^2 \\ &= \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)} ({y_i}^2 - 2 y_i f_{\boldsymbol{w}}(\boldsymbol{x}_i) + f_{\boldsymbol{w}}(\boldsymbol{x}_i)^2) \\ &= \frac{1}{2} \sum_{i=1}^{n} (0 - 2 y_i + 2 f_{\boldsymbol{w}}(\boldsymbol{x}_i)) \\ &= \sum_{i=1}^{n} (f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i) \end{aligned}$$

$$\begin{aligned} \frac{\partial f_{\boldsymbol{w}}(\boldsymbol{x}_i)}{\partial w_j} &= \frac{\partial (\boldsymbol{w}^\top \boldsymbol{x}_i)}{\partial w_j} = \frac{\partial}{\partial w_j} (w_j * x_{ij} + w_1 * x_{i1}) \\ &= x_{ij} \end{aligned}$$

GMOペパボ

誤差関数の勾配

$$\nabla L(\boldsymbol{w}) = \left(\frac{\partial L(\boldsymbol{w})}{\partial w_0}, \frac{\partial L(\boldsymbol{w})}{\partial w_1}\right)$$

$$= \left(\sum_{i=1}^{n}(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i)x_{i0}, \sum_{i=1}^{n}(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i)x_{i1}\right)$$

パラメータの更新

$$\boldsymbol{w} = \boldsymbol{w} - \alpha \nabla L(\boldsymbol{w})$$

$$= \begin{pmatrix} w_0 - \alpha \sum_{i=1}^{n}(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i)x_{i0} \\ w_1 - \alpha \sum_{i=1}^{n}(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i)x_{i1} \end{pmatrix}$$

“まとめ”

- 入力から出力を推測する関数のパラメータを求める手法を学んだ
  - 目標を誤差関数として定義した
  - 最小の誤差関数の値を効率的に求めるため勾配を理解した
  - 勾配を利用した勾配降下法によるパラメータの更新を学んだ

- 誤差や勾配は学習データを全て用いて求めるため効率が悪い
  - 最急降下法と呼ばれる
  - ランダムな100~1000程度のデータを利用するミニバッチ勾配降下法が良く利用される
  - 特に1つだけのデータを利用する場合、確率的勾配降下法と呼ばれる

“参考”

GMOペパボ

- 本資料における勾配降下法の導出は以下の文献を参考にしました。
- より詳細、発展的な説明が必要であれば、精読し、理解を深めてみてください。

- LINE Fukuoka株式会社 立石 賢吾, やさしく学ぶ 機械学習を理解するための数学のきほん ~アヤノ&ミオと一緒に学ぶ 機械学習の理論と数学、実装まで ~, マイナビ出版, 2017年09月20日. ISBN:978-4-8399-6352-1
- 三宅 悠介, Goによる勾配降下法 - 理論と実践 -, プログラマのための数学勉強会 @福岡 #5, 2016年8月

レッスン＆ハンズオン

Machine Learning Crash Course | Google Developers

- レッスン＆ハンズオンとして、Googleが公開している機械学習の短期集中コースを利用します
  - https://developers.google.com/machine-learning/crash-course
- Day4は、GeneralizationからClassificationまで行います。
  - 英語のテキストのみです
  - 各セッションのビデオは飛ばして構いません。後半のテキストにもほぼ同じ内容が含まれており冗長なためです（時間もギリギリなので）
  - 自分のペースで進めてください（休憩含む）
  - わからないところは逐次聞いてください！！！（ **重要**）

GMOペパボ

1. Generalization

15min。モデルの汎化性能について用語や判断基準などを学びます。

2.Training and Test Sets

25min。データセットの分割の必要性と利点について学びます。

3.Validation Set

35min。パラメータチューニングを伴うモデル訓練でのデータセットの分割について学びます。

4.Representation

35min。データを特徴量として表現するための方法を学びます。

5.Logistic Regression

20min。明日のハンズオンで使うロジスティック回帰について学びます。

6.Classification

32min。ロジスティック回帰などを用いた分類タスクの評価指標を学びます。ThresholdingからCheck Your understanding: Accuracy, Precision, Recallまで。

See you tomorrow!