

機械学習研修 Day1

(研修全体ガイダンス・機械学習イントロ・ハンズオン)

渡辺 龍二 / ペパボ研究所
ペパボエンジニア研修2022
2022.08.02

GMOペパボ

1. 研修ガイダンス (5 min)
2. 機械学習イントロダクション(20 min)
3. Let's ハンズオン! (140 min)
4. 学んだことを各自まとめる(10 min)

機械学習研修の目的はこれだ！

サービス価値の向上のために「再帰化」を実現する上で、実装手段として機械学習を検討できるようになる。

- 目的の背景:ペパボのエンジニアとして機械学習を学ぶ動機は、
機械学習を「再帰化」の実現アプローチとして利用することにある
 - サービス価値の向上において「再帰化」は重要な考え方^[栗林2022]
 - 「再帰化 ≠ 機械学習の利用」であることを理解して
再帰化を実現するために機械学習の利用するか否かを判断できることを目指す

研修目的を達成するために以下のカリキュラムを設定しています

研修の目的

サービス価値の向上のために「**再帰化**」を実現する上で、**実装手段として機械学習を検討できる**ようになる。

	ゴール	実施内容
Day1	機械学習とは何か、 自分の言葉で説明できる	<ul style="list-style-type: none"> - 機械学習イントロダクション(座学) - Machine Learning Crash Course (ハンズオン)
Day2	再帰化を実現するために、 機械学習をどう活用できるか イメージできる	<ul style="list-style-type: none"> - Machine Learning Crash Course (ハンズオン) - 再帰化と機械学習(座学)
Day3	サービスの実課題を 機械学習で解決する 一連の流れの例を挙げられる	<ul style="list-style-type: none"> - Machine Learning Crash Course (ハンズオン) - 機械学習のサービス連携事例紹介(座学)

機械学習イントロダクション

1. 研修ガイダンス
- 2. 機械学習イントロダクション(20 min)**
 - a. ざっくりとした「機械学習とは何か」
 - b. 予測タスクからイメージする機械学習
 - c. 機械学習で取り組むタスクいろいろ
3. Let's ハンズオン！(140 min)
4. 学んだことを各自まとめる(10 min)

機械学習とは…

アルゴリズムによって

データに内在する規則や構造を抽出することで

タスクを解くための技術や方法論のこと

「機械学習」というネーミングのニュアンスは以下のような感じです

※「機械」が「学習」する、というのはあくまでメタファー

アルゴリズムによって

② コンピュータで実行可能な手続きを適用し…「機械」(コンピュータ)が実行する

データに内在する規則や構造を抽出することで

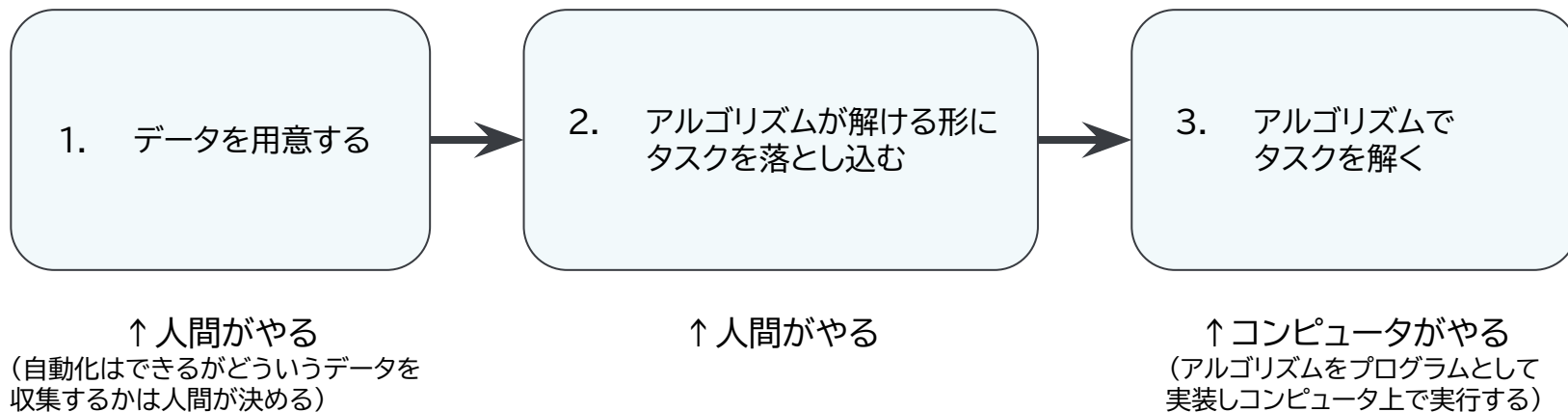
① 先にデータがあり ③有用な規則や構造を得る…なんか人間の「学習」っぽい

タスクを解くための技術や方法論のこと

④何らかのタスクを解く

アルゴリズムでタスクを解くために、
人間によるデータの用意やタスクの落とし込みが必要となる

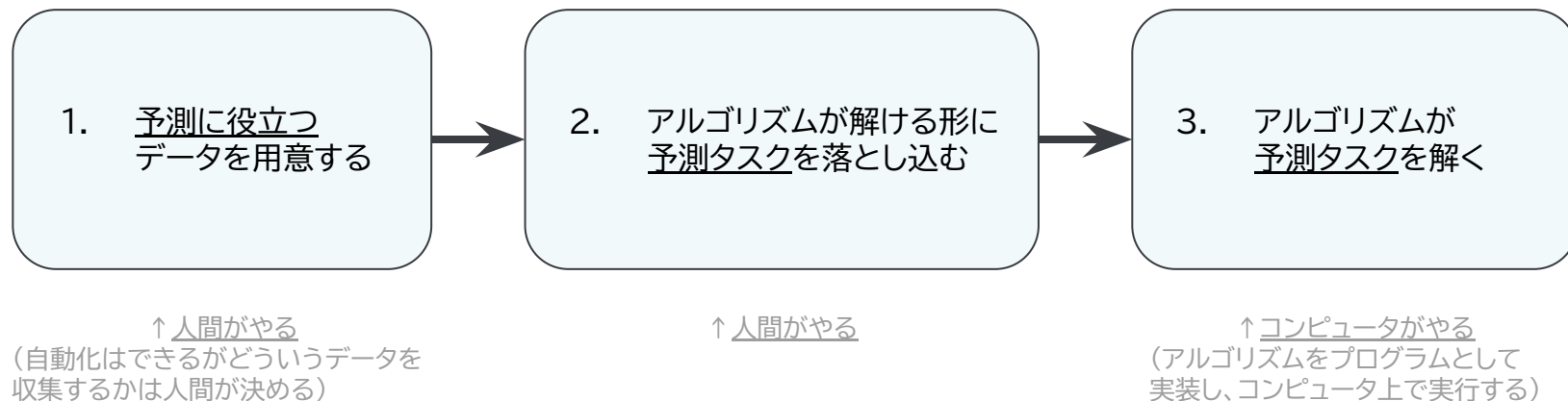
何か解きたいタスクがあるとして、機械学習では以下の手順を踏む



※意外と人間がやることは多いです(機械学習に全部お任せ！ではない)

これから「予測タスク」を題材に、以下の流れの具体例を見ていきます！

何か解きたいタスクがあるとして…



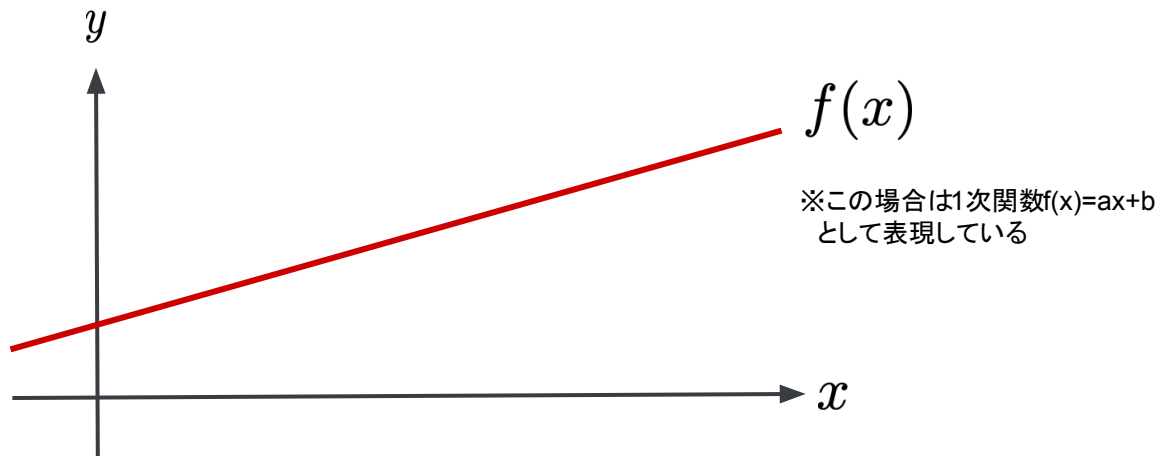
※意外と人間がやることは多いです(機械学習に全部お任せ！ではない)

その日の気温からアイスクリームへの支出額を予測したい、というタスクを想定する

- 想定:あなたは新しく開店するアイスクリーム屋の店長です
- 願望:毎日の在庫管理を過不足なく行いたい
- 着想:
 - 「気温が高いとアイスクリームの売れ行きが上がる」という関係がありそう
 - 気温からアイスへの支出額の予想ができれば、在庫管理に有用なのでは？

気温と支出額の間に関数の関係があると仮定し、関数を得て支出額を予測する

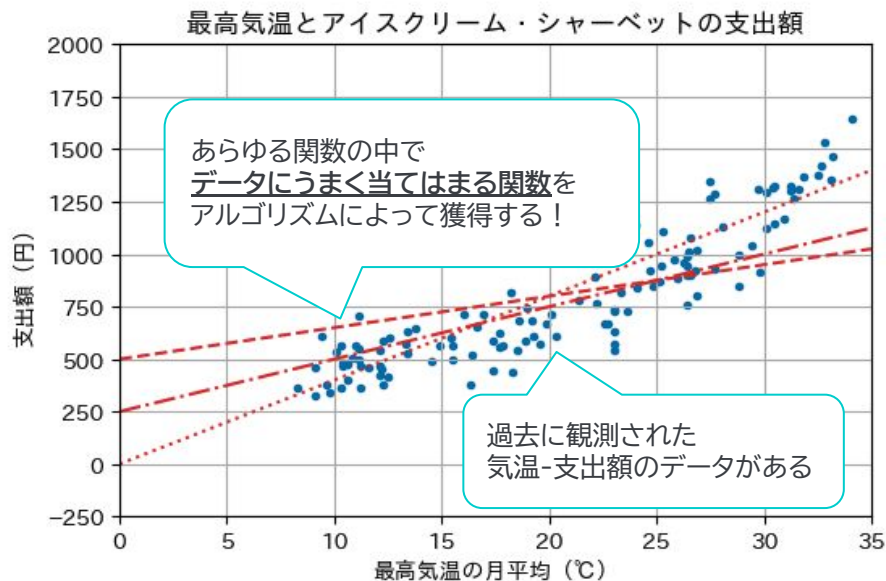
- 気温を x , 支出額を y とし $y=f(x)$ という関係があるとする



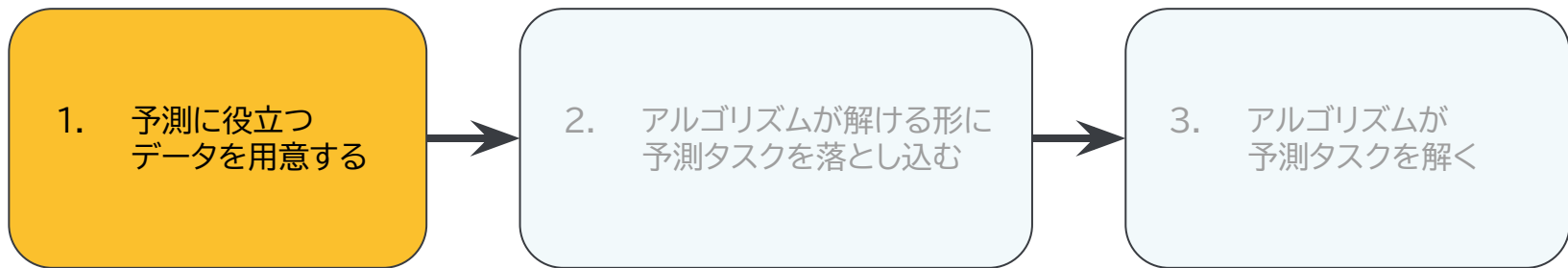
- この f を獲得できれば、任意の x に対して y を予測できる
- よって、予測タスク≡関数獲得タスクと言える

機械学習では、データにうまく当てはまるような関数を獲得することを考える

機械学習の定義の「データに内在する規則を抽出」に相当

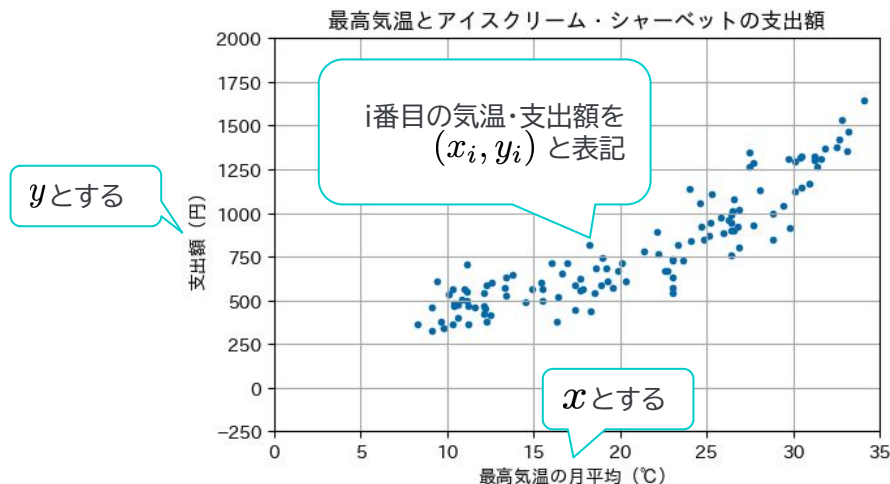


「データに当てはまる関数」を得るために、まずはデータの用意から行う



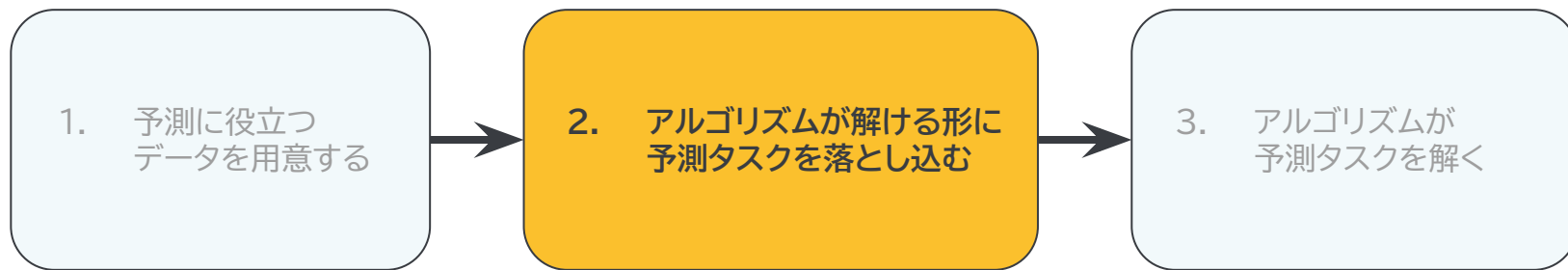
データを用意する

- 今回の具体例におけるデータ: $D = \{(x_i, y_i)\}_{i=1}^N$ と表記



- 一般に、どんなデータを使うかは人間が決定し、計算機で収集・蓄積する

次は、アルゴリズムが解ける形に予測タスクを落とし込んでいく

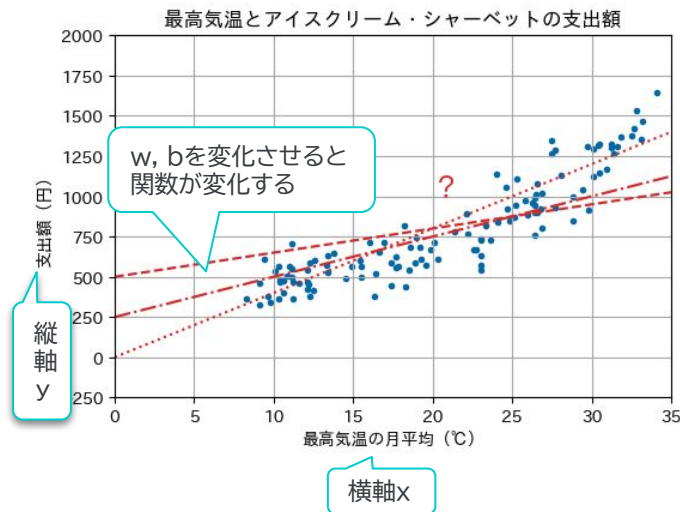


関数をより具体的に数式で表現する(モデル化)

- 気温 x と消費額 y の関係(モデル)を以下のように定義する(この手順をモデル化と呼ぶ)

$$y = f_{w,b}(x) \\ = wx + b$$

関数は傾き w , 切片 b の
一次関数であると仮定し
それぞれフリーパラメータとする



- このモデルにおいてタスク「予測＝関数獲得」は「パラメータ w, b を求める」になる
 - 推定対象を具体的に定式化することで、アルゴリズムで解きやすくなる

データに対するパラメータの評価指標(損失関数)を定義する

- やりたいこと: アルゴリズムが良いパラメータを見つけるために
パラメータによって定まる関数がデータをどれだけ当てはまっているかを定量的に評価したい
- やり方: データと関数の予測値のズレ $\epsilon_i = y_i - f_{w,b}(x_i)$ (残差) を元に以下の関数を定義する

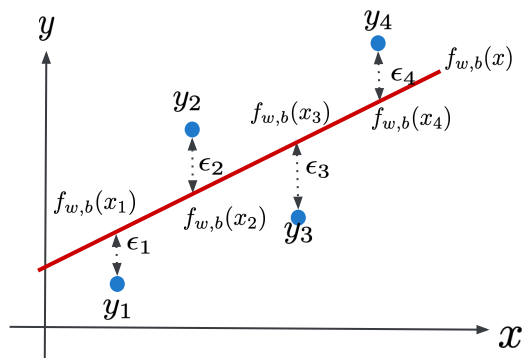
$$F(w, b | D) = \frac{1}{2} \sum_{i=1}^N \epsilon_i^2$$

w, b の変化に伴い
ズレも変動するため
w, b の関数となる

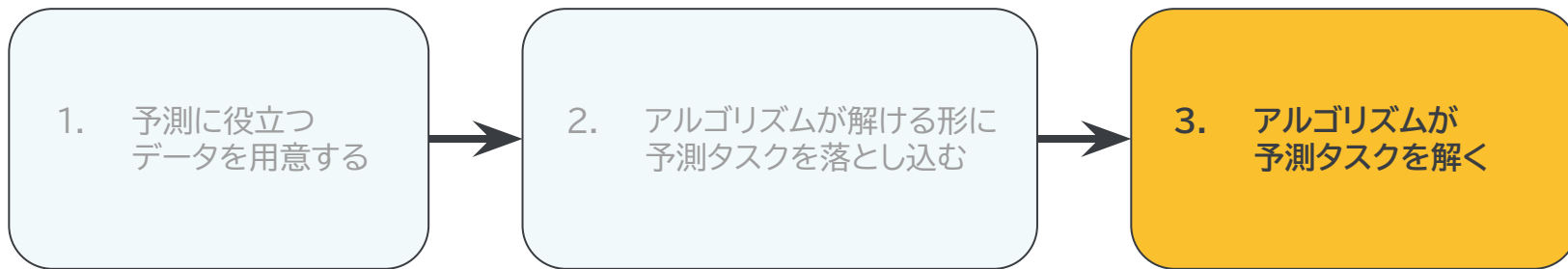
2乗を取って
正の値にして
総和を取る

このような関数を
損失関数 (loss function) と呼ぶ

ズレ(残差)の図示



準備ができたので、アルゴリズムにタスクを解いてもらう



評価指標(損失関数)を最適化するパラメータをアルゴリズムによって求める

- 定義した損失関数 $F(w, b | D)$ が最小となるようなパラメータを求める
 - いわゆる「数理最適化問題(厳密には連続最適化問題)」であり、アルゴリズムで解ける

$$\hat{w}, \hat{b} = \arg \min_{w, b} F(w, b | D)$$

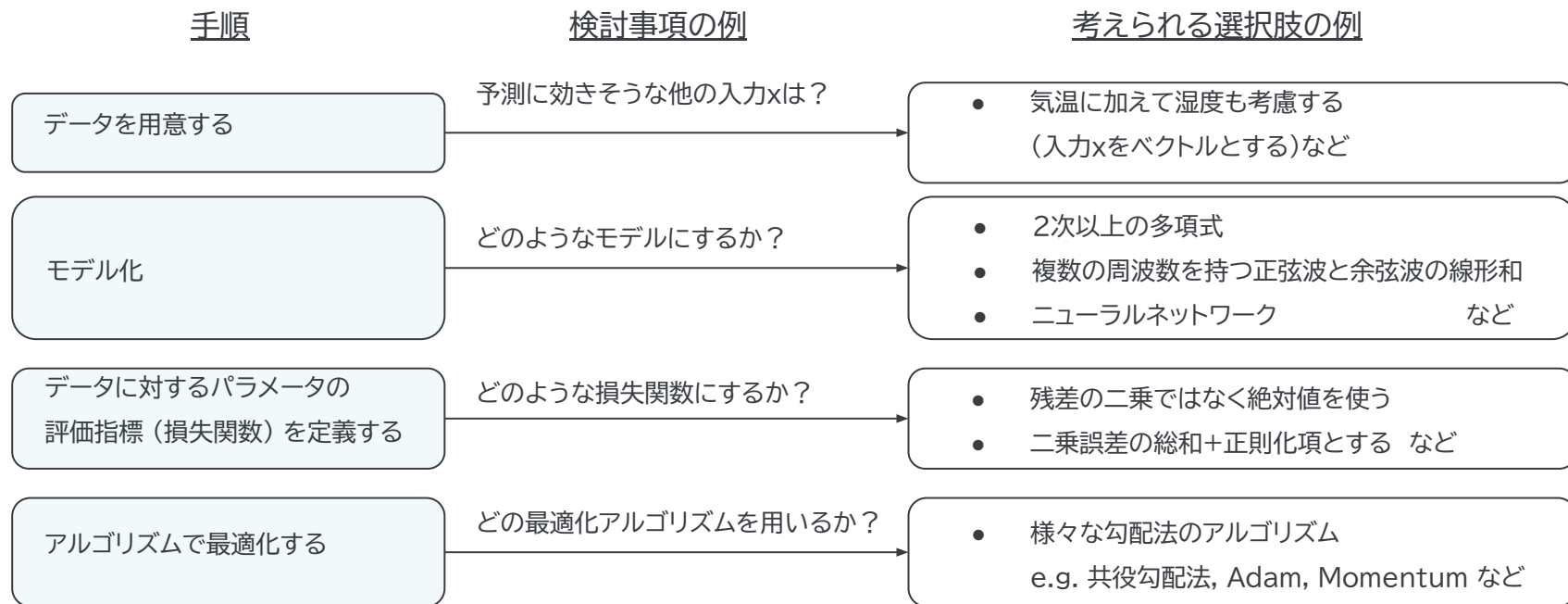
最終的な推定値を
^付きで表現

ただし
$$F(w, b | D) = \frac{1}{2} \sum_{i=1}^N (y_i - f_{w, b}(x_i))^2$$
$$f_{w, b}(x) = wx + b$$

- どんなアルゴリズムを用いるかは人間が決める
 - 関数の性質や与えられる制約によってある程度は絞られる
 - 機械学習では勾配法が用いられることが多い

詳しくはハンズオンで学びます

各手順において複数の選択肢があるため、良い予測のためには試行錯誤が必要



機械学習で解くタスクはさまざまある(関数獲得ばかりではない)

		出力yの取る値	
		連続値	離散値
データ中に教師yがあるか？	ある	回帰 (Regression)	分類 (Classification)
	ない	次元削減 (Dimensionality Reduction)	クラスタリング (Clustering)

どちらも予測＝関数獲得タスク
まとめて「教師あり学習」と呼ぶ
(前述のアイスクリームの例は回帰)

その他のタスク

- ・ 強化学習
- ・ 協調フィルタリング、異常検出などなどなど…

1. 研修ガイダンス
2. 機械学習イントロダクション(20 min)
 - a. ざっくりとした「機械学習とは何か」
 - b. 予測タスクからイメージする機械学習
 - c. 機械学習で取り組むタスクいろいろ
3. **Let's ハンズオン！(140 min)**
4. 学んだことを各自まとめる(10 min)

手を動かしながら機械学習を学んでいきましょう！

- [Machine Learning Crash Course](#)
 - Software Engineerが
機械学習の理論・実装を
学ぶためのハンズオン
- 今日進めて欲しい単元
 - ML concepts - Introduction to ML (3 min)
 - Framing (15 min)
 - Descending into ML (20 min)
 - Reducing loss (60 min)
 - First steps with tf (65 min)
 - 進められるところまで！

Machine Learning Crash Course with TensorFlow APIs

Google's fast-paced, practical introduction to machine learning, featuring a series of lessons with video lectures, real-world case studies, and hands-on practice exercises.

[Start Crash Course](#)[View prerequisites](#)

30+ exercises



25 lessons



15 hours



Lectures from Google
researchers



Real-world case studies



Interactive visualizations

進める上での注意点

- ハンズオン資料は英語です
- 動画は飛ばしてください！ ※”Introduction to ML” は除く
- 休憩を適宜取ってください
- 早く終わった場合はどんどん先に進んで大丈夫です！
- Day1ゴール「機械学習とは何か、自分の言葉で説明できる」が達成できるようにがんばりましょう！

1. 研修ガイダンス
2. 機械学習イントロダクション(20 min)
 - a. ざっくりとした「機械学習とは何か」
 - b. 予測タスクからイメージする機械学習
 - c. 機械学習で取り組むタスクいろいろ
3. Let's ハンズオン！(140 min)
4. **学んだことを各自まとめる(10 min)**

学んだことをまとめる時間

- 以下のことをまとめてみてください～！
 - 学んだこと
 - 感じたこと
 - 考えたこと
- 「機械学習を自分の言葉で説明」にトライしてみてください！