

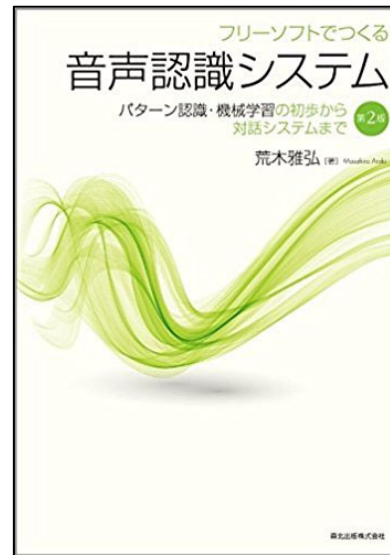
# 機械学習講座 概要版

本日のスライド

<https://github.com/MasahiroAraki/MLCourse>

# 自己紹介

- 荒木雅弘
- 京都工芸繊維大学 情報工学・人間科学系 准教授
- 専門：音声対話処理
- 著書



# この講座の目的

- 機械学習技術の全体像を広く・浅く知ること、  
自社製品・サービスが機械学習技術によってどのように変革できるか、またそのためにどのような技術者・開発者を育成するべきかが見通せるようになることを目的とします。

# テキスト



2018年8月発行

オーム社

韓国語版、中国語版

(簡体字・繁体字)

発売予定

# 本日の予定

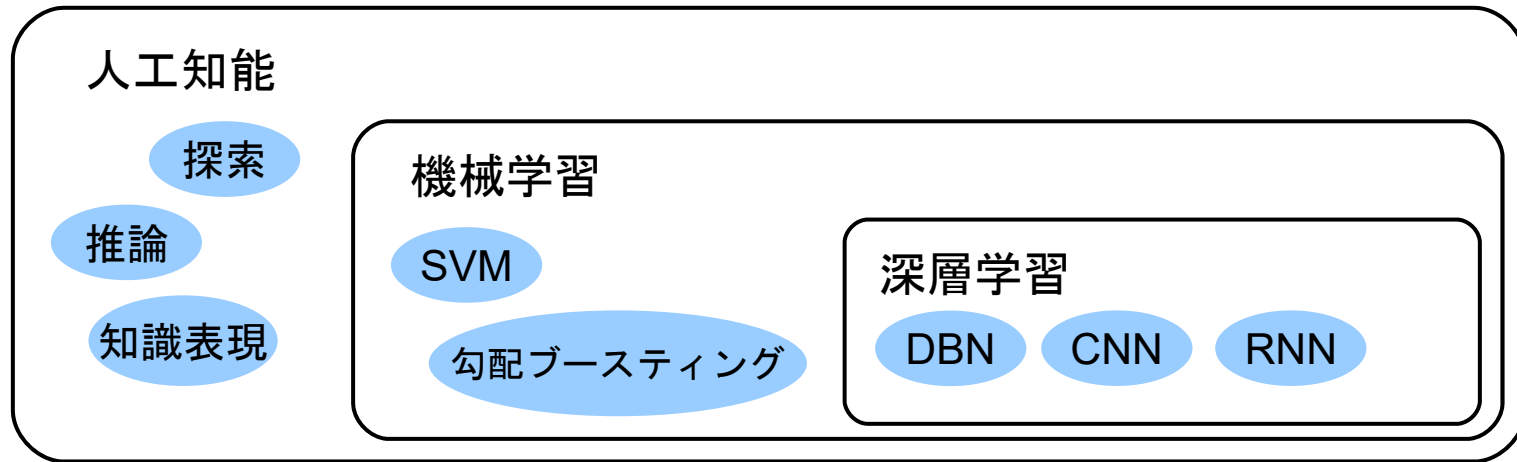
- 9:30-10:30 機械学習の概要、回帰
- 10:45-11:45 基礎的な識別、学習結果の評価
- 13:00-14:00 深層学習
- 14:15-15:15 高度な識別、教師なし学習
- 15:30-16:30 機械学習エンジニア育成のために

# 序章、1章のストーリー

- 清原は市役所に勤務する新社会人
- 清原は、市が主催するイベントの参加人数をできるだけ正確に予測したくて、研究室の先輩さやかのもとを訪れる
- さやかは清原に機械学習全般についてと、データから数値を予測する回帰手法について教える

# 機械学習の概要（序章）

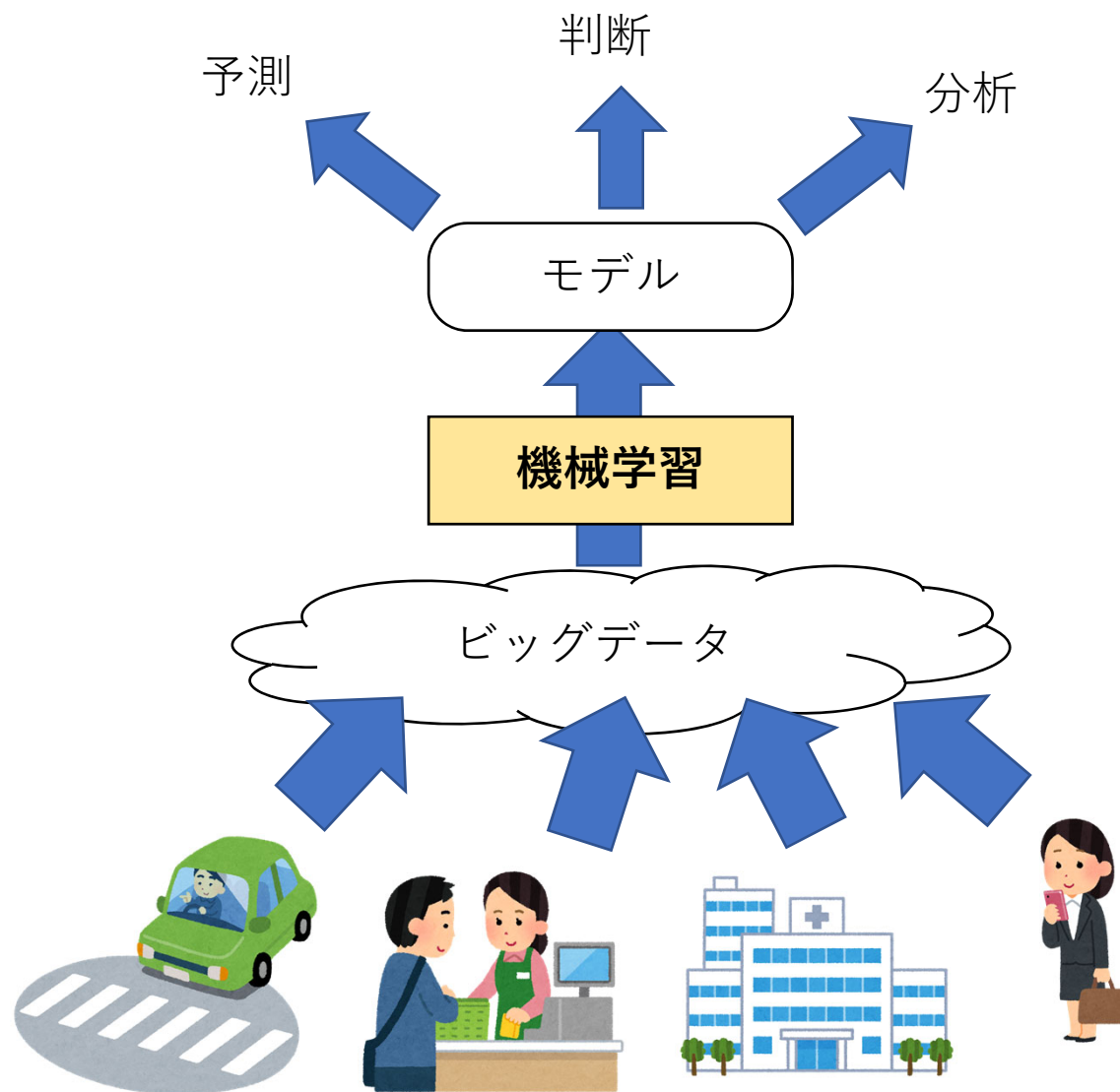
# 人工知能・機械学習・深層学習



- 人工知能
  - ある種の 現在、人が行っている知的な判断を代わりに行う技術
- 機械学習
  - 大量のデータから予測や判断を行うモデルを作成する技術
  - 深層学習はその一手段



# 機械学習の位置づけ



# 機械学習の分類



3章 評価

1章 回帰

2章 ロジスティック識別、決定木

4章 ディープラーニング

5章 アンサンブル学習

機械学習

6章 クラスタリング、行列分解

教師あり学習

中間的学習

教師なし学習

正解付きデータ

年齢	性別	時刻	購入
35	男	16	Yes
24	男	9	Yes
22	女	21	No



正解なしデータ

長さ	幅	重さ
15	6	16
24	8	19
32	7	18

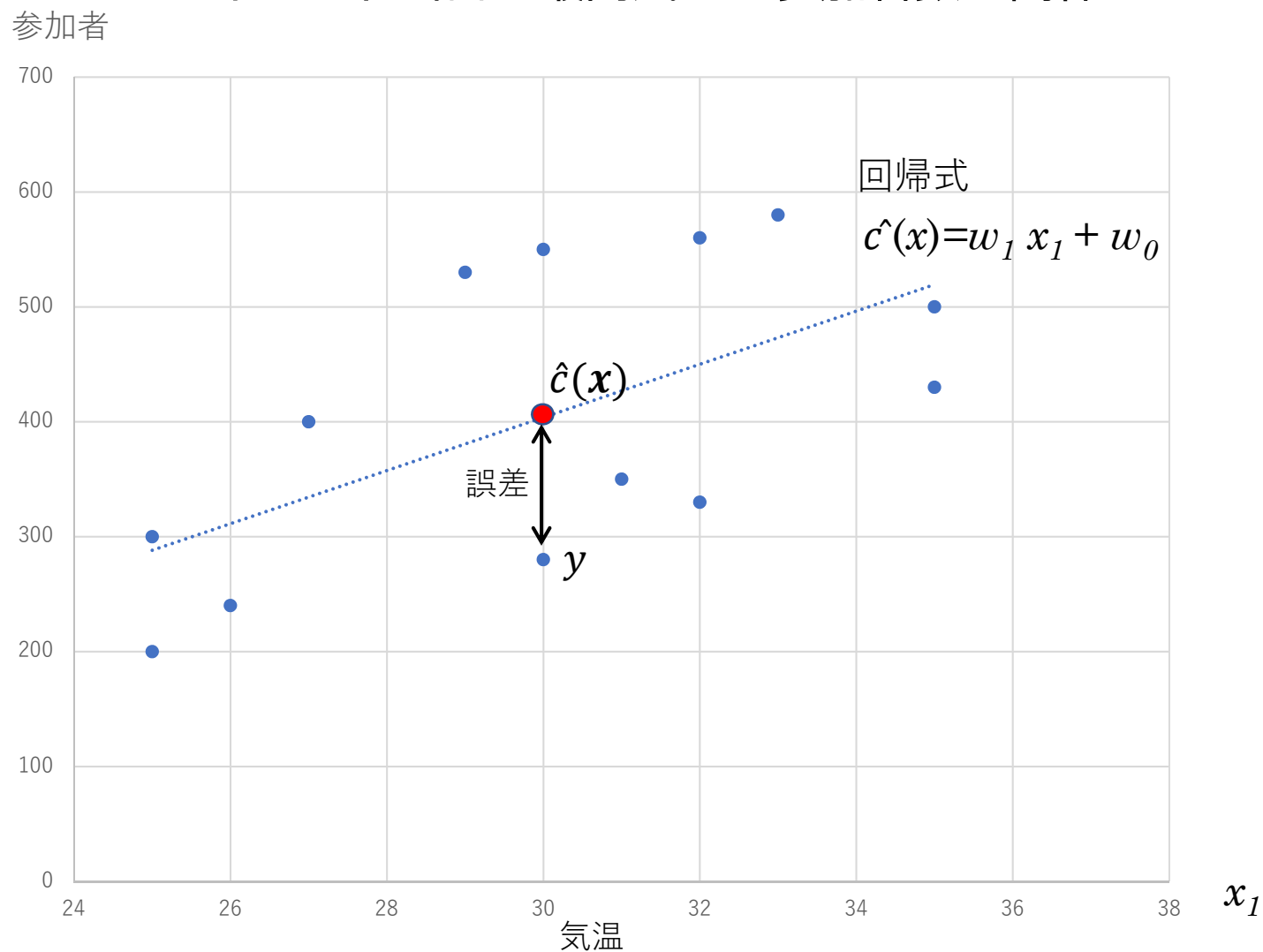
# 回帰（1章）

- 回帰とは
  - 教師あり学習問題
  - 特徴から数値を予測する

p.11 3コマ目

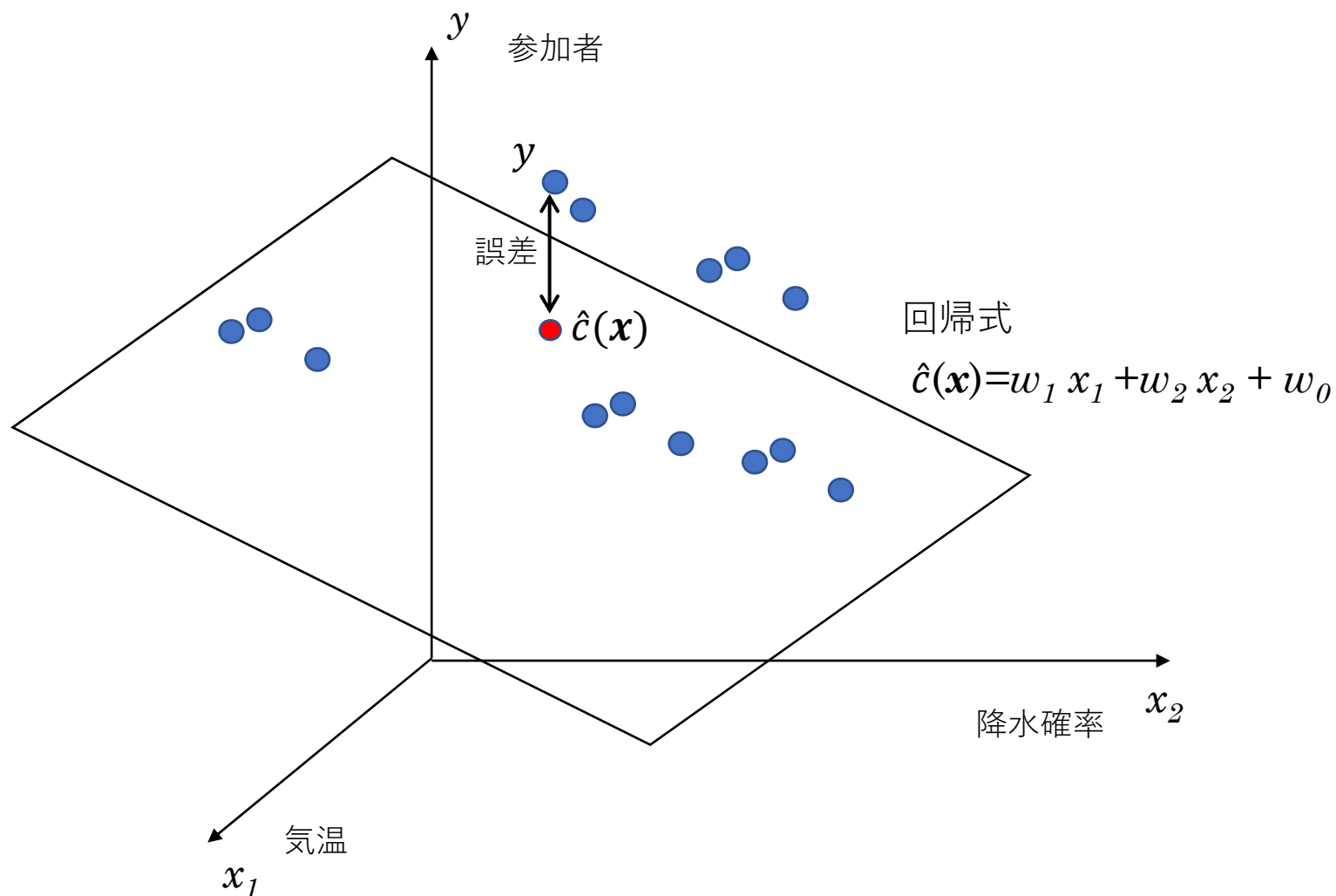
# 単純な回帰問題

イベント当日の最高気温と参加者数の関係



# やや複雑な回帰問題

イベント当日の最高気温・降水確率と参加者数の関係



# 線形回帰

- 問題の定義

直線・平面

- 入力  $\mathbf{x}$  から出力  $\hat{c}(\mathbf{x})$  を求める回帰式を1次式に限定

$$\hat{c}(\mathbf{x}) = \sum_{i=0}^d w_i x_i$$

$d$  : 特徴の次元数  
 $x_0$  : 0に固定

- 学習データに対してなるべく誤差の少ない直線（あるいは平面）の係数  $w$  を求める

# 最小二乗法による解法

- 推定の基準：誤差の二乗和  $E$  を最小化

$$E(\boldsymbol{w}) = \sum_{i=1}^N (y_i - \hat{c}(\boldsymbol{x}_i))^2$$

$N$ : 全データ数

$y_i$ : 正解

$$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})$$

$\boldsymbol{X}$ : 学習データを並べた行列

$\boldsymbol{y}$ : 正解を並べたベクトル

$\boldsymbol{w}$ : 係数を並べたベクトル

- $E$ が最小となるのは  $\boldsymbol{w}$  で偏微分したものが0となるとき

$$\boldsymbol{X}^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) = 0$$

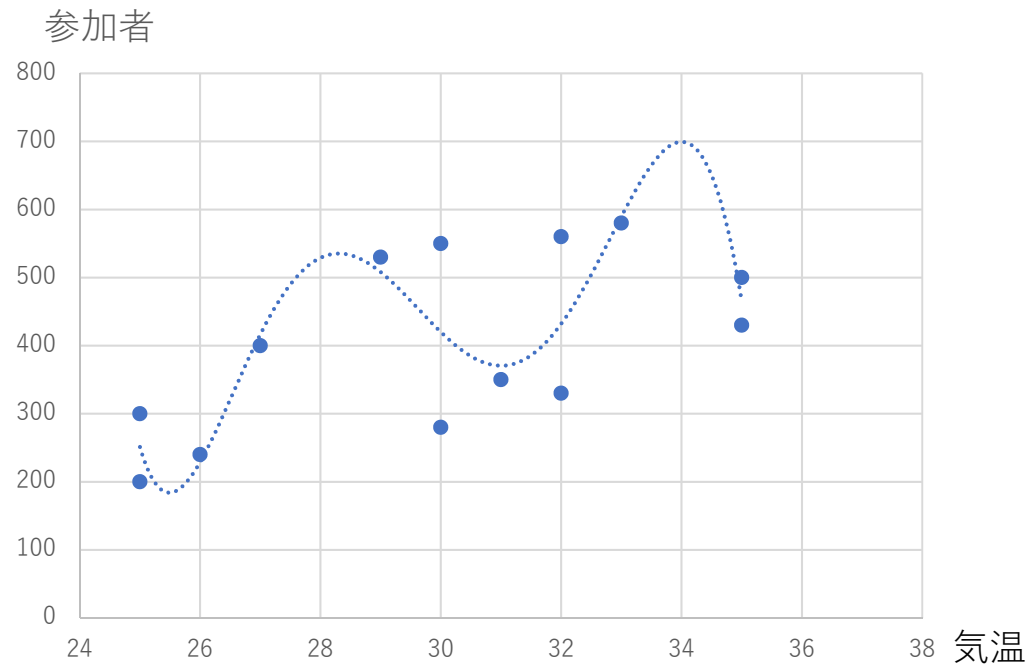
$$\Leftrightarrow \boldsymbol{w} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

$\boldsymbol{w}$  が行列の計算のみで求まる

# 正則化

- 過学習

- 最小二乗法は係数が線形であれば高次式でも適用可
- 特徴の次数を上げたり、特徴の次元数を増やしたりすると、複雑な回帰式で解を近似することになる



学習データだけに  
当てはまる不自然な  
回帰式が求まって  
しまう



# 正則化

- 過学習への対処

- 過学習した回帰式とは

- ⇒ 入力が少し動いただけで出力が大きく動く

- ⇒ 回帰式の係数  $w$  が大きい

- 正則化

多少誤差が大きくなることと  
引き換えに  $w$  を小さくする

p.23 3コマ目

# Ridge回帰

- 係数 $w$ の2乗を正則化項として誤差の式に加える
  - 全体的に係数が小さくなり、極端な値の変動がなくなる

正解に合わせて  
こちらを小さく  
しようとすると...

係数が  
大きくなる



$$E(w) = (y - Xw)^T (y - Xw) + \alpha w^T w$$



$\alpha$  : 誤差と正則化項の  
バランス

係数の値を  
小さくしすぎると...

正解から大きく  
離れてしまう

# Lasso回帰

- 係数  $w$  の絶対値を正則化項として誤差の式に加える
  - 値が0となる係数が多くなり、出力に影響を与えている特徴を絞り込むことができる

$$E(\boldsymbol{w}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) + \alpha \sum_{j=1}^d |w_j|$$

# 回帰式的具体例

- Bostonデータ

- 犯罪発生率、部屋数、立地など13の条件から不動産価格を推定

線形回帰	Ridge	Lasso
CRIM : -0.11	CRIM : -0.10	CRIM : -0.02
ZN : 0.05	ZN : 0.05	ZN : 0.04
INDUS : 0.02	INDUS : -0.04	INDUS : -0.00
CHAS : 2.69	CHAS : 1.95	CHAS : 0.00
NOX : -17.80	NOX : -2.37	NOX : -0.00
RM : 3.80	RM : 3.70	RM : 0.00
AGE : 0.00	AGE : -0.01	AGE : 0.04
DIS : -1.48	DIS : -1.25	DIS : -0.07
RAD : 0.31	RAD : 0.28	RAD : 0.17
TAX : -0.01	TAX : -0.01	TAX : -0.01
PTRATIO : -0.95	PTRATIO : -0.80	PTRATIO : -0.56
B : 0.01	B : 0.01	B : 0.01
LSTAT : -0.53	LSTAT : -0.56	LSTAT : -0.82

# 回帰の実用化事例

- NEC

- 日配品（主に冷蔵が必要なもの）の需要予測

<https://jpn.nec.com/ai/solution/value.html>

- 健診結果予測シミュレーション

<https://wisdom.nec.com/ja/technology/2018031501/index.html>

- 心疾患リスクスコアの推定 [Ganz et al. 16]

- 血液中の1130種類のタンパク質から心疾患に関連する9種のタンパク質を特定

prognostic index

$$\begin{aligned} &= 16.61 - 1.55 \times \text{ANGPT2} + 1.22 \times \text{GDF8/11} - 2.12 \times \text{C} \\ &7 + 2.64 \times \text{SERPINF2} - 0.57 \times \text{CCL18} - 1.02 \times \text{ANGPTL4} \\ &- 1.43 \times \text{SERPINA3} - 0.72 \times \text{MMP12} - 0.59 \times \text{TNN13} \end{aligned}$$