機械学習を解釈する技術2

廣田雄亮

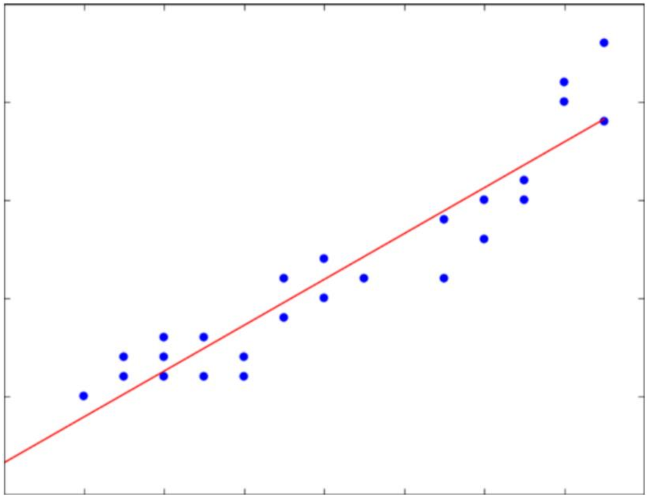
2章
「解釈性」の理解

前提

線形回帰モデルを解釈する方法を学ぶ.

回帰問題とは
目的変数が連続地の場合に予測を行うタスク

線形回帰モデルはシンプル
ふるまいに対する透明性が高い
→　解釈性が高い



線形回帰モデルが備える解釈性

特徴量と予測値の関係が明らかであること

1. 特徴量とモデルの予測値の平均的な関係

→　PD

2. 特徴量の重要度

→　PFI

3. 特徴量と予測値のインスタンス毎の関係

→　ICE

4. モデルの出したインスタンス毎の予測値の理由

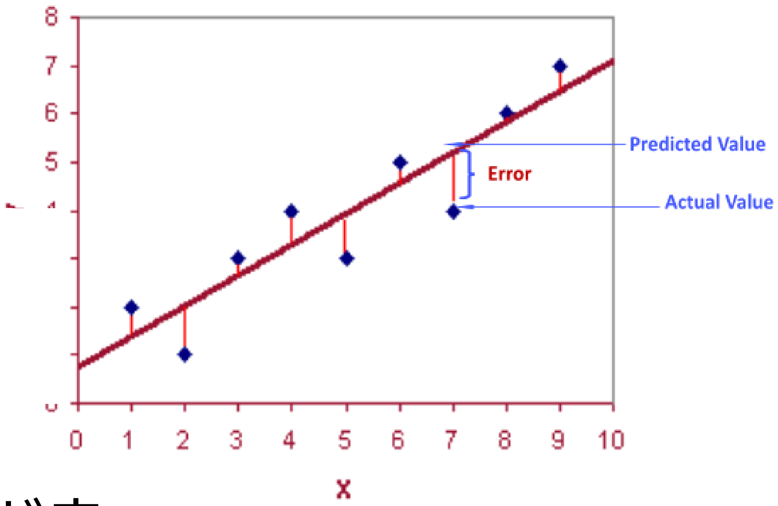
→　SHAP

線形回帰モデルの評価指標

- **RMSE**　小さいほど良い

平均的な予測誤差

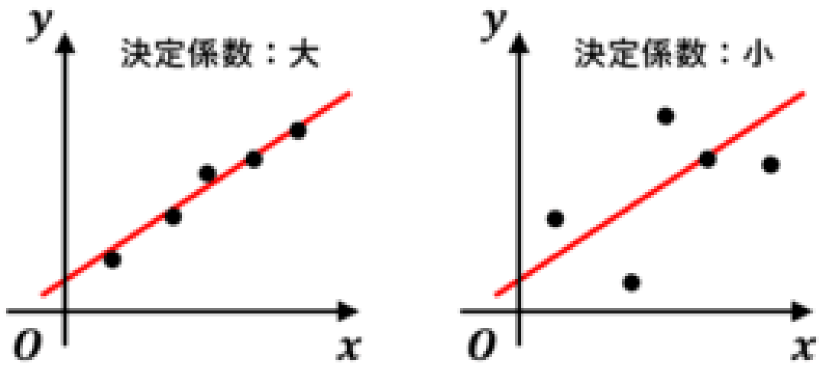
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$



- **R^2（決定係数）**大きいほど良い

単純に平均値を返すモデルと比較して，
どの程度精度が良いか

$$R^2 \equiv 1 - \frac{\sum_{i=1}^{N}(y_i - f_i)^2}{\sum_{j=1}^{N}(y_j - \bar{y})^2}$$



特徴量とモデルの予測値の平均的な関係

$$f(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

特徴量X_1が1単位増加すると予測値はβ_1増加する

このことは全てのインスタンスで共通

⇓

回帰係数は特徴量とモデルの
予測値の**平均的な (インスタンス毎ではない)**
関係を解釈している

特徴量と予測値のインスタンス毎の関係(1/3)

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

特徴量x が1単位増加すると予測値の変化量は
上記の式を微分すると分かる

$$\frac{\partial f(x)}{\partial x} = \beta_1 + 2\beta_2 x$$

特徴量と予測値のインスタンス毎の関係(2/3)

$(\beta_1, \beta_2) = (1, 2)$ のとき
X = 1 のインスタンスでは，X が1増加した場合
予測値の変化量は $1 + 2 \times 2 \times 1 = 5$

X = 10 のインスタンスでは，X が1増加した場合
予測値の変化量は $1 + 2 \times 2 \times 10 = 41$

特徴量と予測値の**各データ毎**の関係が明らか

特徴量と予測値のインスタンス毎の関係(3/3)

交互作用がある場合

$$f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \underline{\beta_2 x_1 x_2}$$

特徴量x が1単位増加すると予測値の変化量は

$$\frac{\partial f(x_1, x_2)}{\partial x} = \beta_1 + \beta_2 x_2$$

特徴量と予測値のインスタンス毎の関係(3/3)

交互作用がある場合

$$f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \underline{\beta_2 x_1 x_2}$$

(β_1, β_2) = (1, 2) のとき
X_2 = 1 のインスタンスでは，X_1 が1増加した場合
予測値の変化量は 1 + 2 × 1 = 3

(β_1, β_2) = (1, 2) のとき
X_2 = 10 のインスタンスでは，X_1 が1増加した場合
予測値の変化量は 1 + 2 × 10 = 21

交互作用がある場合

$$f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \underline{\beta_2 x_1 x_2}$$

インスタンス毎の
説明が可能

(β_1, β_2) = (1, 2) のとき
X_2 = 1 のインスタンスでは，X_1 が1増加した場合
予測値の変化量は 1 + 2 × 1 = <u>3</u>

(β_1, β_2) = (1, 2) のとき
X_2 = 10 のインスタンスでは，X_1 が1増加した場合
予測値の変化量は 1 + 2 × 10 = <u>21</u>

回帰係数による重要度比較

$$f(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

特徴量 (X_1, X_2, X_3) が1単位ずつ増加した場合の予測値の変化量はそれぞれ (β_1, β_2, β_3)

(β_1, β_2, β_3) = (0, 1, 10) の場合

- 最も予測値が変動するのは X_3
- X_1 は全く影響はない

ように見える.

回帰係数による重要度比較

回帰係数とは
特徴量が1単位増加した場合の予測値に与える影響

特徴量毎に「1単位増加する」ことの意味が
大きく異なるケースがある

⇓

特徴量のスケールを揃えてから比較するべき

特徴量のスケールが異なる場合

例

現職年収予測値(万円) =

1 × 前職年数(万円) + 100 × 現職経験年数

前職年収は，300万 ~ 1000万 といった場合に
考えられる範囲が広い

経験年数は，1年 ~ 60年 といった場合に
考えられる範囲が狭い

⇒ 正答な比較が難しい

特徴量の重要度(2/3)

標準化

特徴量のスケールが異なる場合

⇒　標準化を行い，各特徴量のスケールを同じにして，回帰係数を比較

標準化

平均値からの偏差を標準偏差で割ること.

平均は0，分散は1となる.

$$\tilde{x}_i = \frac{x_i - \bar{x}}{\widehat{SD}(x)}$$

平均：$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

標準偏差：$\widehat{SD}(x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$