

Example: Cross fitting and it's applications

機械学習の経済学への応用

川田恵介

交差推定 (Cross fitting)

- “交差推定” は機械学習の応用において、極めて重要
 - 予測問題: HyperParameter Tuning, Algorithm 選択, Stacking
 - 母集団の推論: 収束性質の改善
- しっかりイメージを持つことが重要
 - 現代的パッケージを用いると (べきだが)、イメージしにくいかも、、、

例: 元データ

```
library(tidyverse)
library(magrittr)

Data <- tibble(
  X = c(4, 2, 1, 1, 2, 5),
  Y = c(10, 19, 10, 2, 10, 20)
)
```

Data

A tibble: 6 x 2

	X	Y
	<dbl>	<dbl>
1	4	10
2	2	19
3	1	10
4	1	2

5	2	10
6	5	20

例: グループ分け

```
Data <- mutate(Data,
  G = sample(1:3,6,replace = TRUE)
)
```

Data

```
# A tibble: 6 x 3
      X     Y     G
  <dbl> <dbl> <int>
1     4    10     2
2     2    19     3
3     1    10     2
4     1     2     3
5     2    10     1
6     5    20     3
```

例: G = 1 への予測値

```
tibble(X = c(4,2,1,1,2,5),
  Y = c(10,19,10,2,10,20),
  G = c(1,2,1,3,3,2),
  MeanPrediction = c(12.75,NA,12.75,NA,NA,NA),
  KawataPrediction = c(10,NA,9,NA,NA,NA)
)
```

```
# A tibble: 6 x 5
      X     Y     G MeanPrediction KawataPrediction
  <dbl> <dbl> <dbl>          <dbl>              <dbl>
1     4    10     1          12.8                10
2     2    19     2           NA                 NA
3     1    10     1          12.8                 9
4     1     2     3           NA                 NA
5     2    10     3           NA                 NA
6     5    20     2           NA                 NA
```

例: $G = 2$ への予測値

```
tibble(X = c(4,2,1,1,2,5),
       Y = c(10,19,10,2,10,20),
       G = c(1,2,1,3,3,2),
       MeanPrediction = c(12.75,8,12.75,NA,NA,8),
       KawataPrediction = c(10,10,9,NA,NA,5)
)
```

A tibble: 6 x 5

	X	Y	G	MeanPrediction	KawataPrediction
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	4	10	1	12.8	10
2	2	19	2	8	10
3	1	10	1	12.8	9
4	1	2	3	NA	NA
5	2	10	3	NA	NA
6	5	20	2	8	5

例: $G = 3$ への予測値

```
tibble(X = c(4,2,1,1,2,5),
       Y = c(10,19,10,2,10,20),
       G = c(1,2,1,3,3,2),
       MeanPrediction = c(12.75,8,12.75,14.75,14.75,8),
       KawataPrediction = c(10,10,9,10,20,5)
)
```

A tibble: 6 x 5

	X	Y	G	MeanPrediction	KawataPrediction
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	4	10	1	12.8	10
2	2	19	2	8	10
3	1	10	1	12.8	9
4	1	2	3	14.8	10
5	2	10	3	14.8	20
6	5	20	2	8	5

- 交差推定完了!!!

例: Algorithm 比較

```
tibble(X = c(4,2,1,1,2,5),
       Y = c(10,19,10,2,10,20),
       G = c(1,2,1,3,3,2),
       MeanPrediction = c(12.75,8,12.75,14.75,14.75,8),
       KawataPrediction = c(10,10,9,10,20,5)
) |>
mutate(MeanEval = (Y - MeanPrediction)^2,
       KawataEval = (Y - KawataPrediction)^2)
```

A tibble: 6 x 7

	X	Y	G	MeanPrediction	KawataPrediction	MeanEval	KawataEval
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	4	10	1	12.8	10	7.56	0
2	2	19	2	8	10	121	81
3	1	10	1	12.8	9	7.56	1
4	1	2	3	14.8	10	163.	64
5	2	10	3	14.8	20	22.6	100
6	5	20	2	8	5	144	225

例: Stacking

```
tibble(X = c(4,2,1,1,2,5),
       Y = c(10,19,10,2,10,20),
       G = c(1,2,1,3,3,2),
       MeanPrediction = c(12.75,8,12.75,14.75,14.75,8),
       KawataPrediction = c(10,10,9,10,20,5)
) %$%
lm(Y ~ MeanPrediction + KawataPrediction)
```

Call:

```
lm(formula = Y ~ MeanPrediction + KawataPrediction)
```

Coefficients:

(Intercept)	MeanPrediction	KawataPrediction
35.9910	-2.4829	0.4897