1. Simulation – Parameter Tuning

<model 1>

$Y_1 = \sin(\beta_1^T X) + \sin(\beta_2^T X) + 0.5\varepsilon$

n=300, p=40

- Parameter tuning

Case (1): $\beta_1 = (1, 0, \ldots, 0)^\mathsf{T}$ $\beta_2 = (0, 1, 0, \ldots, 0)^\mathsf{T}$.

λ1 = 1, 0.1, λ2 = 0.1

cor(X β1, X β1hat), cor(X β2, X β2hat) 결과

```
> sdr.cor1
  [1] 1.00000000 0.63256202 0.94285063 0.86771067 0.97415693 0.86163925 0.97474127 0.63842329
  [9] 0.98965313 0.75973234 0.98549293 0.49489651 0.88445943 0.82932165 0.71034932 1.00000000
 [17] 0.99804892 1.00000000 0.87855379 1.00000000 0.94456308 0.74799164 0.97132643 1.00000000
 [25] 0.99998860 1.00000000 0.90771863 0.97942711 1.00000000 1.00000000 0.85856093         NA
 [33] 0.80075432 1.00000000 0.94115398 1.00000000 1.00000000 0.81400027         NA 0.87593045
 [41] 0.60939298 0.68832506 0.96054537 0.99943082 0.81517864 1.00000000 1.00000000 0.93818075
 [49] 0.04359431 0.99961630 0.71507125 1.00000000         NA 0.94027010 0.92677681 0.91388008
 [57] 1.00000000 0.92805944 0.95721935 0.84215137 0.97565969 0.70883596 0.84582198 1.00000000
 [65] 0.75876659 0.64641064 0.83642208 0.44188717 0.89310112 0.92445595 0.04191160 1.00000000
 [73] 0.86888683 0.99999996 0.66150154         NA 0.96708420 0.66308496 1.00000000         NA
 [81] 0.93330154 1.00000000 0.95841304 0.87253646 0.97168465 0.77213758 0.97629129 0.93327310
 [89] 0.99736680 0.99997803         NA         NA 0.95214068 0.99999980 0.99140544 0.79544477
 [97] 0.99999173 0.96038084 0.83901865 0.95812290
> sdr.cor2
  [1]         NA         NA         NA         NA         NA         NA         NA         NA
  [9]         NA         NA         NA         NA         NA         NA         NA         NA
 [17]         NA         NA         NA         NA         NA         NA         NA         NA
 [25]         NA         NA         NA         NA         NA         NA         NA 0.02620701
 [33]         NA         NA         NA         NA         NA         NA 1.00000000         NA
 [41]         NA         NA         NA         NA         NA         NA         NA         NA
 [49]         NA         NA         NA         NA 1.00000000         NA         NA         NA
 [57]         NA         NA         NA         NA         NA         NA         NA         NA
 [65]         NA         NA         NA         NA         NA         NA         NA         NA
 [73]         NA         NA         NA 0.03056433         NA         NA         NA 0.08330481
 [81]         NA         NA         NA         NA         NA         NA         NA         NA
 [89]         NA         NA 0.00191585 0.05083876         NA         NA         NA         NA
 [97]         NA         NA         NA         NA
> 
```

E((X β1-X β1hat)^2), E((X β2-X β2hat)^2),

```
> sdr.mse1
  [1] 0.000000e+00 7.476178e-01 1.261891e-01 2.728944e-01 5.521411e-02 2.900899e-01 5.826591e-02
  [8] 7.182753e-01 2.407531e-02 5.130315e-01 3.494688e-02 9.887507e-01 2.403862e-01 3.434899e-01
 [15] 5.949989e-01 0.000000e+00 7.667778e-03 1.909669e-03 2.501187e-01 0.000000e+00 1.348478e-01
 [22] 5.267177e-01 6.622695e-02 0.000000e+00 0.000000e+00 0.000000e+00 1.903280e-01 4.612699e-02
 [29] 0.000000e+00 1.157094e-03 3.085684e-01 1.008170e+00 4.193278e-01 0.000000e+00 1.505737e-01
 [36] 0.000000e+00 0.000000e+00 3.907449e-01 1.838989e+00 2.608024e-01 7.936916e-01 5.989798e-01
 [43] 8.964304e-02 2.120690e-03 3.955302e-01 9.046974e-05 0.000000e+00 1.553022e-01 1.008170e+00
 [50] 1.234875e-03 6.005986e-01 0.000000e+00 1.892675e+00 1.569183e-01 1.603256e-01 1.787350e-01
 [57] 0.000000e+00 1.749605e-01 8.999813e-02 3.379163e-01 5.713474e-02 6.018052e-01 3.221518e-01
 [64] 0.000000e+00 4.632470e-01 7.172863e-01 3.317833e-01 1.049284e+00 2.175747e-01 1.625136e-01
 [71] 1.008170e+00 0.000000e+00 2.609267e-01 0.000000e+00 6.685040e-01 7.867541e-02 8.195144e-02
 [78] 6.576702e-01 0.000000e+00 1.008170e+00 1.738385e-01 0.000000e+00 1.077306e-01 2.797278e-01
 [85] 5.646986e-02 4.793600e-01 5.905189e-02 1.683341e-01 1.337356e-02 0.000000e+00 2.724739e-01
 [92] 1.008170e+00 9.839401e-02 2.138596e-04 2.346760e-02 4.177630e-01 0.000000e+00 8.824515e-02
 [99] 3.343588e-01 9.252100e-02
> sdr.mse2
  [1] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [10] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [19] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [28] 0.9707594 0.9707594 0.9707594 0.9707594 1.9888279 0.9707594 0.9707594 0.9707594 0.9707594
 [37] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [46] 0.9707594 0.9707594 0.9707594 0.0000000 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [55] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [64] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.0000000 0.9707594
 [73] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 1.7164334 0.9707594
 [82] 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
 [91] 0.9707594 2.0940471 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594 0.9707594
[100] 0.9707594
```

β 추정치가 모두 0이 나오는 경우가 있어 cor(X β, X βhat) = NA가 나온다.

이런 경우 simulation 횟수를 늘려서 NA를 제거하고 결과를 구해야하는건가?

Case (1)	Sparse DR

Case (1)

: $\beta_1 = (1, 0, \ldots, 0)^{\mathsf{T}}$ $\beta_2 = (0, 1, 0, \ldots, 0)^{\mathsf{T}}$

λ1 = 0.1, 0.1

λ2 = 0.01

slice: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.7092513	0.1216416	2.602801	1.9654668
sparse dr	77.93	0.8915245	0.7511124	2.755059	0.9020436

NA가 50개 이상 나오기 때문에 NA를 제거하고 나머지로 평균을 낸 결과

Case (1)

: $\beta_1 = (1, 0, \ldots, 0)^{\mathsf{T}}$ $\beta_2 = (0, 1, 0, \ldots, 0)^{\mathsf{T}}$

λ1 = 0.1, 0.1

λ2 = 0.1

slice: 4개(slice 개수를 늘리면

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.7092513	0.1216416	2.602801	1.9654668
sparse dr	78.03	0.8804943	0.8106900	2.735991	0.8831386

NA가 50개 이상 나오기 때문에 NA를 제거하고 나머지로 평균을 낸 결과

Case (1)

: $\beta_1 = (1, 0, \ldots, 0)^{\mathsf{T}}$ $\beta_2 = (0, 1, 0, \ldots, 0)^{\mathsf{T}}$

λ1 = 0.1, 0.05

λ2 = 0.05

slice 개수: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.7092513	0.12164155	2.602801	1.965467
sparse dr	77.16	0.8865939	0.06533786	2.749936	1.929122

NA는 없음

Case (1)

: $\beta_1 = (1, 0, \ldots, 0)^{\mathsf{T}}$ $\beta_2 = (0, 1, 0, \ldots, 0)^{\mathsf{T}}$

λ1 = 0.1, 0.01

λ2 = 0.01

slice 개수: 10개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.693193	0.1045944	2.838022	1.963565
sparse dr	62.33	0.892805	0.4696206	2.952623	2.441872

NA는 없음

Case (1) –가장 처음 결과

: $\beta_1 = (1, 0, \ldots, 0)^{\mathsf{T}}$ $\beta_2 = (0, 1, 0, \ldots, 0)^{\mathsf{T}}$

λ1 = 0.01, 0.01

λ2 = 0.01

slice 개수: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.7092513	0.12164155	2.602801	1.965467
sparse dr	59.94	0.7548557	0.07432311	2.724729	2.125324

NA 없음

Case (2)	Sparse DR

Case (2)

: β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)

β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)

λ1 = 0.05, 0.01

λ2 = 0.01

slice: 10개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.1373907	0.1228866	2.074601	1.904770
sparse dr	62.66	0.1829674	0.1011862	2.230346	2.137052

NA없음

Case (2)

: β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)

β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)

λ1 = 0.05, 0.01

λ2 = 0.01

slice: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.1391564	0.1217259	2.140710	1.942375
sparse dr	66.65	0.1422226	0.1467648	2.127528	2.093356

NA 없음

Case (2)

: β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)

β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)

λ1 = 0.05, 0.05

λ2 = 0.01

slice: 4개

method	p	corr1	corr2	mse1	mse2
dr	0	0.1416286	0.1105666	2.086301	1.882642
sparse dr	78	0.1094201	0.1119920	2.069400	1.972753

NA 나오지 않음 – lambda를 늘렸더니 beta 0의 개수가 너무 많게 나옴

Case (2)—가장 처음 결과

: β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)

β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)

λ1 = 0.01, 0.01

λ2 = 0.01

slice: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.1395863	0.1152813	2.099900	1.933958
sparse dr	55.27	0.1792127	0.1335033	2.245751	2.133104

NA나오지 않음

Case (3)	Sparse DR

Case (3)

: β_1 =(1,1,1,1,1,1,1,1,1,1 ,0,0...0,0)

β_2 =(0,0...0,0,0,1,1,1,1,1,1,1,1,1,1)

λ1 = 0.05, 0.05

λ2 = 0.05

slice: 10개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.1543846	0.1349226	2.146090	1.944538
sparse dr	77.96	0.1269302	0.1231912	2.113108	1.934631

NA 나오지 않음 – lambda를 늘렸더니 beta 0의 개수가 너무 많게 나옴

Case (3)

: β_1 =(1,1,1,1,1,1,1,1,1,1 ,0,0...0,0)

β_2 =(0,0...0,0,0,1,1,1,1,1,1,1,1,1,1)

λ1 = 0.1, 0.01

λ2 = 0.01

slice: 10개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.1610913	0.1147005	2.078947	1.923570
sparse dr	63.01	0.2124721	0.1095257	2.088260	2.162515

NA나오지 않음

Case (3)—가장 처음 결과

: β_1 =(1,1,1,1,1,1,1,1,1,1 ,0,0...0,0)

β_2 =(0,0...0,0,0,1,1,1,1,1,1,1,1,1,1)

λ1 = 0.01, 0.01

λ2 = 0.01

slice: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.1407005	0.1310424	2.094963	1.959305
sparse dr	55.28	0.2043405	0.1515334	2.277170	2.051917

NA나오지 않음

<model 2>

$Y_2 = \frac{\beta_1^T X}{0.5+(2\beta_2^T X+0.3)^2} + 0.3\varepsilon$, n=300, p=50

case	Sparse DR

Case (1)

: $\beta_1 = (1, 0, \ldots, 0)^{\top}$, $\beta_2 = (0, 1, 0, \ldots, 0)^{\top}$

λ1 = 0.1, 0.1

λ2 = 0.1

slice: 10개

method	p	corr1	corr2	mse1	mse2
dr	0	0.8344419	0.1473274	0.3430334	2.111377
sparse dr	98	1.0000000	0.5477459	0.0000000	1.223341

NA 없음

Case (1) – 가장 처음 결과

: $\beta_1 = (1, 0, \ldots, 0)^{\top}$, $\beta_2 = (0, 1, 0, \ldots, 0)^{\top}$

λ1 = 0.01, 0.01

λ2 = 0.01

slice: 4개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.8219249	0.2578787	0.36904806	1.966568
sparse dr	79.11	0.9833684	0.4017083	0.03622688	2.077622

NA 없음

case	Sparse DR

Case (2)

β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)

β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)

λ1 = 0.05, 0.05

λ2 = 0.01

slice: 10개

method	p	corr1	corr2	mse1	mse2
dr	0.00	0.7676112	0.1834386	3.420411	2.004947
sparse dr	97.84	0.5064309	0.1217545	2.915175	2.048526

NA 없음 - lambda를 늘렸더니 beta 0의 개수가 너무 많게 나옴

Case (2)

β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)

β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)

λ1 = 0.05, 0.01

λ2 = 0.01

method	p	corr1	corr2	mse1	mse2
dr	0.0	0.7676112	0.1834386	3.420411	2.004947
sparse dr	80.7	0.4086287	0.1507352	2.542006	2.208538

NA없음

<table>
<tr><td>slice: 10개</td><td></td></tr>
<tr><td>Case (2) – 가장 처음 결과
β_1 =(1,1,1,1,1,0.1,0.1,0.1,0.1,0.1,0,0...0,0)
β_2 =(0,0...0,0,0.1,0.1,0.1,0.1,0.1,1,1,1,1,1)
λ1 = 0.01, 0.01
λ2 = 0.01
slice: 4개</td><td>
<table>
<tr><th>method</th><th>p</th><th>corr1</th><th>corr2</th><th>mse1</th><th>mse2</th></tr>
<tr><td>dr</td><td>0.00</td><td>0.6001518</td><td>0.1900158</td><td>2.916600</td><td>2.112913</td></tr>
<tr><td>sparse dr</td><td>75.04</td><td>0.6409069</td><td>0.2345991</td><td>2.955934</td><td>2.252680</td></tr>
</table>
NA 없음</td></tr>
</table>

<table>
<tr><th>case</th><th>Sparse DR</th></tr>
<tr><td>Case (3)
β_1 =(1,1,1,1,...,1,1,1,0,0...0,0)
β_2 =(0,0...0,0,0,1,1,1,1,...,1,1)
λ 1 = 0.05, 0.01
λ2 = 0.01
slice: 4개</td><td>
<table>
<tr><th>method</th><th>p</th><th>corr1</th><th>corr2</th><th>mse1</th><th>mse2</th></tr>
<tr><td>dr</td><td>0.00</td><td>0.3479419</td><td>0.1719531</td><td>2.255180</td><td>2.009467</td></tr>
<tr><td>sparse dr</td><td>85.56</td><td>0.2734218</td><td>0.1649450</td><td>2.192834</td><td>2.147575</td></tr>
</table>
NA 없음</td></tr>
<tr><td>Case (3)
β_1 =(1,1,1,1,...,1,1,1,0,0...0,0)
β_2 =(0,0...0,0,0,1,1,1,1,...,1,1)
λ1 = 0.05, 0.01
λ2 = 0.01
slice: 10개</td><td>
<table>
<tr><th>method</th><th>p</th><th>corr1</th><th>corr2</th><th>mse1</th><th>mse2</th></tr>
<tr><td>dr</td><td>0.00</td><td>0.4743199</td><td>0.1638061</td><td>2.259702</td><td>2.030333</td></tr>
<tr><td>sparse dr</td><td>81.02</td><td>0.2930573</td><td>0.1406950</td><td>2.327612</td><td>2.197306</td></tr>
</table>
NA없음</td></tr>
<tr><td>Case (3) – 가장 처음 결과
β_1 =(1,1,1,1,...,1,1,1,0,0...0,0)
β_2 =(0,0...0,0,0,1,1,1,1,...,1,1)
λ1 = 0.01, 0.01
λ2 = 0.01
slice: 4개</td><td>
<table>
<tr><th>method</th><th>p</th><th>corr1</th><th>corr2</th><th>mse1</th><th>mse2</th></tr>
<tr><td>dr</td><td>0.00</td><td>0.3479419</td><td>0.1719531</td><td>2.255180</td><td>2.009467</td></tr>
<tr><td>sparse dr</td><td>73.48</td><td>0.3639681</td><td>0.1779225</td><td>2.360421</td><td>2.225217</td></tr>
</table>
NA 없음</td></tr>
</table>

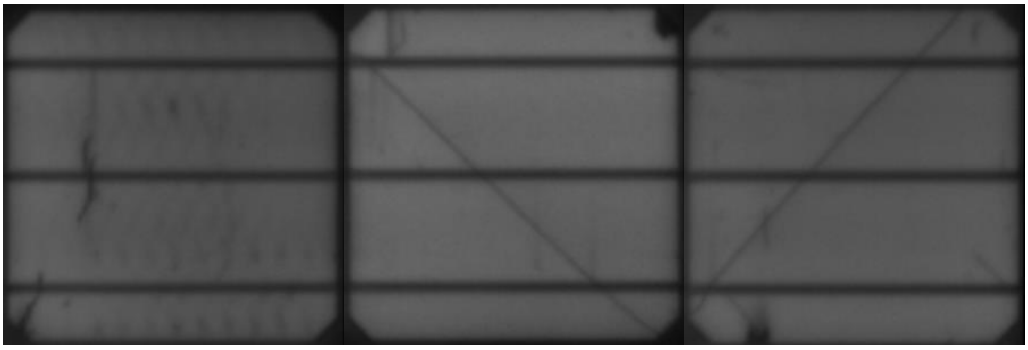
2. Application

Data: 태양 전지 결함 데이터 – 전기 발광 이미지 (Visual Identification of Defective Solar Cells in Electroluminescence Imagery)

https://github.com/zae-bayern/elpv-dataset : 300*300, n=2624

44개의 서로 다른 태양 전지로부터 추출한 이미지이다.(정상/결함)

Images file (.png)



cell0001.png cell0002.png cell0003.png

labels.csv

attribute	Description
Image url	이미지 저장 이름
Defect probability	결함 확률(0, 0.33333, 0.66666, 1) 1로 가까워질수록 결함이 있는 태양 전지이다.
Solar module type	태양 전지의 종류(mono/poly) Mono: 단일 실리콘 결정으로 만들어진 검은색 태양전지를 가지고 있으며, 일반적으로 효율이 더 높다. 가격이 비싸다. Poly: 여러 개의 실리콘 결정으로 만들어지며 파란색 셀을 가지고 있다. 효율은 다소 떨어지지만 가격이 저렴하다.

- Mono와 poly를 구분해야할지? (둘 중 하나만 써야할지?)
- 왜 readpng함수를 쓰면 image가 grayscale로 안 받아지고 빨간색으로 나오는지?
- 메모리 문제