転移学習
Transfer learning

明治大学 理工学部 応用化学科
データ化学工学研究室　金子 弘昌

どんなときに転移学習が有効か？

✓あるデータセットを用いて、回帰モデル・クラス分類モデル $y=f(x)$ を構築し、x の値から y の値を推定したい

✓しかし、そのデータセットのサンプル数が小さく、適切なモデルが得られるか心配

✓そのデータセットのサンプルと、全く同じ環境で得られたというわけではないが、似た環境で得られたある程度サンプル数のあるデータセットがあり、2 つのデータセットで y の種類や x の変数が同じ

- 例) y や x を測定した装置が異なるデータセット
- 例) 実スケールのデータセットとパイロットスケールのデータセット

何を“転移”させるか？

✓ある程度の数があるサンプルを転移させる

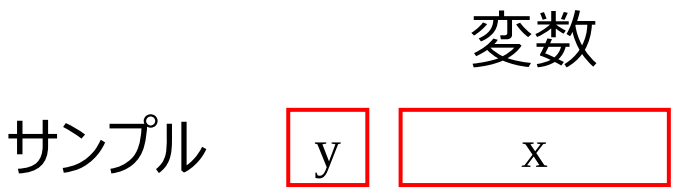
✓ある程度の数があるサンプルで学習したモデルを転移させる

- サンプルを転移させる方法に着目

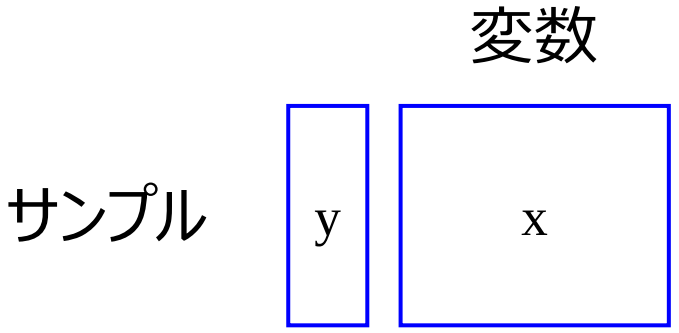
http://www.kamishima.net/archive/2009-tr-jsai_dmsm1-PR.pdf

2 種類のデータセットを有効に活用しよう！

ターゲットのデータセット (サンプル数 小)

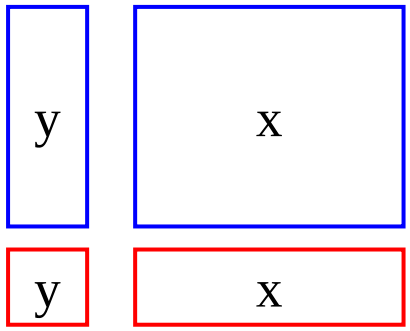


サポート用のデータセット (サンプル数 大)



一般的な解析

2 つのデータセットをサンプル方向につなげて、

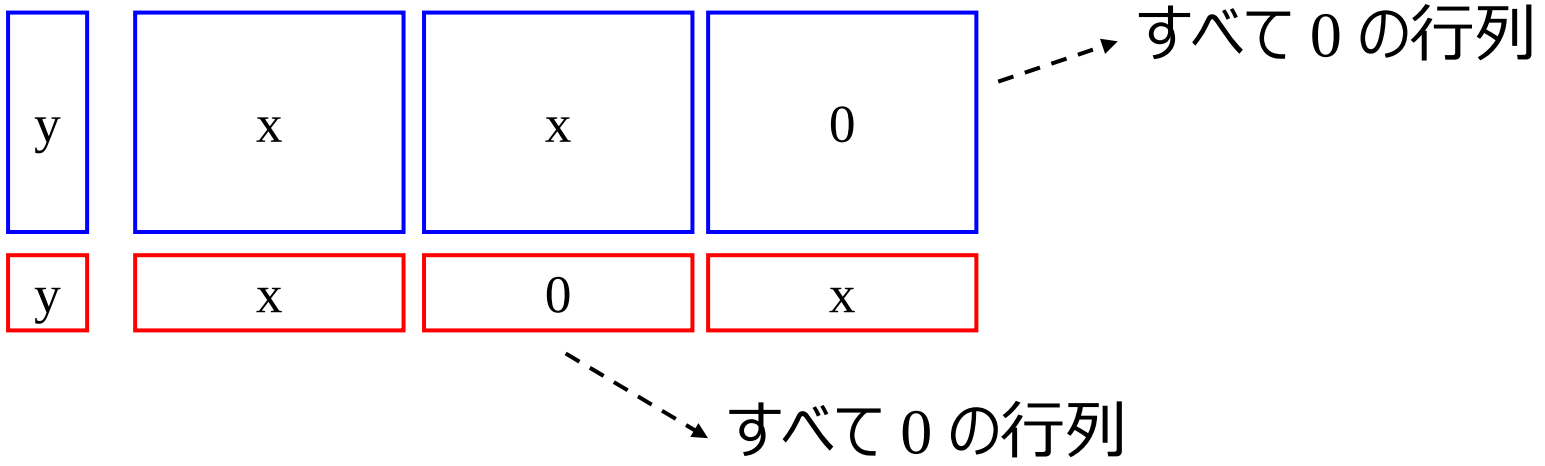


回帰分析手法・クラス分類手法でモデル y = f(x) 構築

モデルに、x の値を入力して、y の値を推定

転移学習

2 つのデータセットを

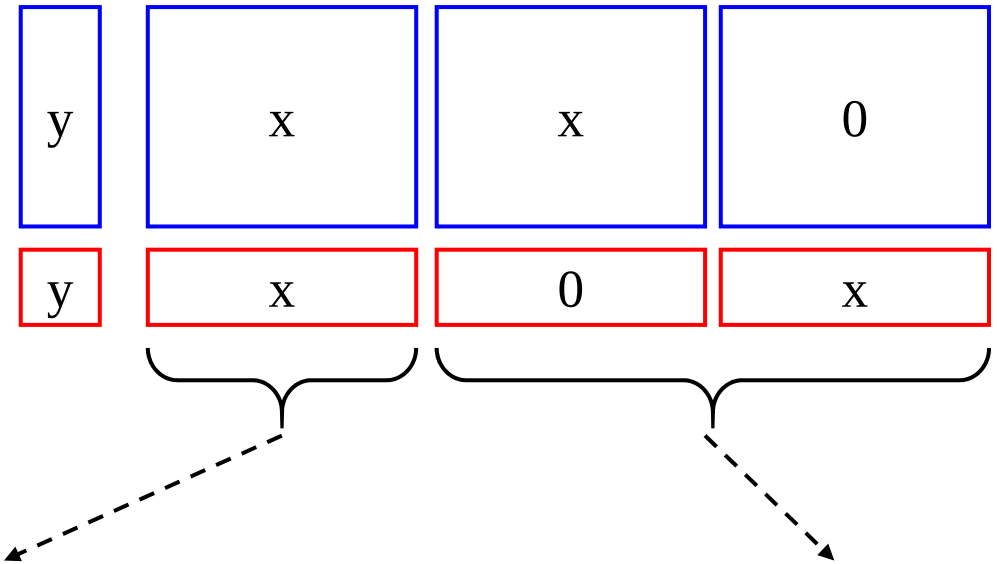


上のようにつなげて、回帰分析手法・クラス分類手法でモデル y = f(x) 構築

モデルに、 x 0 x の形式にした x の値を

入力して、y の値を推定

転移学習で期待すること



2 つのデータセットで共通する
x と y との間の関係を学習

2 つのデータセットで異なる
x と y との間の関係を学習

サポート用のデータセットも活用することで、共通する x と y との
関係を学習することで、異なる関係だけならターゲットのデータセットの
少ないサンプルでも学習できるか！？

数値シミュレーションデータで確認！

✓ ターゲットのデータセット 3 サンプル
✓ サポート用のデータセット 100 サンプル

の状況において、新たなターゲットの 100 サンプルを
正確に推定できるか？？

✓ケース1: x と y の間の傾きが、ターゲット・サポート用のデータセットで変化

- ターゲット: $y = 2x_1 + 4x_2 + 1$
- サポート用: $y = 2x_1 + 3x_2 + 1$

✓ケース2: 定数項 (y切片) が、ターゲット・サポート用のデータセットで変化
[x と y は非線形]

- ターゲット: $y = 2(x_1-2)^3 + 3x_2^2 + 3$
- サポート用: $y = 2(x_1-2)^3 + 3x_2^2 + 1$

比較した手法

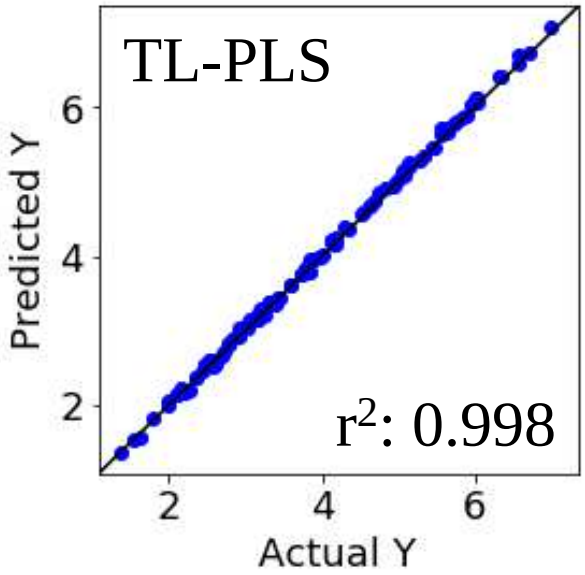
1. TL: Transfer Learning, 転移学習
2. OT: Only Target dataset, ターゲットのデータセット (3 サンプル) のみ使用
3. BD: Both target and supporting Dataset, ターゲットのデータセット (3 サンプル) とサポート用のデータセット (100 サンプル) 使用 [p. 4 の方法]

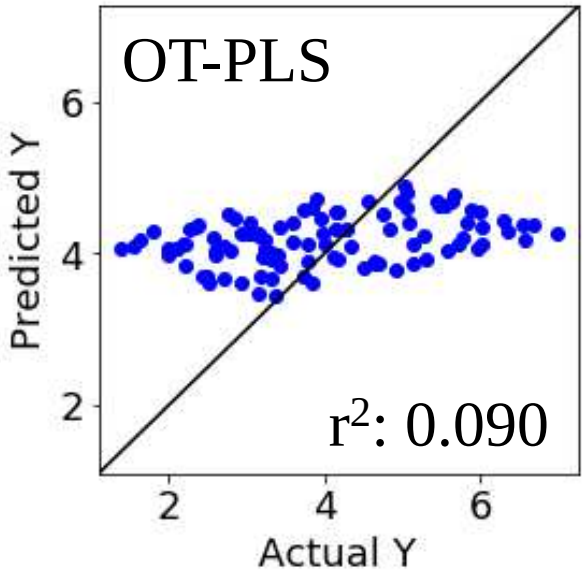
回帰分析手法は、

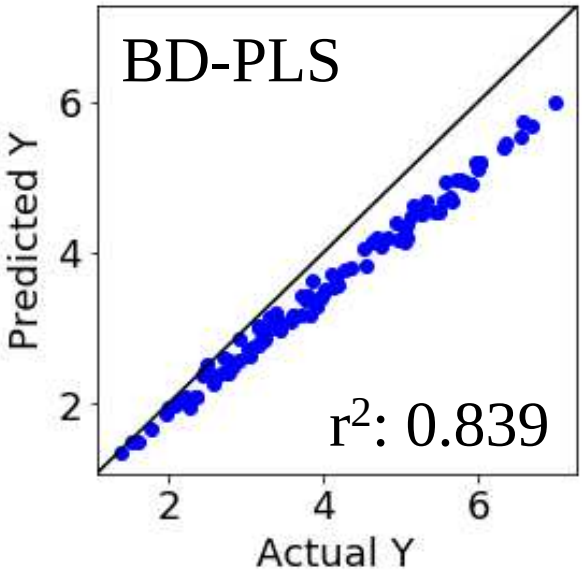
✓ Partial Least Squares (PLS)
✓ Gaussian Process Regression (GPR)

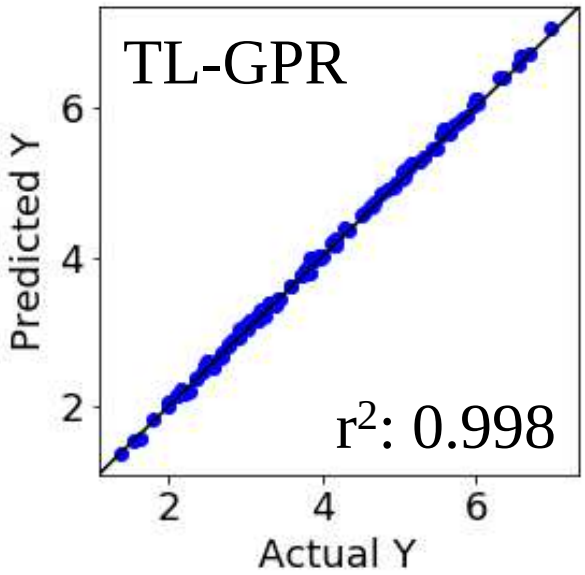
詳しくはこちら https://datachemeng.com/summarydataanalysis/

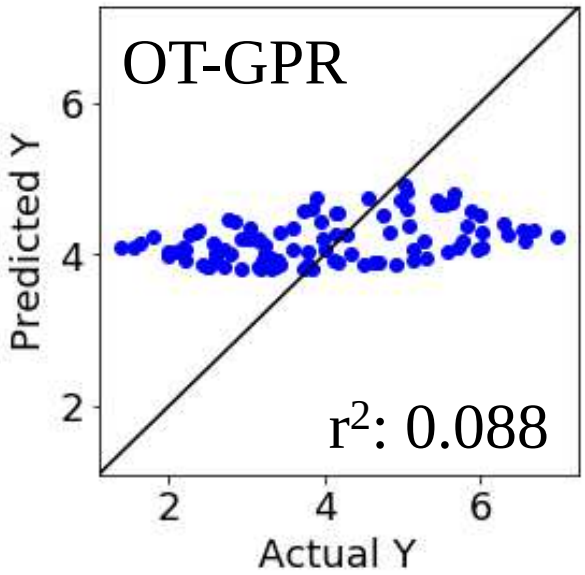
ケース1

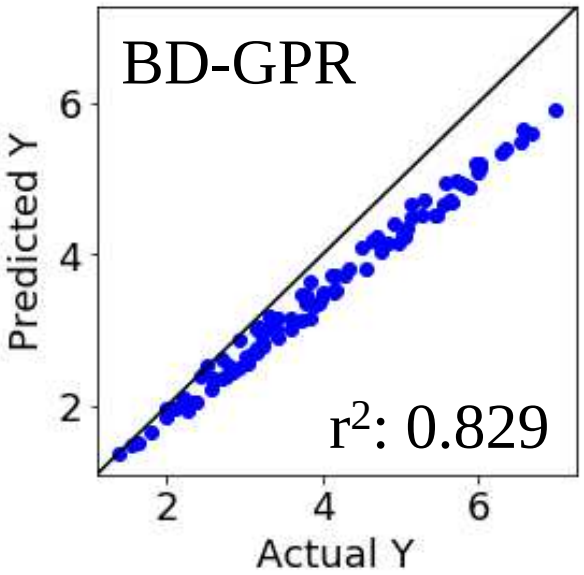

TL-PLS
r²: 0.998
Predicted Y
Actual Y
2
4
6


OT-PLS
r²: 0.090
Predicted Y
Actual Y
2
4
6

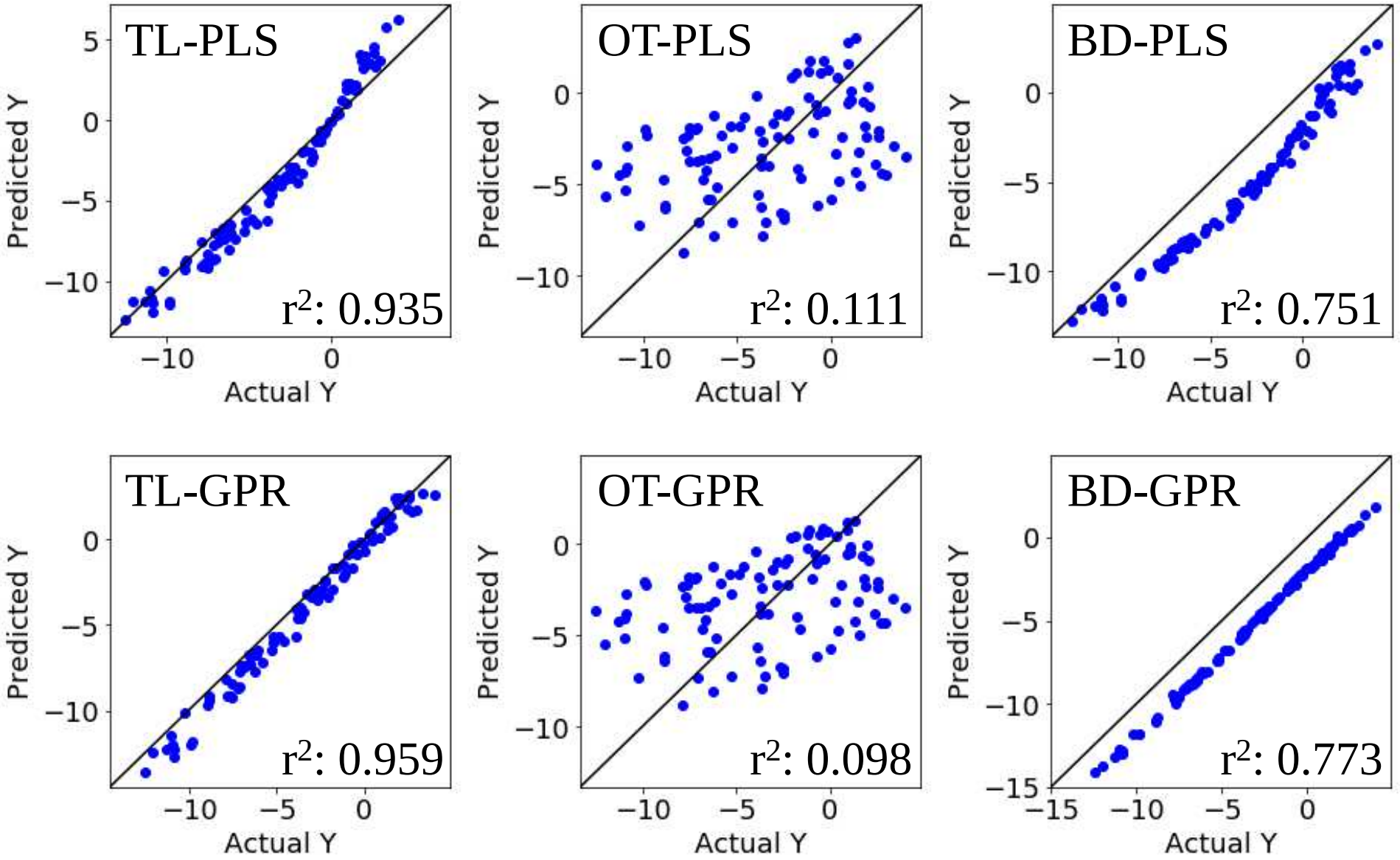

BD-PLS
r²: 0.839
Predicted Y
Actual Y
2
4
6


TL-GPR
r²: 0.998
Predicted Y
Actual Y
2
4
6


OT-GPR
r²: 0.088
Predicted Y
Actual Y
2
4
6


BD-GPR
r²: 0.829
Predicted Y
Actual Y
2
4
6

ケース2


TL-PLS
r²: 0.935
OT-PLS
r²: 0.111
BD-PLS
r²: 0.751
TL-GPR
r²: 0.959
OT-GPR
r²: 0.098
BD-GPR
r²: 0.773
Predicted Y
Actual Y

実際のデータセットで検証

✓Shootout 2012 のデータセット

- y : 医薬品中のAPIの重量パーセント濃度 [wt %]
- x : NIRスペクトル (ABB Bomem FT-NIR model MB-160)
  952.42, 953.12, ..., 1309.33 nm (372変数)

✓228 個の錠剤

- ラボスケール装置で製造 : 89 サンプル (shootout_2012_laboratory_scale.csv)
- パイロットスケール装置で製造 : 72 サンプル (shootout_2012_pilot_scale.csv)
- 実スケール装置で製造 : 67 サンプル (shootout_2012_full_scale.csv)

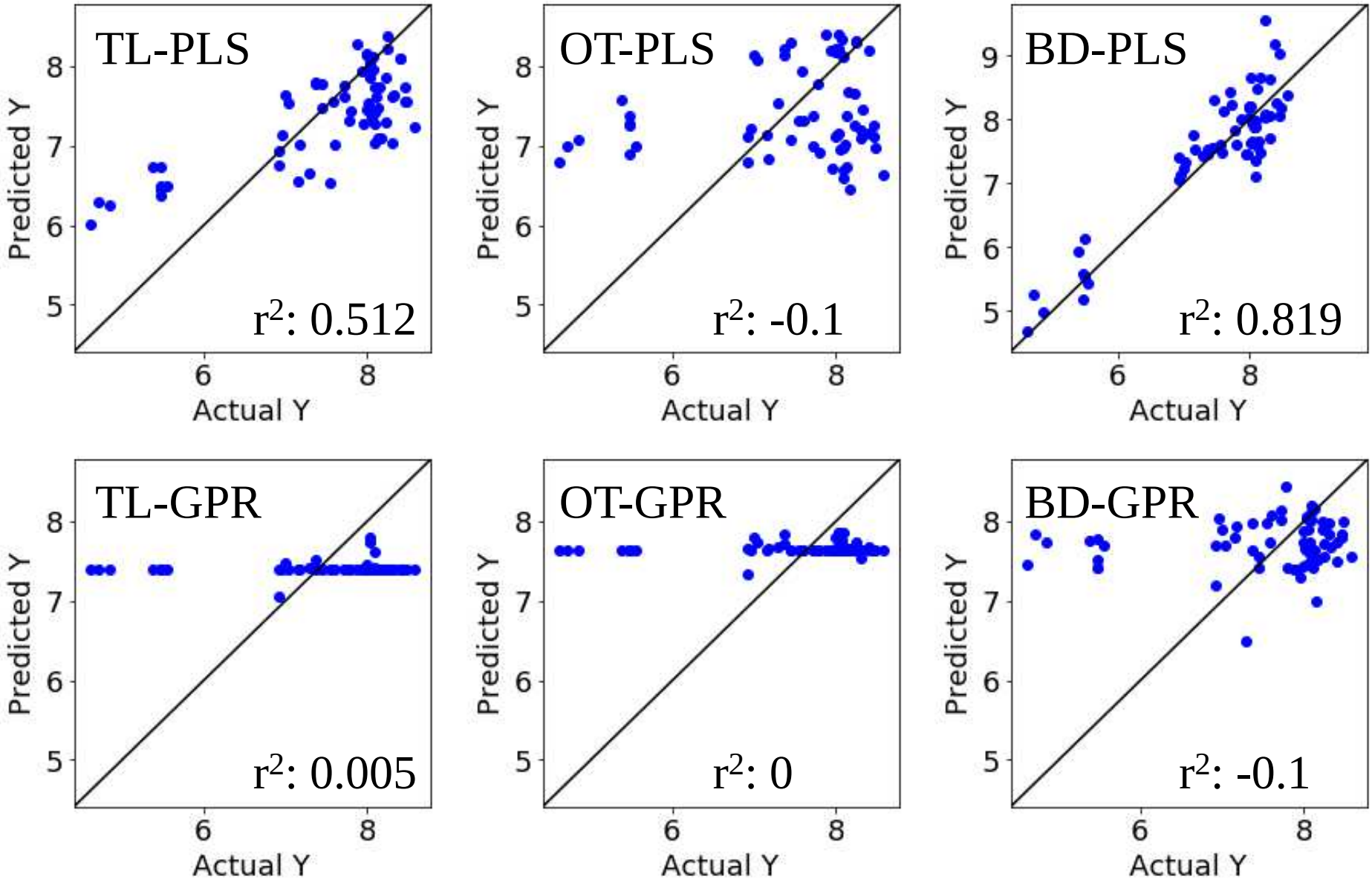
http://www.idrc-chambersburg.org/shootout-2012.html

想定したシチュエーション1

- ✓ 実スケール装置で製造されたターゲットのデータセット 3 サンプル
- ✓ パイロットスケールで製造されたサポート用のデータセット 72 サンプル

の状況において、新たなターゲットの 64 サンプルを
正確に推定できるか？？

結果1


TL-PLS
r²: 0.512
OT-PLS
r²: -0.1
BD-PLS
r²: 0.819
TL-GPR
r²: 0.005
OT-GPR
r²: 0
BD-GPR
r²: -0.1
Predicted Y
Actual Y

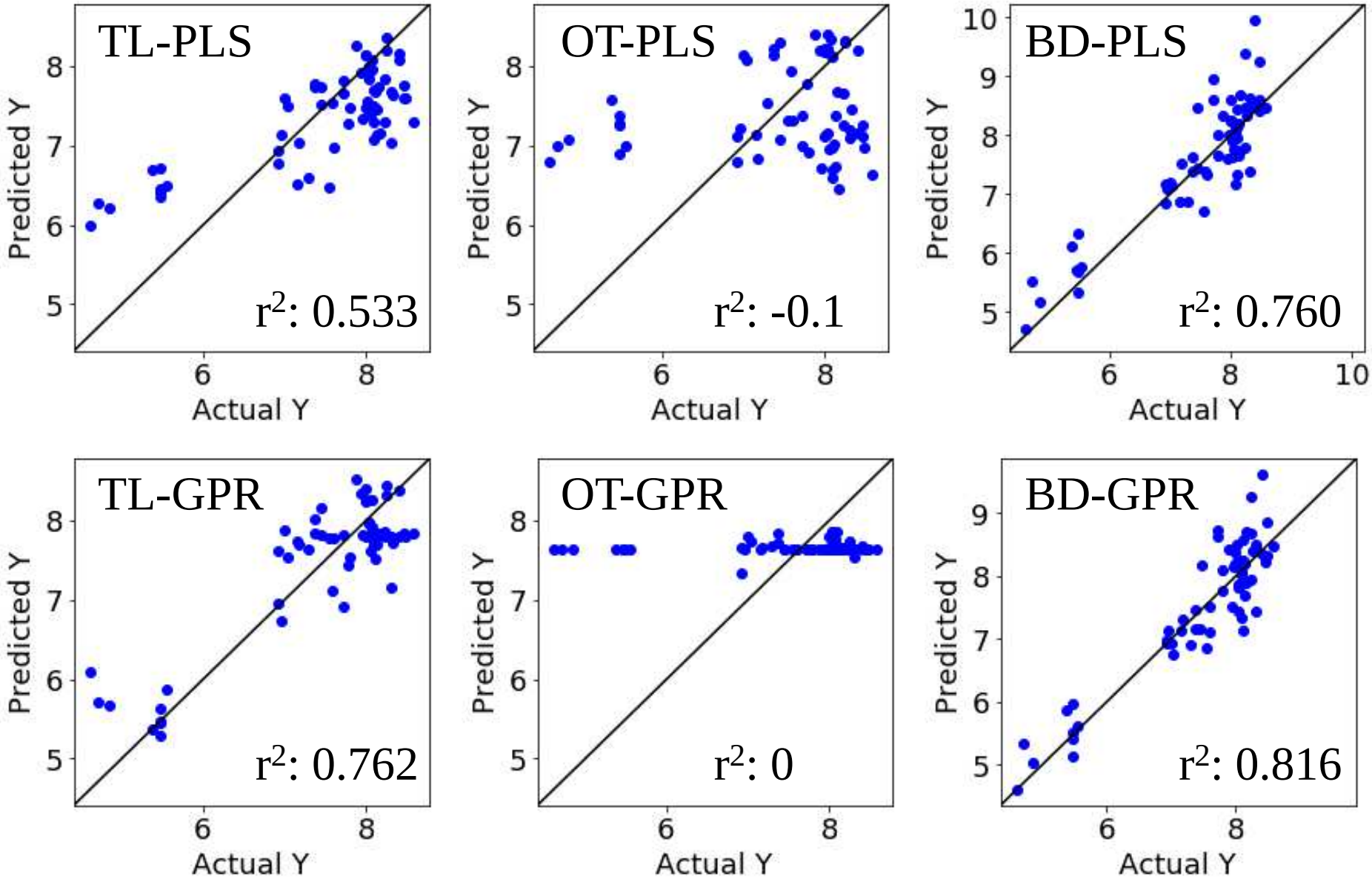
想定したシチュエーション2

- ✓ 実スケール装置で製造されたターゲットのデータセット 3 サンプル
- ✓ パイロットスケールで製造されたサポート用1のデータセット 72 サンプル、
  ラボスケールで製造されたサポート用2のデータセット 89 サンプル

の状況において、新たなターゲットの 64 サンプルを
正確に推定できるか？？

y	x	x	0	0
y	x	0	x	0
y	x	0	0	x

結果2


TL-PLS
r²: 0.533
OT-PLS
r²: -0.1
BD-PLS
r²: 0.760
TL-GPR
r²: 0.762
OT-GPR
r²: 0
BD-GPR
r²: 0.816
Predicted Y
Actual Y

考えごと

✓転移学習のときのハイパーパラメータの決定をどうするか

- いつも通りクロスバリデーションでよい？
- ターゲットのデータセットをよく推定できるように決める？
  - オーバーフィットしそう？
  - サポート用のサンプルがあるから問題ない？

✓スペクトル解析においては波長選択をしたほうがよさそう

✓評価関数をどうするか？

- クロスバリデーション後の r^2 ？
- ターゲットのデータセットにおける r^2 ？
  - オーバーフィットしそう？
  - サポート用のサンプルがあるから問題ない？