

Genetic Algorithm-based
Partial Least Squares
GAPLS

Genetic Algorithm-based
Support Vector Regression
GASVR

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

GAPLS, GASVR とは？

- ✓ GAPLS: 遺伝的アルゴリズム (GA) と PLS とを組み合わせた変数選択手法
- ✓ PLS でクロスバリデーションを行ったときの r^2 が大きくなるように変数が選択される
- ✓ GASVR: GA と SVR とを組み合わせた変数選択手法
- ✓ SVR でクロスバリデーションを行ったときの r^2 が大きくなるように変数が選択される
- ✓ 染色体にSVRのハイパーパラメータも入れることで、高速化している

PLS, SVR

✓ PLS, SVRについてはこちら

- PLS : <https://datachemeng.com/partialleastquares/>
- SVR : <https://datachemeng.com/supportvectorregression/>

遺伝的アルゴリズム (GA) の流れ

遺伝的アルゴリズム (Genetic Algorithm, GA) :

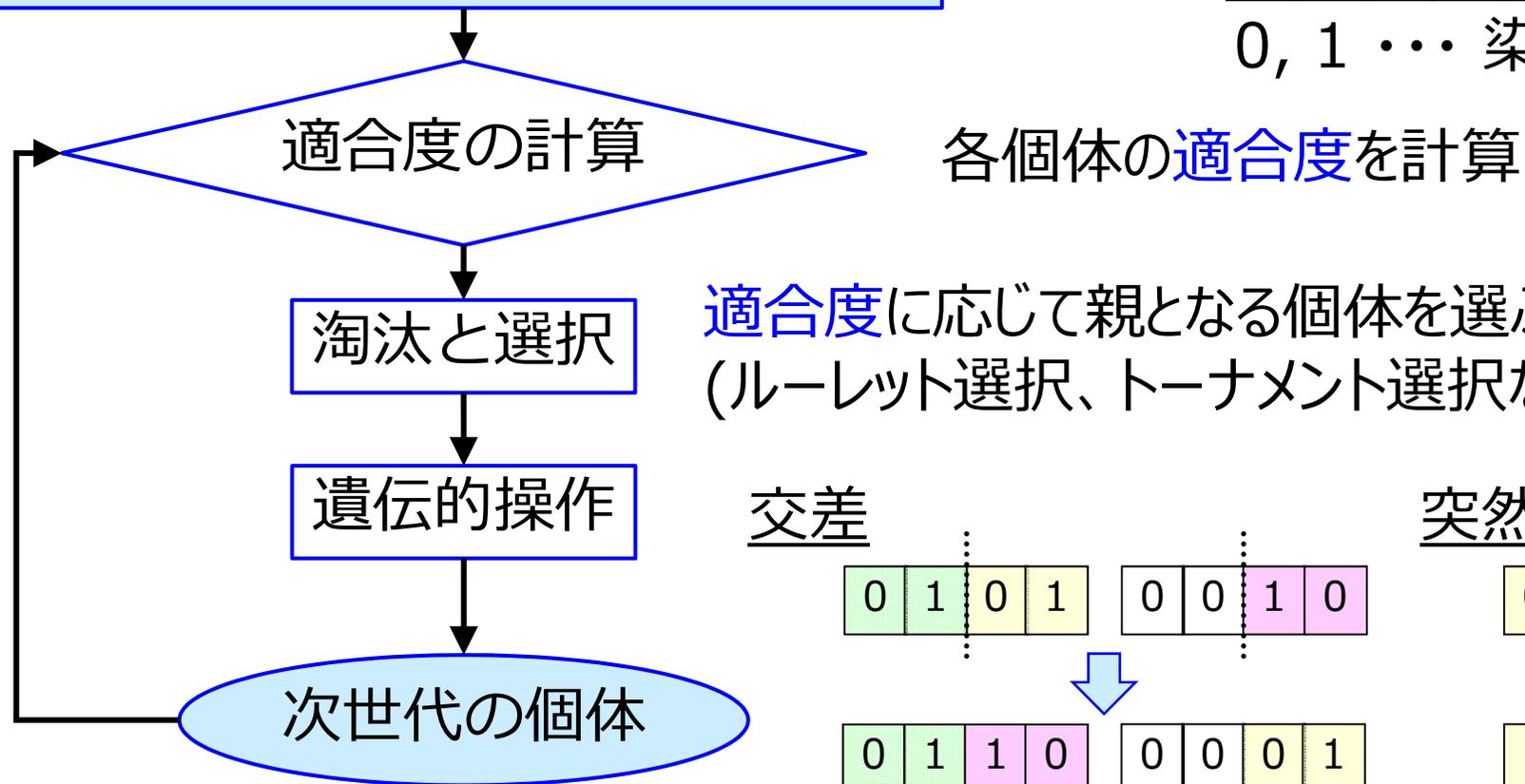
生物の遺伝の様子を模倣した、最適化のためのアルゴリズム

ランダムに初期化された個体の群

各個体

0	1	0	1	1	...
---	---	---	---	---	-----

0, 1 ... 染色体

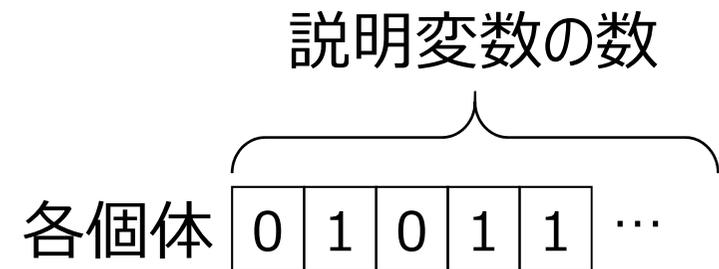


適合度の大きな個体を効率的に得ることができる

GAPLS 個体の表現方法

✓ 個体の染色体である 1, 0 を、

- 1 : 選択された説明変数
- 0 : 選択されない説明変数



とする

- ✓ 個体が実数の GA では、それぞれ 0 から 1 の間の実数で表し、たとえば “0.5” より大きい染色体の番号を、選択された説明変数とする
- ✓ “0.5” の値を変更することで、選択される説明変数の数を調整できる
- ✓ 染色体の 0 ~ 1 の数が大きい順に m 個の説明変数を選択すれば、選択される説明変数の数を決められる

GAPLS 適合度

選択された説明変数のみを用いて、目的変数との間で
PLS によるクロスバリデーションを行う

最適成分数におけるクロスバリデーション推定値を用いた r^2 を
適合度とする

クロスバリデーションについてはこちら：

<https://atachemeng.com/modelvalidation/>

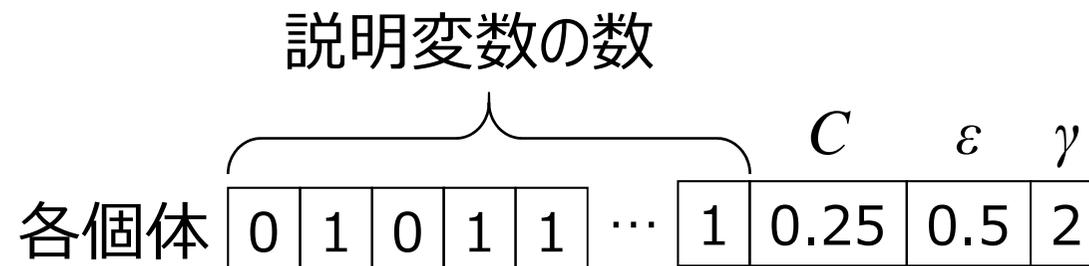
GASVR 個体の表現方法 1/2

✓ 個体の前半の染色体である $1, 0$ を、

- 1 : 選択された説明変数
- 0 : 選択されない説明変数

とする

✓ 個体の後半の染色体を、 C, ε, γ とする



GASVR 個体の表現方法 2/2

- ✓ 個体の実数の GA では、それぞれ 0 から 1 の間の実数で表し、たとえば “0.5” より大きい染色体の番号を、選択された説明変数とする
- ✓ “0.5” の値を変更することで、選択される説明変数の数を調整できる
- ✓ 染色体の 0 ~ 1 の数が大きい順に m 個の説明変数を選択すれば、選択される説明変数の数を決められる

GASVR 適合度

選択された説明変数のみを用いて、目的変数との間で
与えられた C, ε, γ で SVR によるクロスバリデーションを行う

最適成分数におけるクロスバリデーション推定値を用いた r^2 を
適合度とする

クロスバリデーションについてはこちら：

<https://atachemeng.com/modelvalidation/>

どうやって実際にGAPLS, GASVRを実行するか？⁹

- ✓ scikit-learn や DEAP を用いて、GAPLS や GASVR のデモを行うプログラムを作成しました

https://github.com/hkaneko1985/gapls_gasvr

- ✓ランダム性があるため、GAPLS, GASVR を行った結果、いつも同じ結果が得られるとは限らない
- ✓クロスバリデーションをして推定した結果がよくなるように変数を選択しているが、モデルがオーバーフィッティングする可能性もある