

寝ててもヌルヌル頭に入って きそうなkNN説明スライド

神田拓実

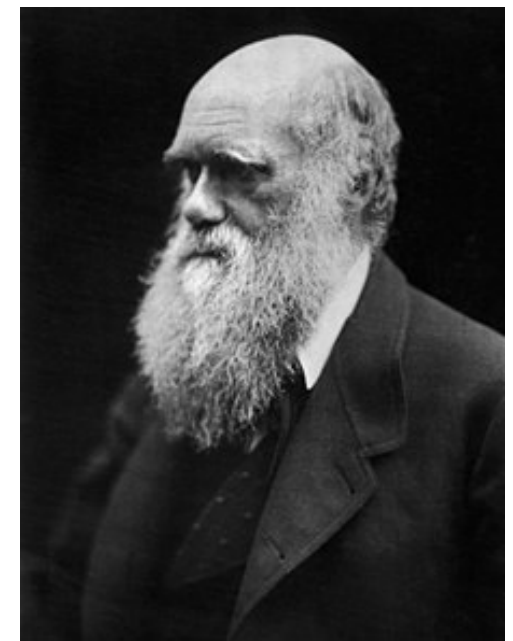
kNN 法とは？

- k-Nearest Neighbor algorithm (k近傍法)
- 「近いデータに多数決を取らせる」法
- **パターン認識**でよく用いられる（音声認識、顔認識）
- 機械学習アルゴリズムの中で最も寝ててもわかる



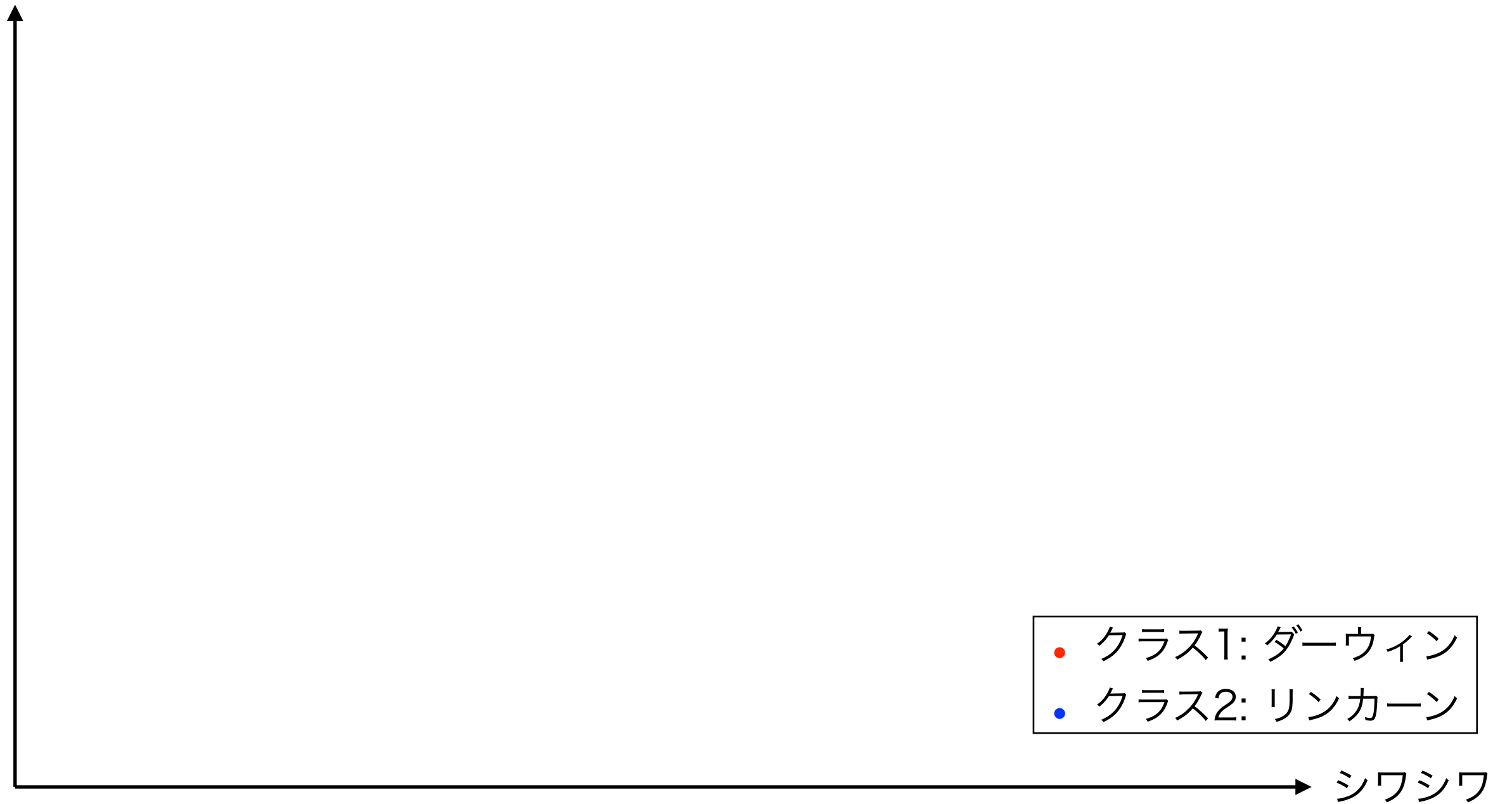
もう帰れます

寝てたけどわかった



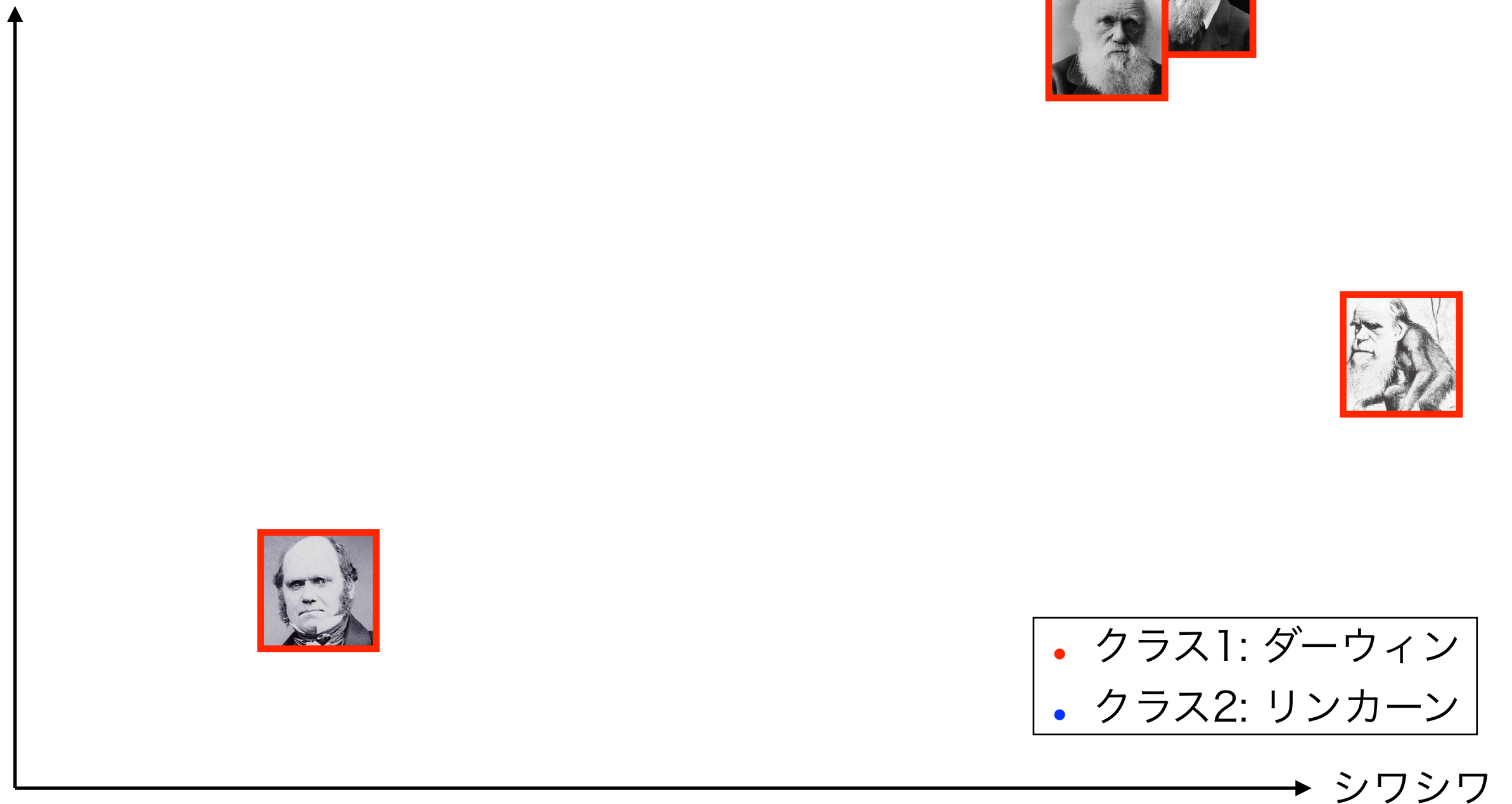
脳みそが進化する図

ヒゲがすごい



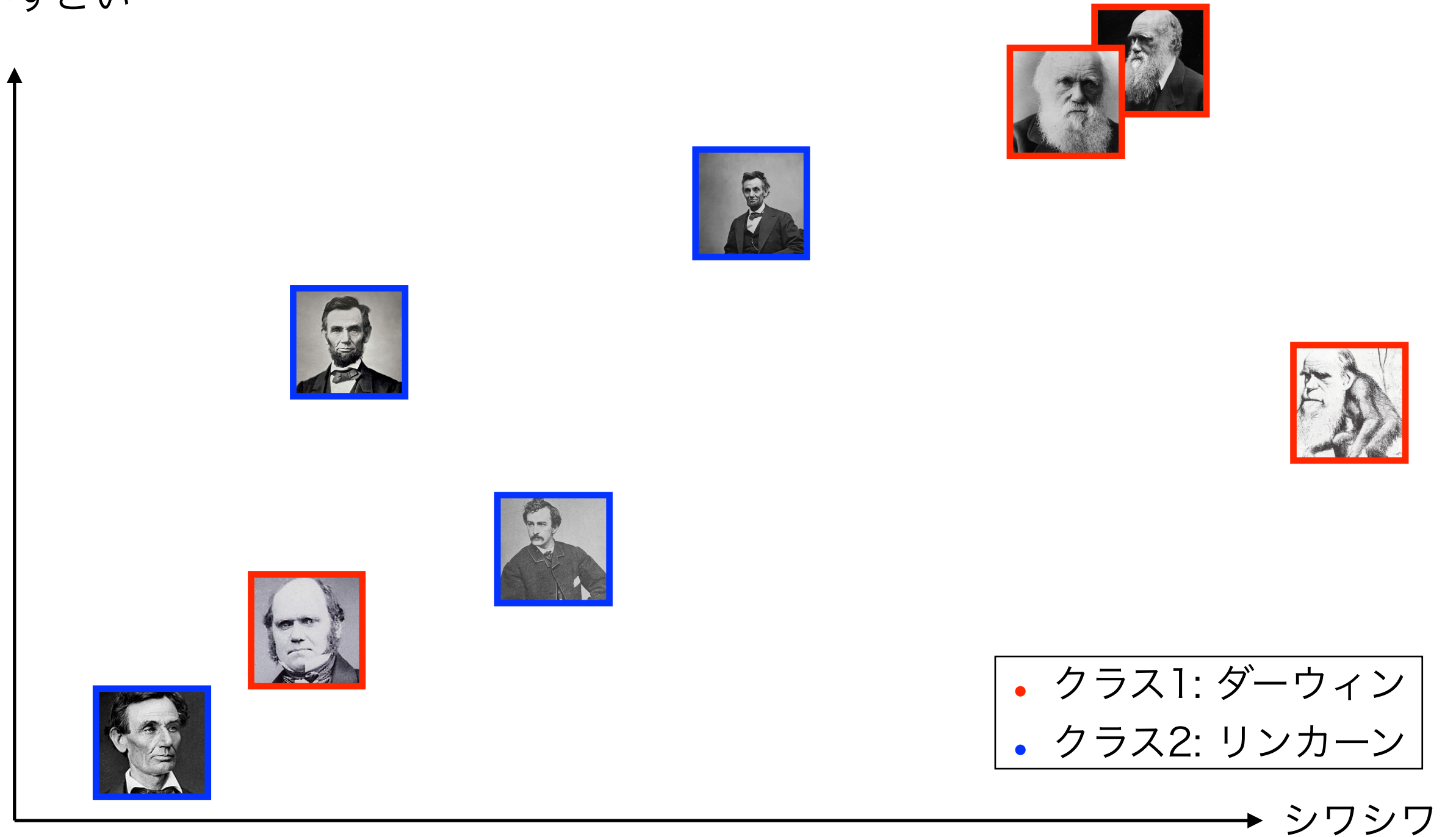
脳みそが進化する図

ヒゲがすごい



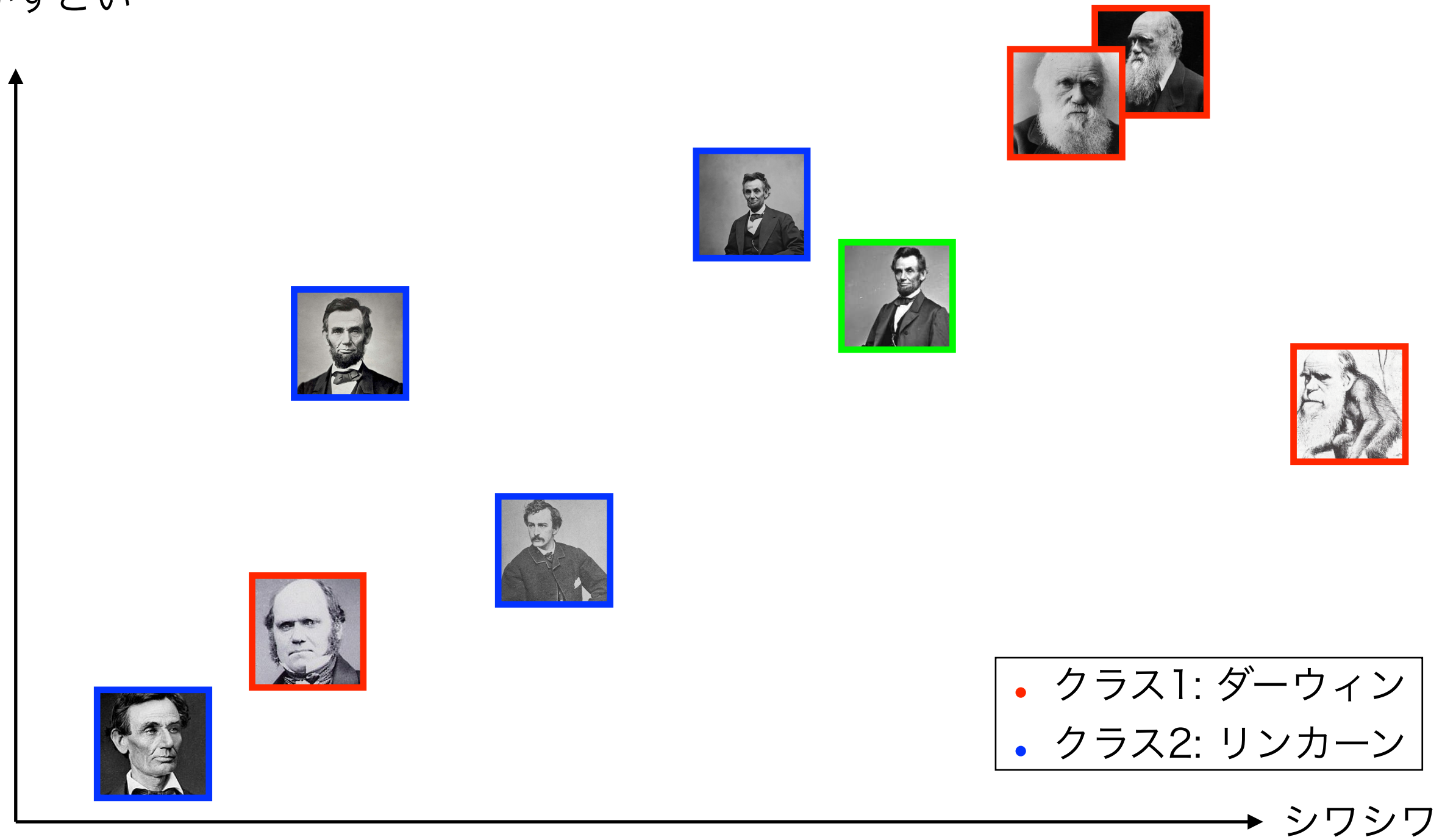
脳みそが進化する図

ヒゲがすごい



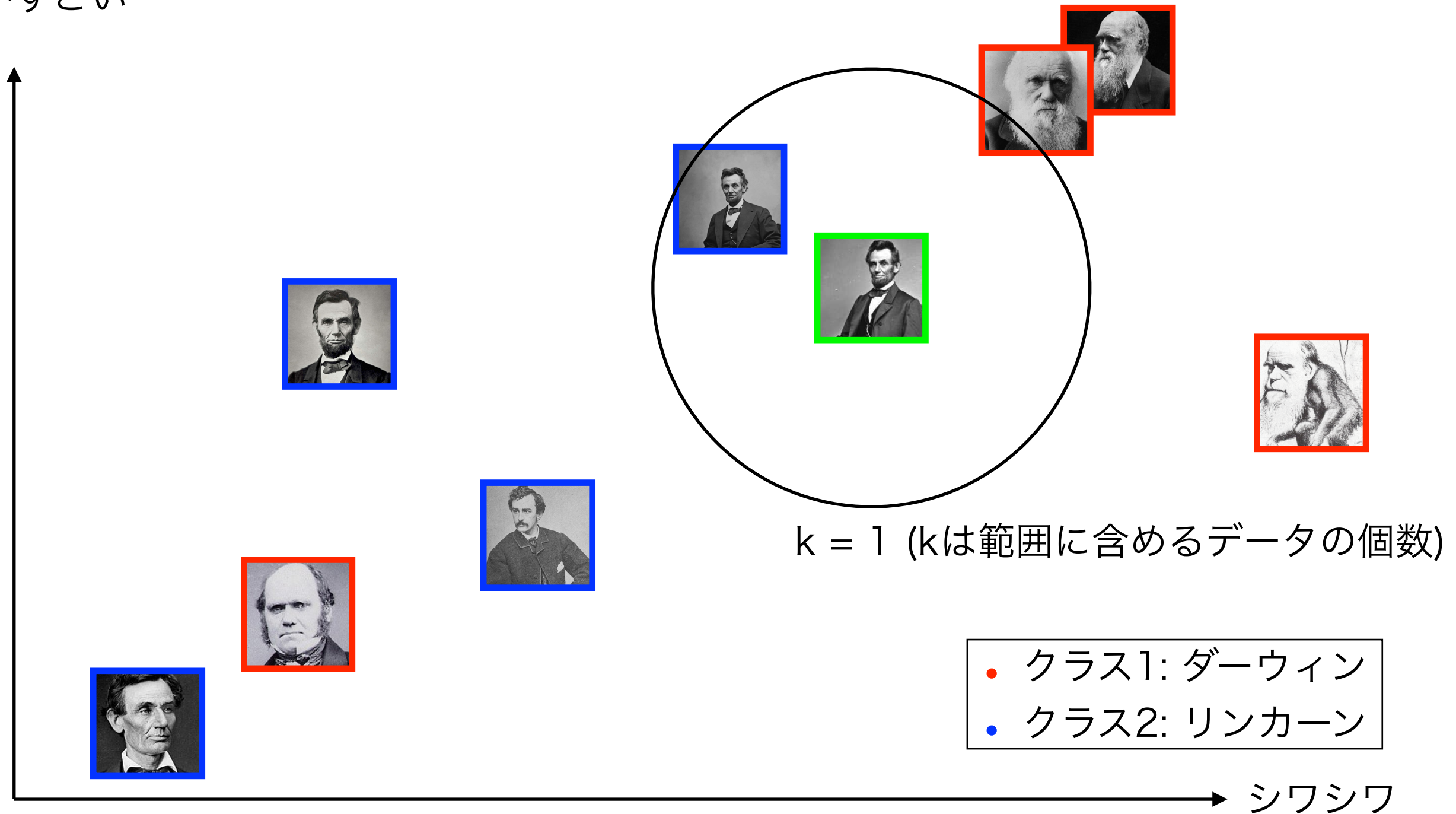
脳みそが進化する図

ヒゲがすごい



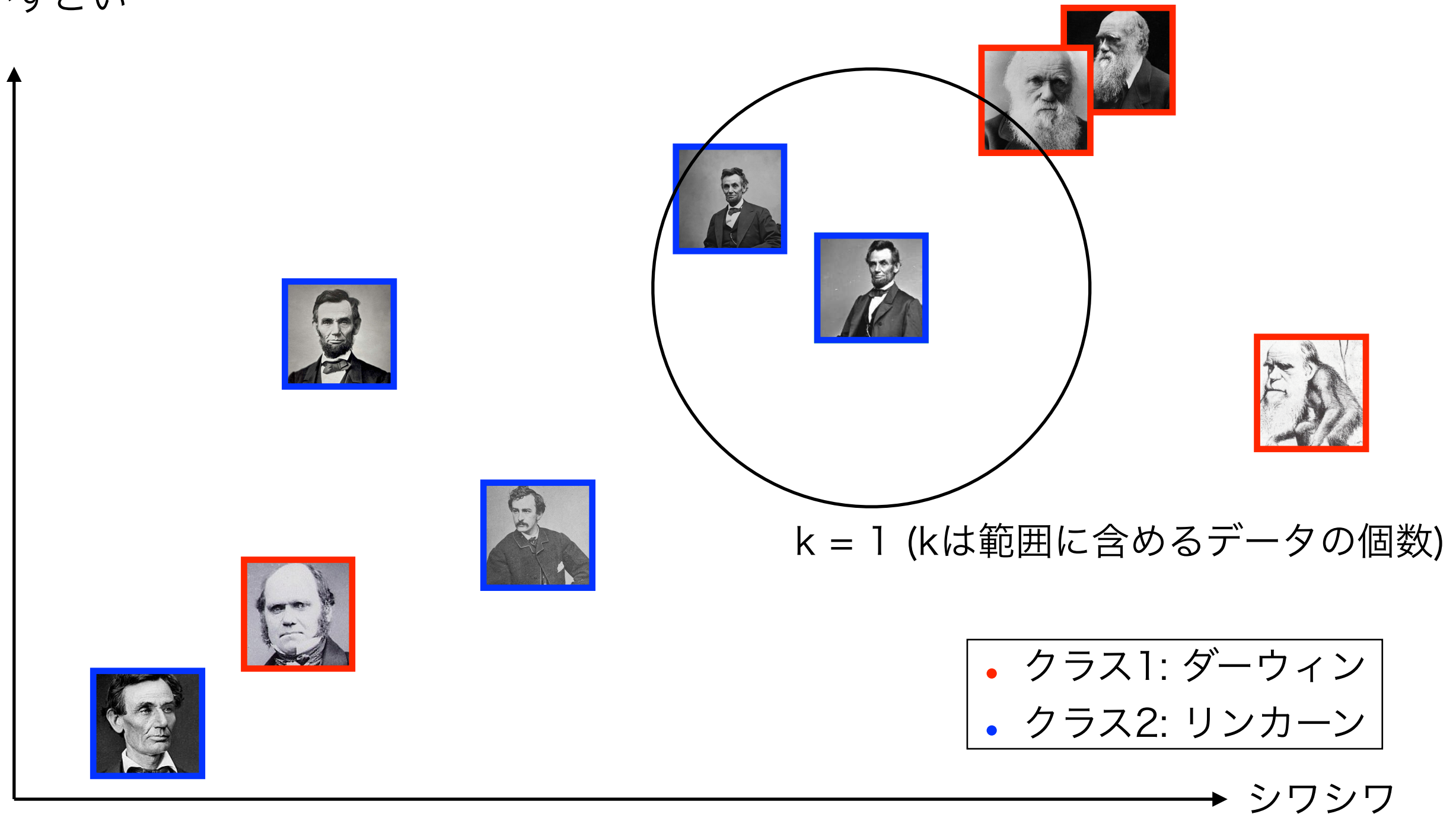
脳みそが進化する図

ヒゲがすごい



脳みそが進化する図

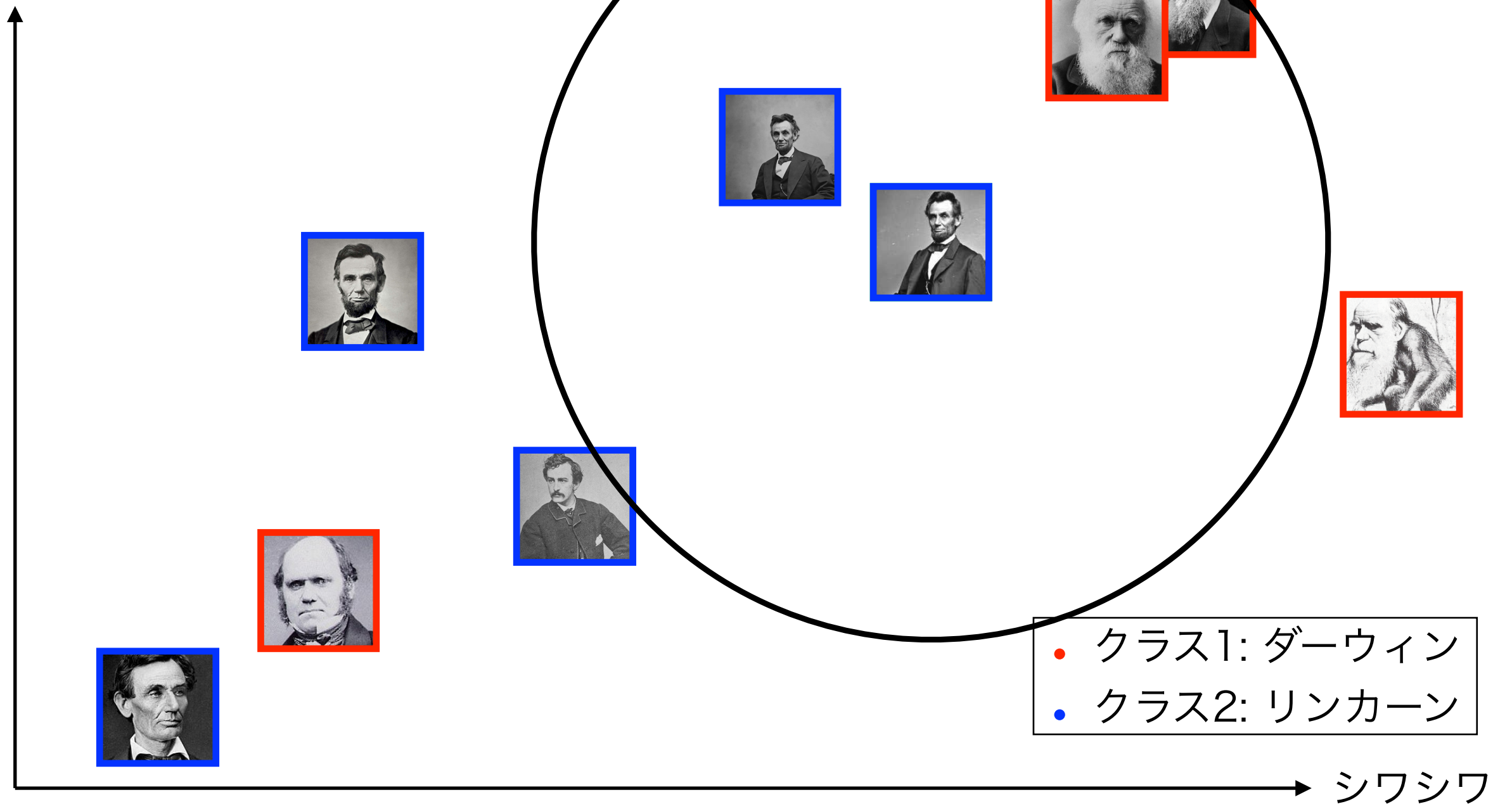
ヒゲがすごい



脳みそが進化する図

ヒゲがすごい

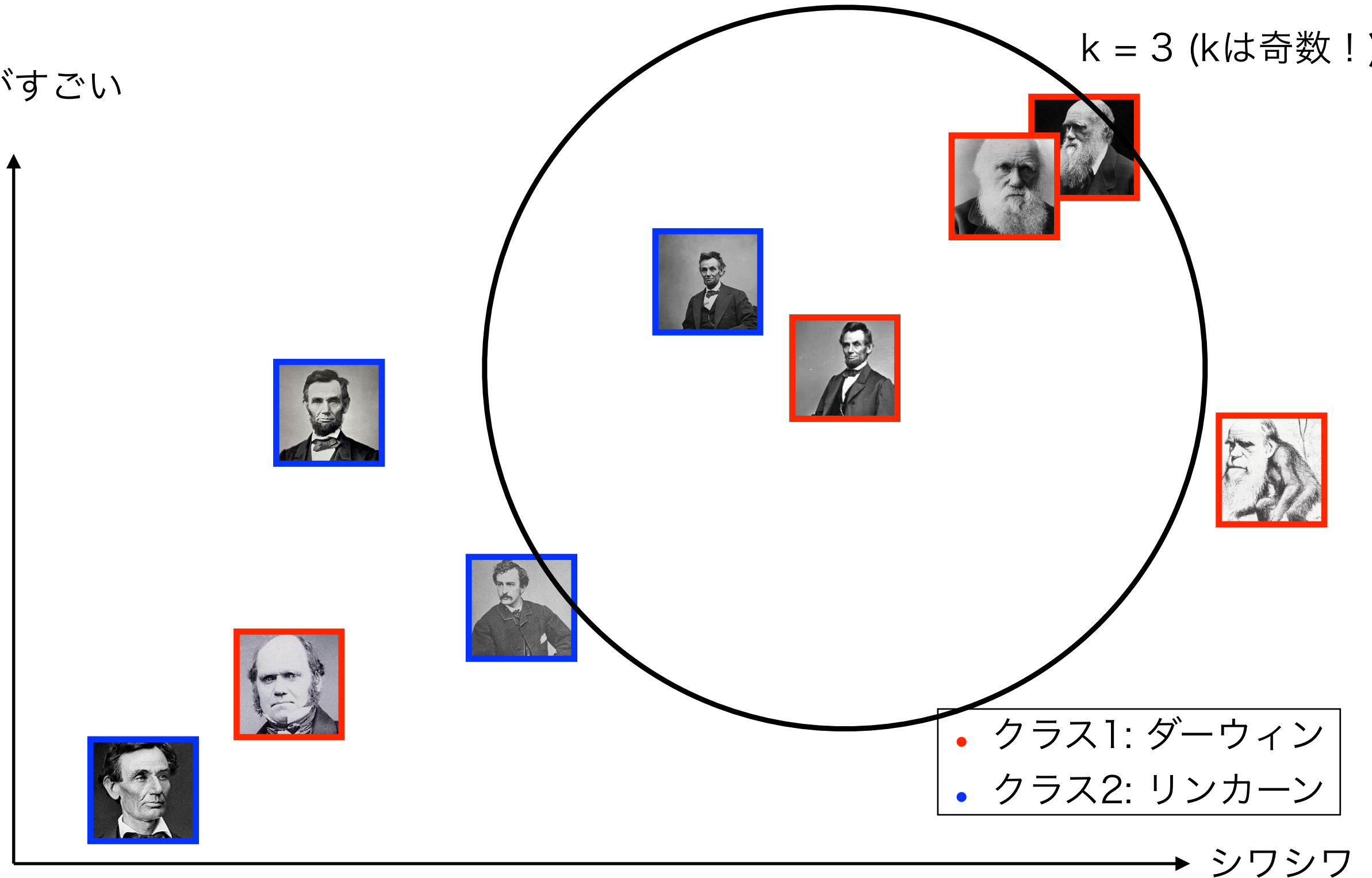
$k = 3$ (k は奇数！)



脳みそが進化する図

ヒゲがすごい

$k = 3$ (k は奇数!)

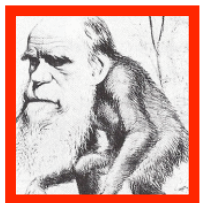
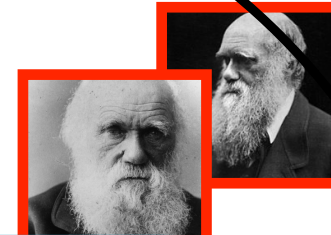
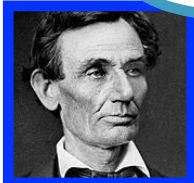


脳みそが進化する図

ヒゲがすごい

$k = 3$ (k は奇数！)

適切な k を選択しなければならない
二値分類の場合は奇数
多値分類の場合は
多数決が取れるよう調整



- クラス1: ダーウィン
- クラス2: リンカーン

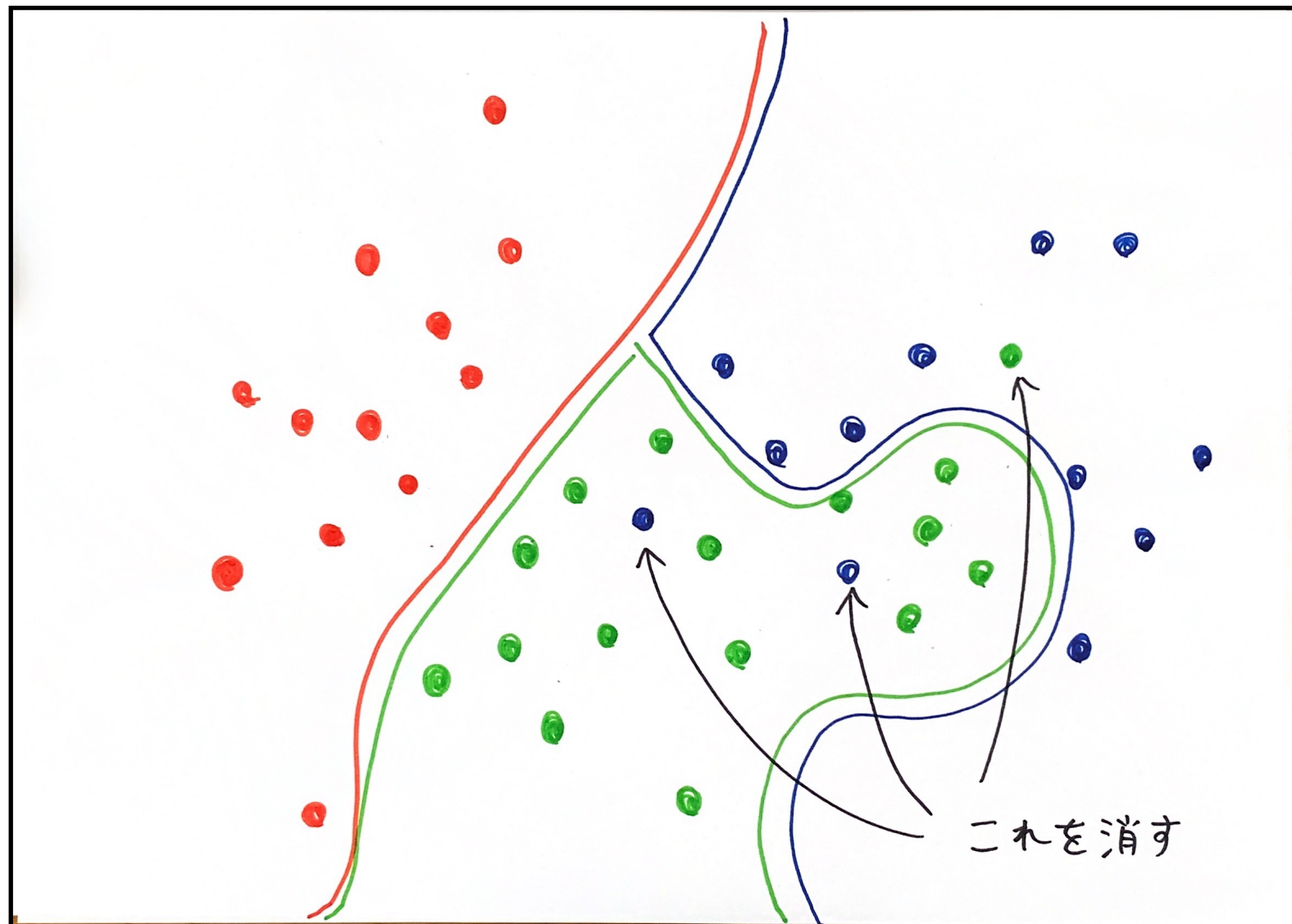
シワシワ

kNN 法の弱点

- **処理が重い**（分類したいデータ毎に学習データを全探索するから当たり前）
 - 改善手法あるよ
 - 誤り削除型kNN
 - 圧縮型kNN
 - 分岐限定法
 - 近似最近傍探索
- **次元の呪いの影響を受ける**

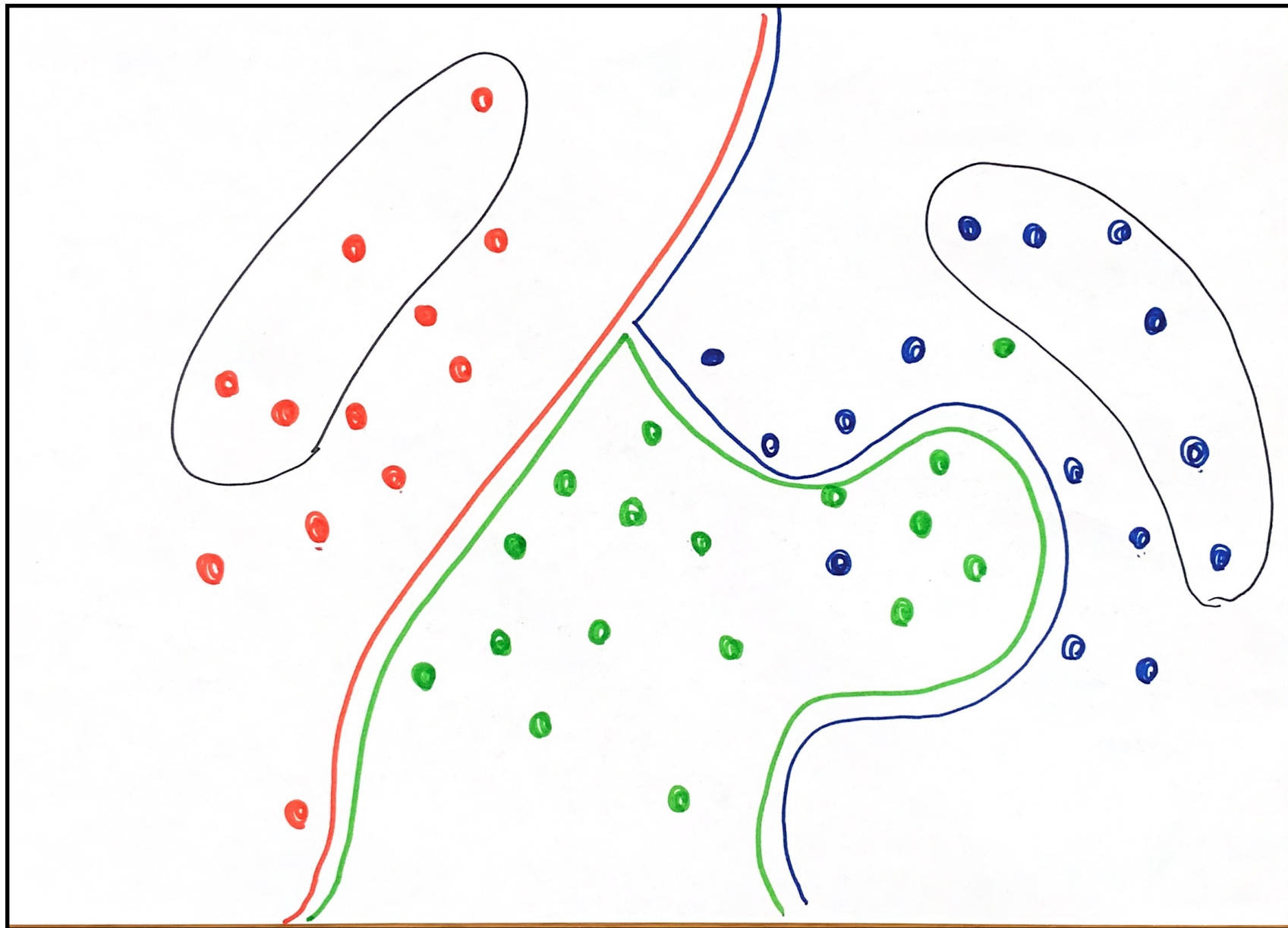
誤り削除型 kNN 法

- 識別境界を作成した後、不正解の識別境界に含まれる教師データを削除する



圧縮型 kNN

- 識別に関係ない教師データを削除する

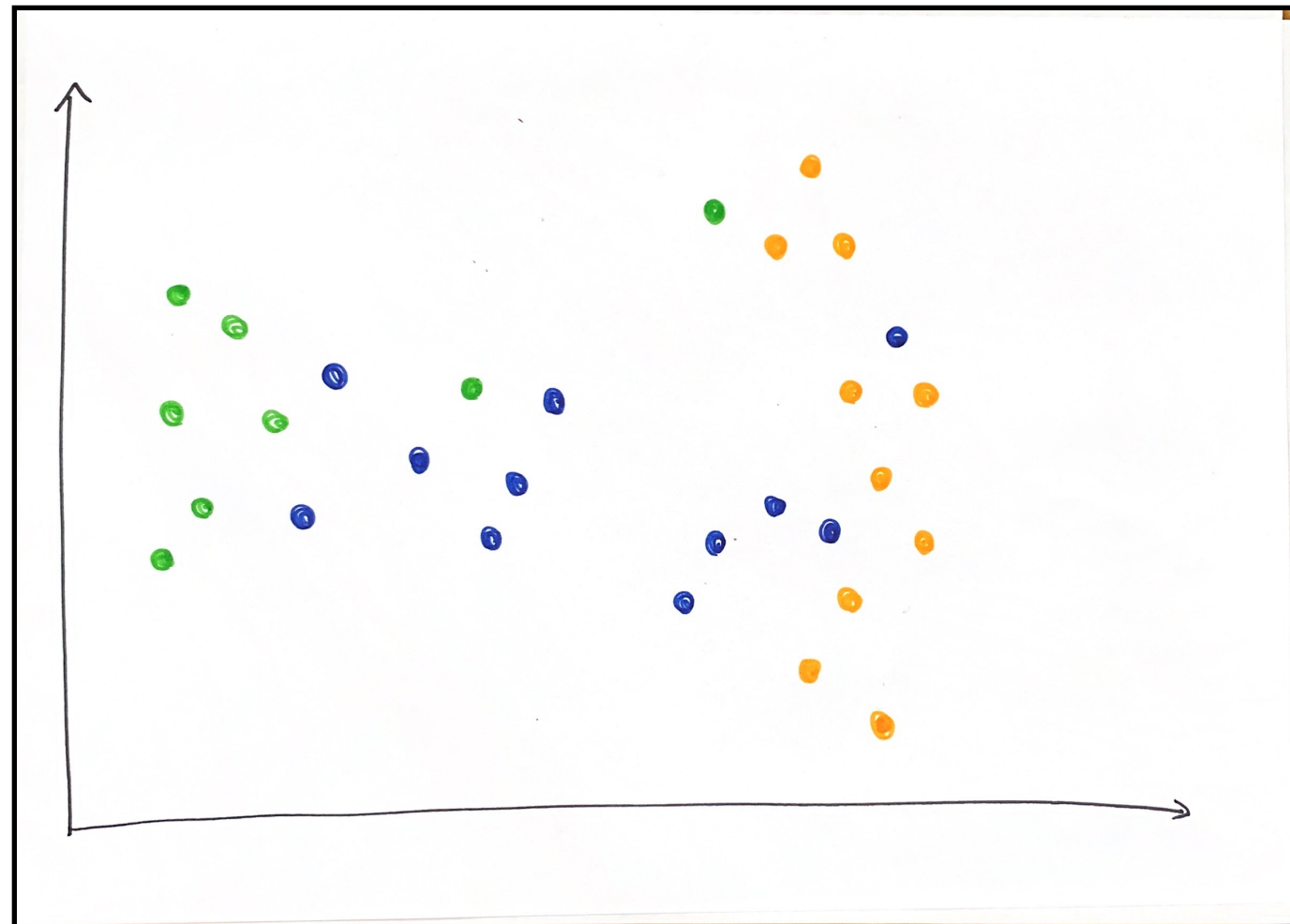


分岐限定法

- 分岐法と限定法を用いて効率的に探索を行う
 - 分岐法：クラスタリングによって**木構造のような形態に組織化**する
 - 限定法：木構造をもとに**範囲を限定**する

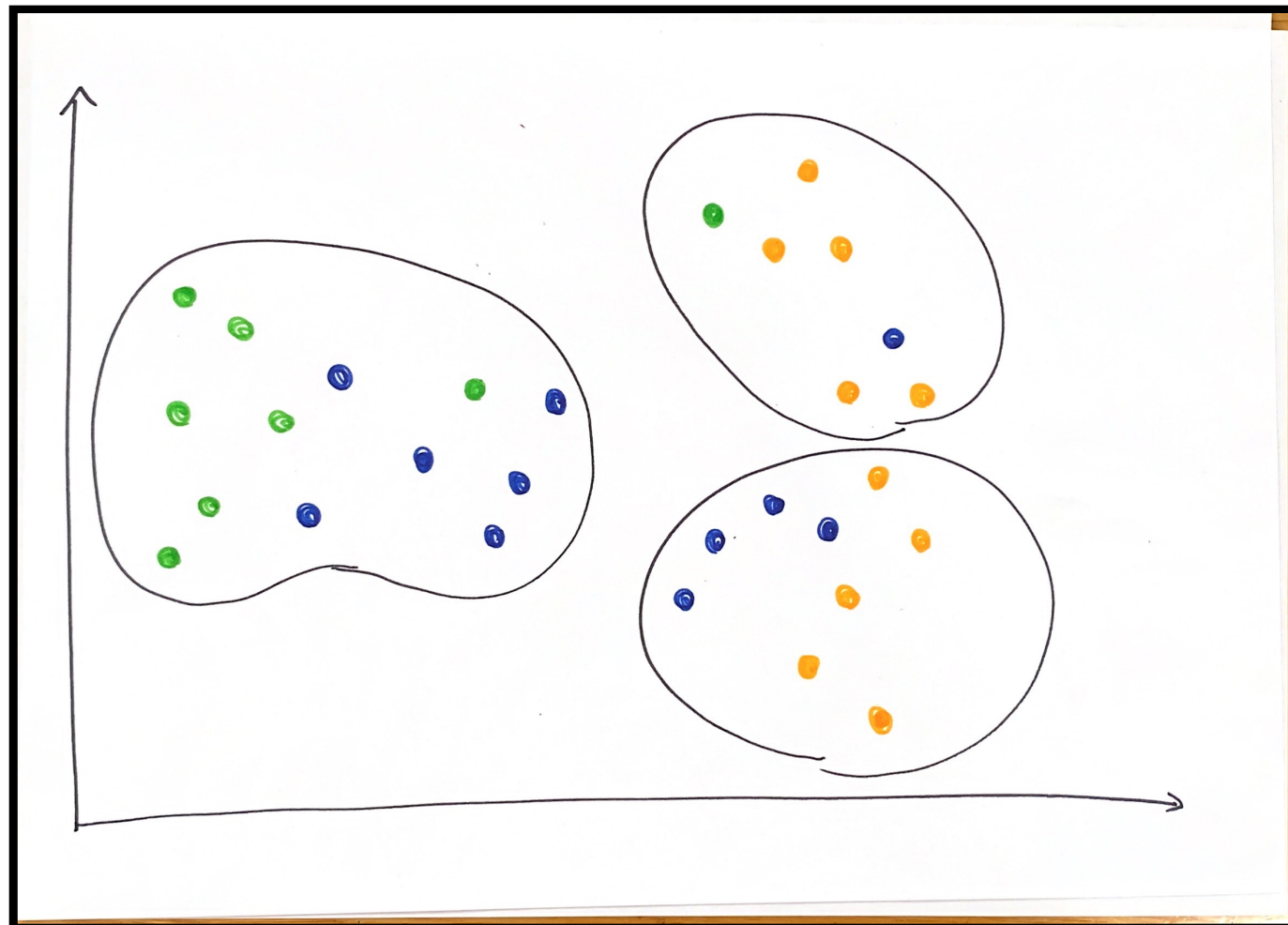
分岐限定法

- 分岐法と限定法を用いて効率的に探索を行う
 - 分岐法：クラスタリングによって**木構造のような形態に組織化**する
 - 限定法：木構造をもとに**範囲を限定**する



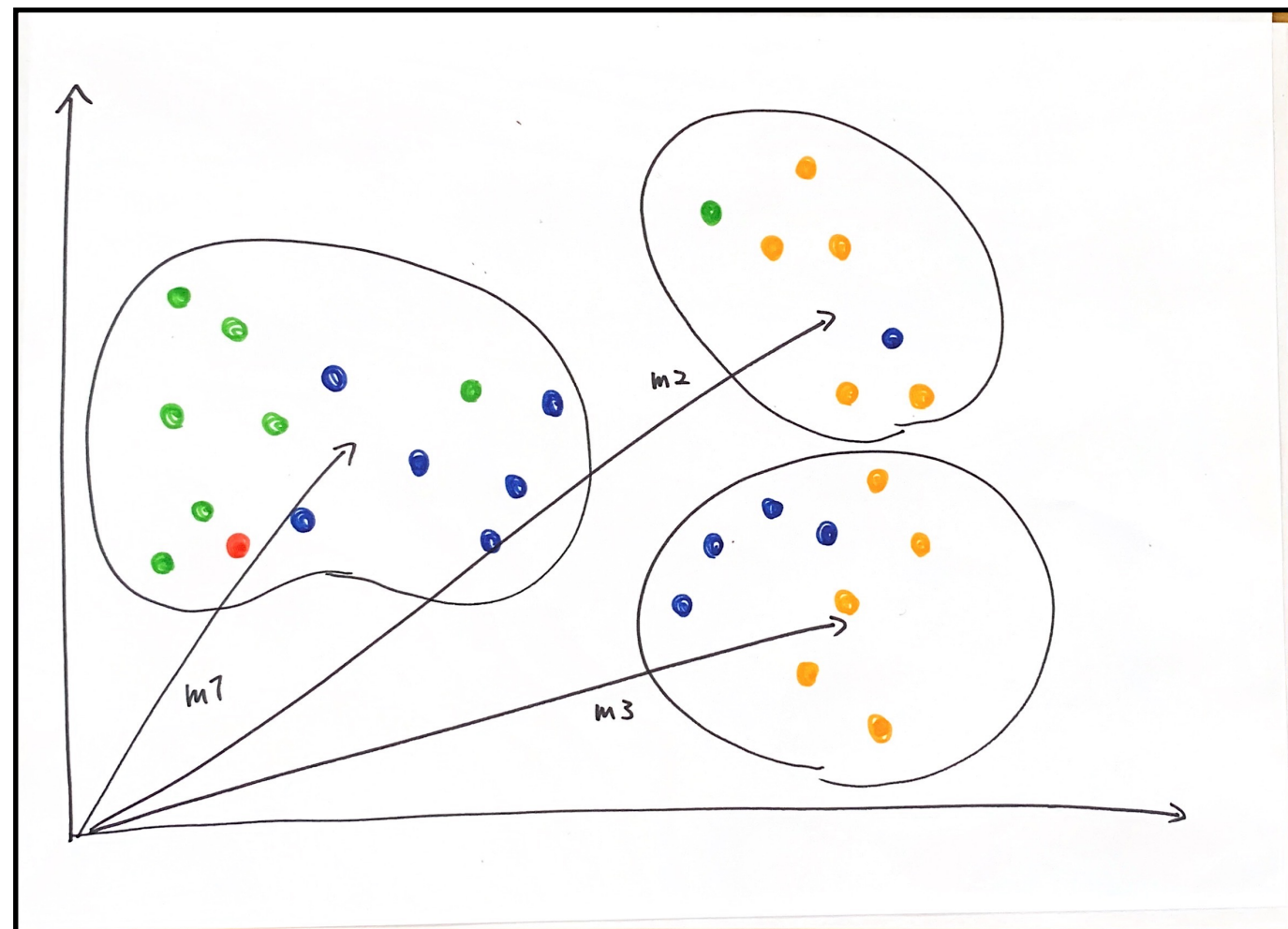
分岐限定法

- 分岐法と限定法を用いて効率的に探索を行う
 - 分岐法：クラスタリングによって**木構造のような形態に組織化**する
 - 限定法：木構造をもとに**範囲を限定**する



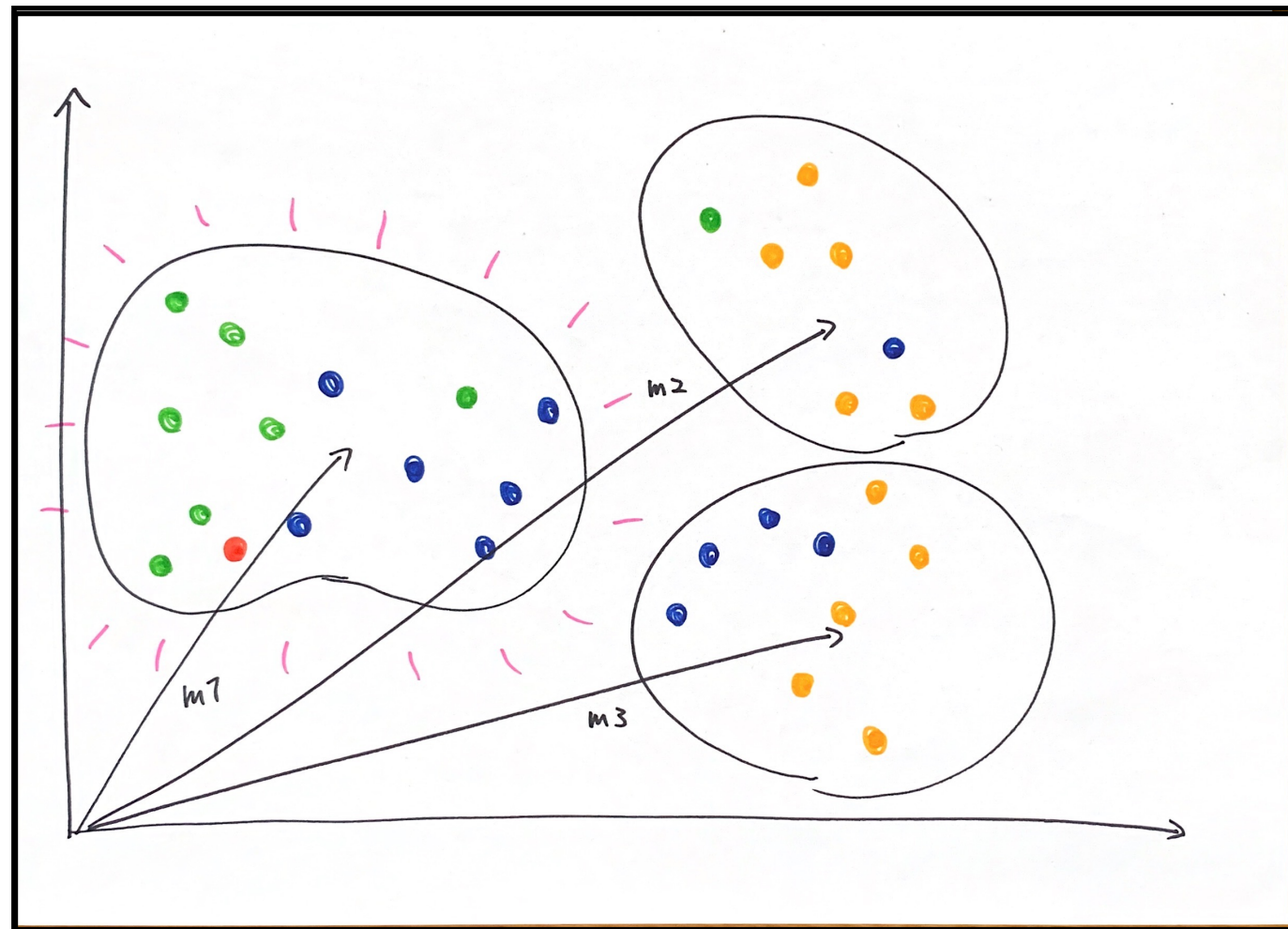
分岐限定法

- 分岐法と限定法を用いて効率的に探索を行う
 - 分岐法：クラスタリングによって木構造のような形態に組織化する
 - 限定法：木構造をもとに範囲を限定する



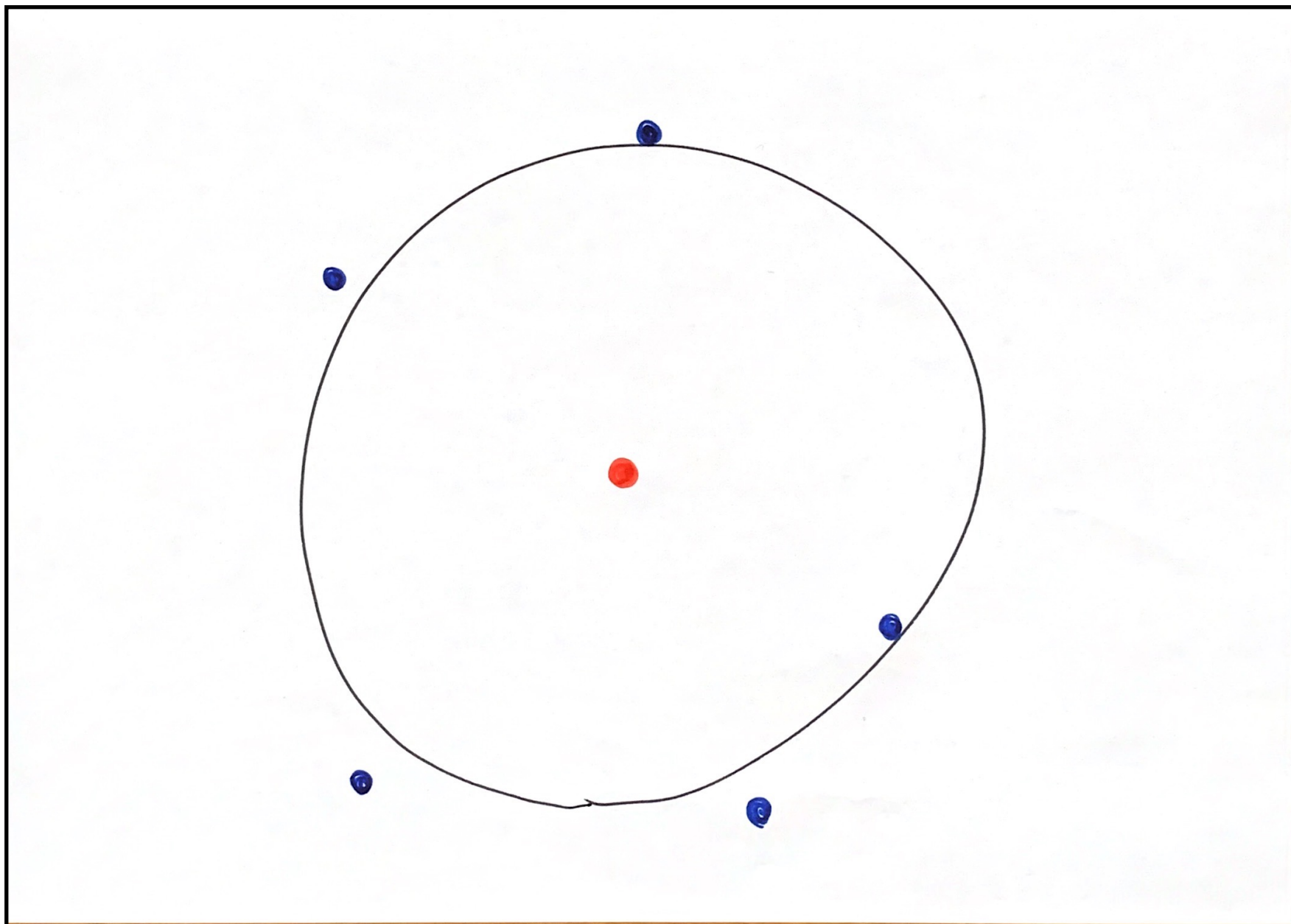
分岐限定法

- 分岐法と限定法を用いて効率的に探索を行う
 - 分岐法：クラスタリングによって木構造のような形態に組織化する
 - 限定法：木構造をもとに範囲を限定する




近似最近傍探索

- 最近傍より少し距離が遠くても許容する



次元の呪い

- 次元が大きくなりすぎるとデータ間の距離を測る時に「どのデータとの距離も大きく、同じようになってしまう」という問題
- サクサクメロンパン問題 (<https://bit.ly/2XWHzOC>)

- kNN は次元数の少ないデータに対して利用するケースが多い