

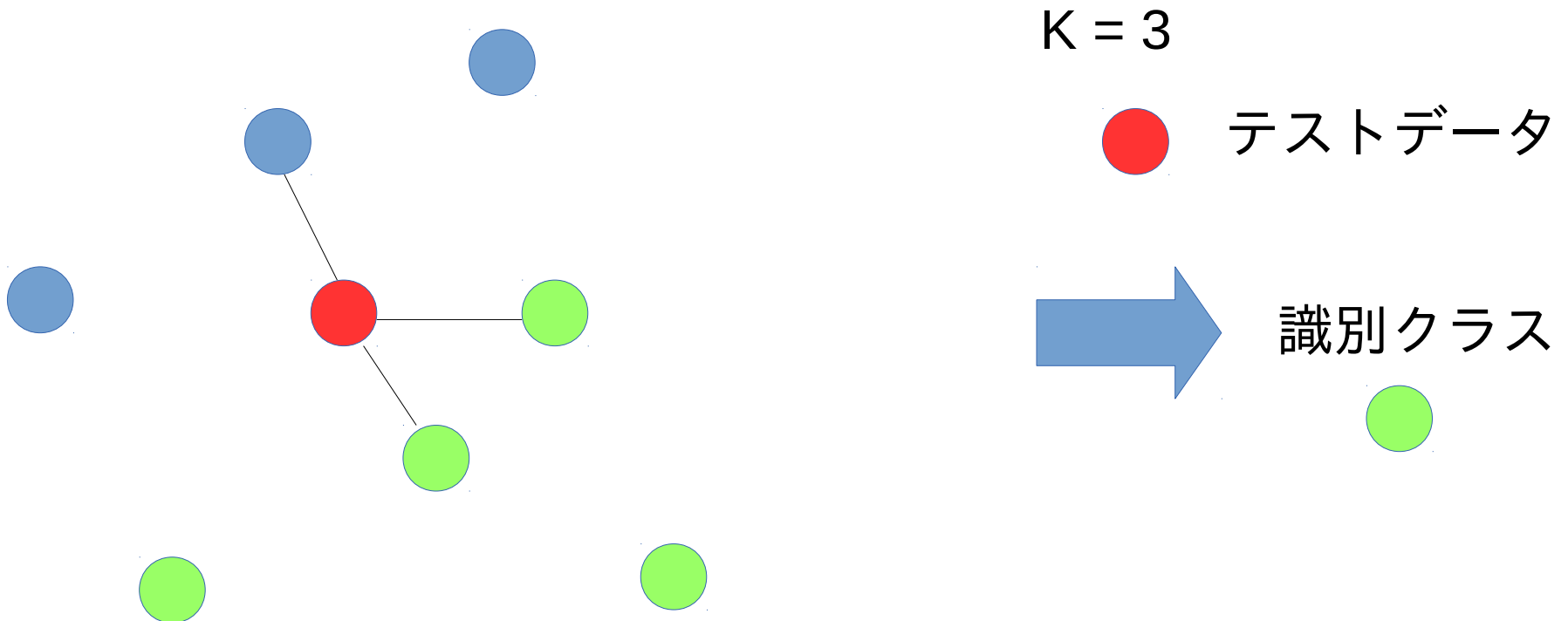
Scikit-learn ゼミ

1.4.2 Nearest Neighbors Classification

新納浩幸

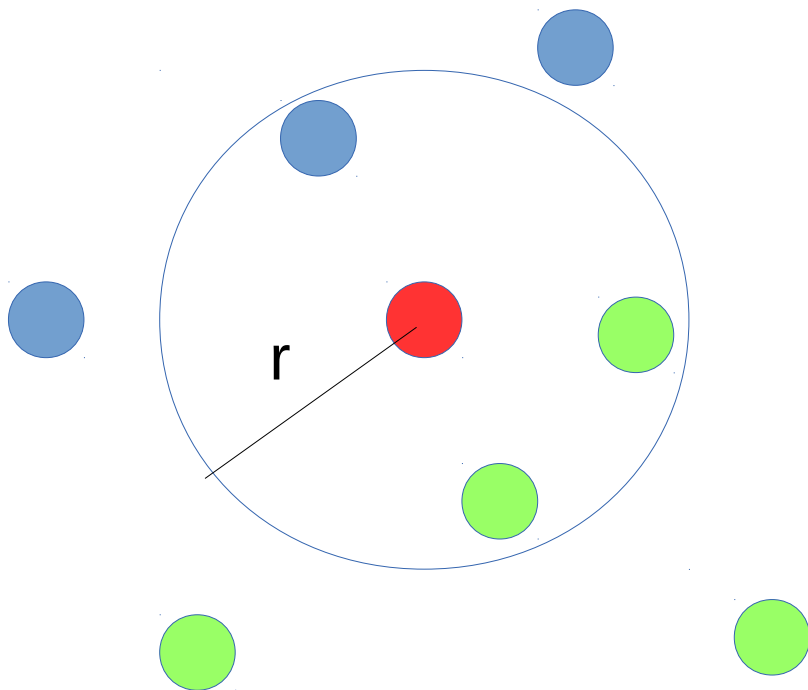
K-NN 法

識別対象のテストデータに対して、
距離の近い順に k 個の訓練データを選択
それらのラベルの多数決で識別クラスを決める



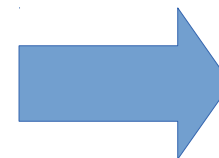
K-NN 法 (2)

識別対象のテストデータを中心にして、
半径 r の円内の訓練データを選択
それらのラベルの多数決で識別クラスを決める



$K = 3$

● テストデータ



識別クラス



K-NN 法の特徴

- ・ K の適切な値はデータに依存、大きければノイズが減るが、識別の境界がゆるくなる
- ・ 訓練データが一様でないときは、半径 r の方の k-NN 法の方が better
- ・ k-NN 法は「次元の呪い」の影響が小さい
- ・ 選択したデータに重みをつけるやりかたもある

scikit-learn の提供関数

KNeighborsClassifier

近いもの k 個

RadiusNeighborsClassifier

半径 r 内のデータ