

1 判別基準評価

1.1 (a)

Nearest Neighbor、ユークリッド距離、類似度および重みつきユークリッド距離法 (重み: (1,5)) を用いてテストデータ A を判別し, その認識正誤表と認識率を表 1 に示す.

表 1 方法別認識正誤表

方法	正誤	認識率 [%]
Nearest	TTTTTTTTTTTTFTTTTTTFTFTTFTTTFTTTTTTTTTTFFF	80.0
ユークリッド距離	FTTTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTT	77.5
重み付きユークリッド	FTTTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTT	77.5
類似度	FTTTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTT	77.5

1.2 (b)

次に, それぞれの方法で識別した散布図, 学習データ平均と識別境界線を一つにまとめた結果図を順に図 1(a)-(e) に示す.

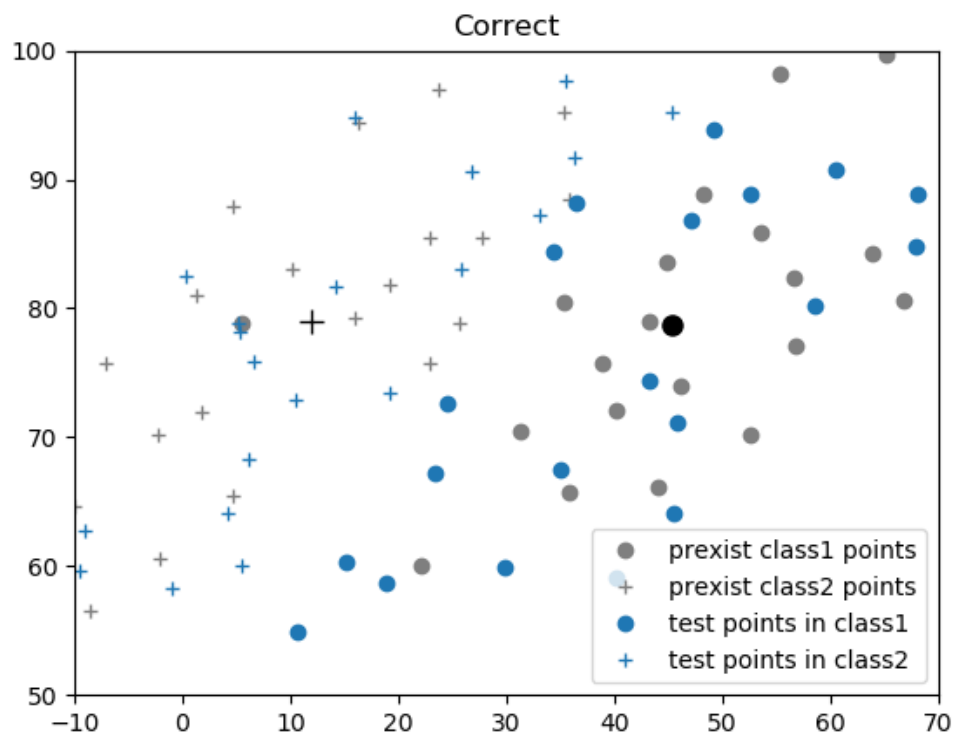
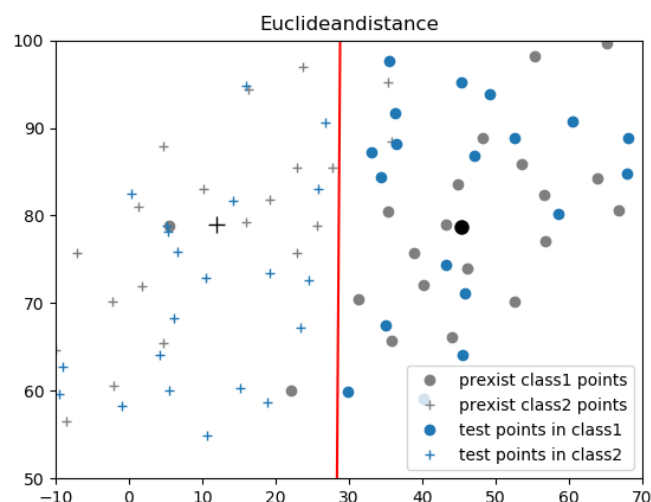


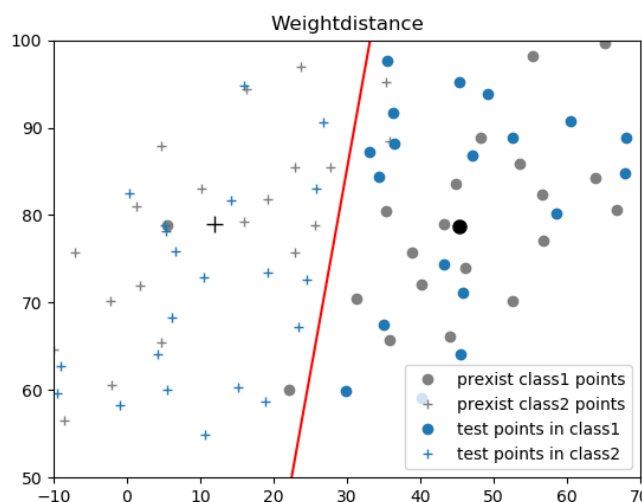
図 1 (a) Correct distribution of test set A.



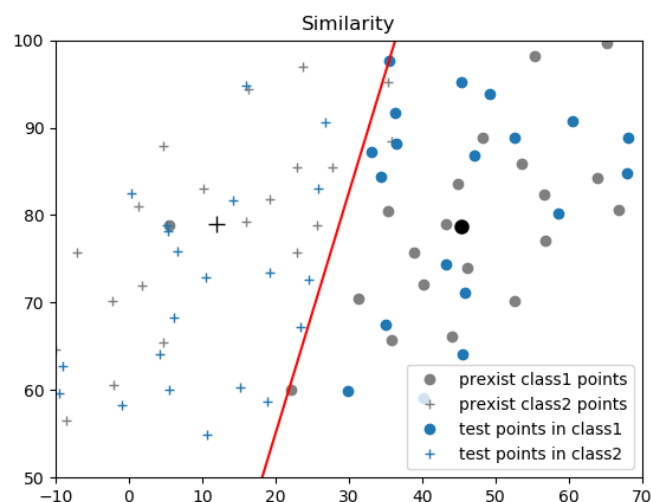
(b) The result of using Nearest Neighbor method.



(b) The result of using EuclideanDistance method.



(d) The result of using Weight Euclidean method.



(e) The result of using Similarity method.

上図から, Nearest Neighbor 法はノイズに弱く, class2 の群れの中に class1 点があっただけでその周りにいる未知点が class1 に判定されてしまう. ユークリッド距離法では各クラス平均の分布状況によって判別基準が大きく左右されることがあることがわかった. 例えば, 図 1(c) に示したように, 今回の学習データ A の class1 と class2 の平均点 (それぞれ大きい黒丸と大きい黒プラス記号で表した) の y 座標がほぼ変わらず, 識別境界線も自明にその中心を結んだものと垂直な線になり, テストデータはほぼ各点の x 座標で分類されたことがわかった.

この問題を改善すべく現れたのが重み付きユークリッド距離法である. 図 1(e) で示したように, この方法ではそれぞれの軸に重みをかけることができ, 今回の学習データ A に対しては平均点がほぼ x 軸に平行な直線上に分布していることを考慮し, 各テストデータの y 座標の影響を引き出すために x 軸に 1, y 軸に 2 の重みをかけた. 類似度判別では, 原点からの角度で判別状況が大きく変わるため, テストデータの分布状況は認識率に大きな影響を与えることが予想できる.

1.3 (c) - データセット B

データセット B, C, D についても上記と同じように分析し結果を下表 2-4 及び図 2-4 に示した.

表 2 方法別認識正誤表 (B)

方法	正誤	認識率 [%]
Nearest	TTTTTTTTTTTTFTTTTTTTTTTTTTFTTTTTTTTTTTTTTT	95.0
ユークリッド距離	TTTTTTTTTTTTFTFTTTFTTTTTTTTTTTTTTTTTTTTTFT	90.0
重み付きユークリッド	FTTFTTFTTTFTTTFTTTFTTTTFFTTFTTTTFTFFTTT	70.0
類似度	TFTTTTTTTTTFTFFTTFFTFFTTTTTTTTFTFTTFTTFTT	70.0

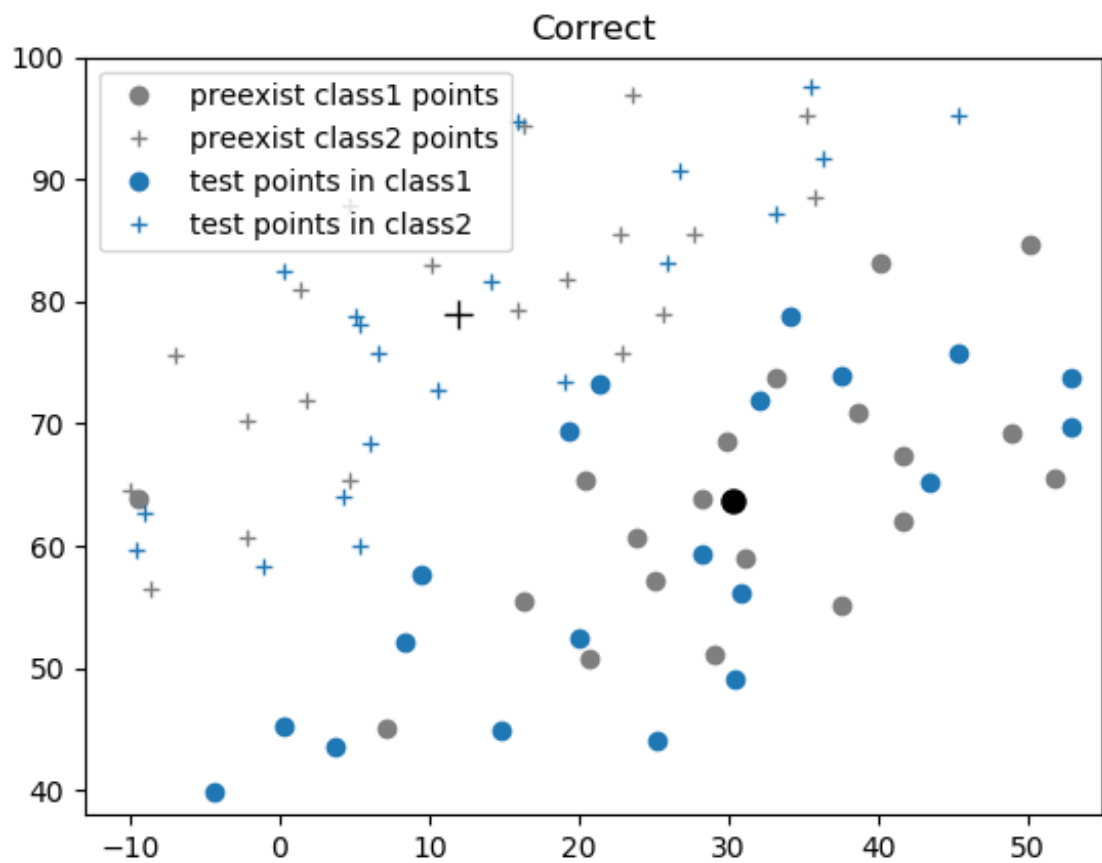
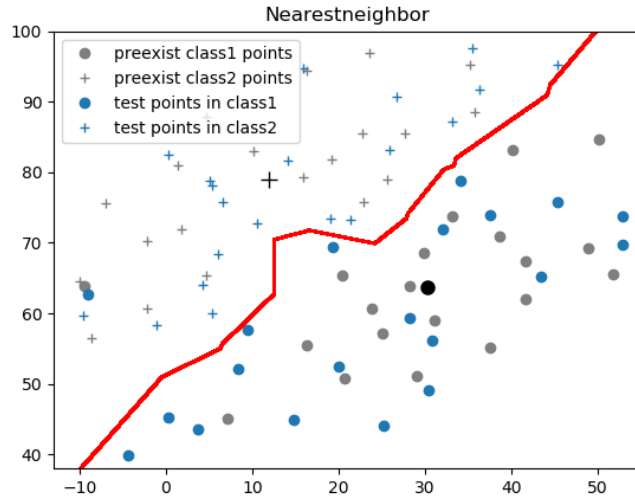
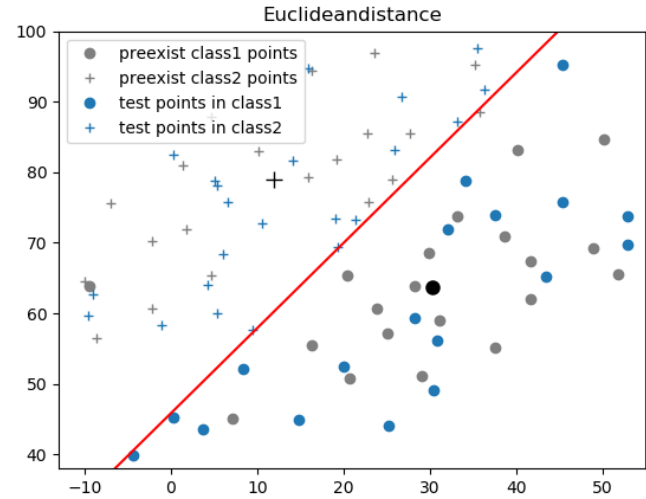


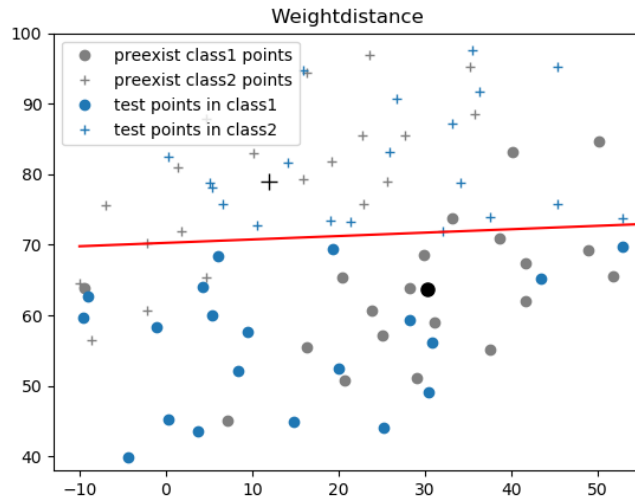
図 2 (a) Correct distribution of test set B.



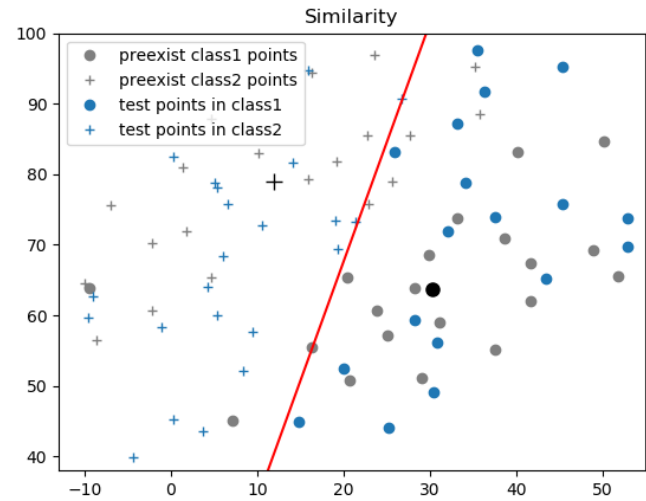
(b) The result of using Nearest Neighbor method. (B)



(b) The result of using EuclideanDistance method.



(d) The result of using Weight Euclidean method. (B)



(e) The result of using Similarity method. (B)

1.2 データセット C

表 3 方法別認識正誤表 (C)

方法	正誤	認識率 [%]
Nearest	TTTTTTTTTTTTFTTTTTTFTFTTFTTTTTTTTTTTFFF	80.0
ユークリッド距離	TFTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTT	77.5
重み付きユークリッド	TFTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTT	77.5
類似度	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTFTFTTTTFTT	92.5

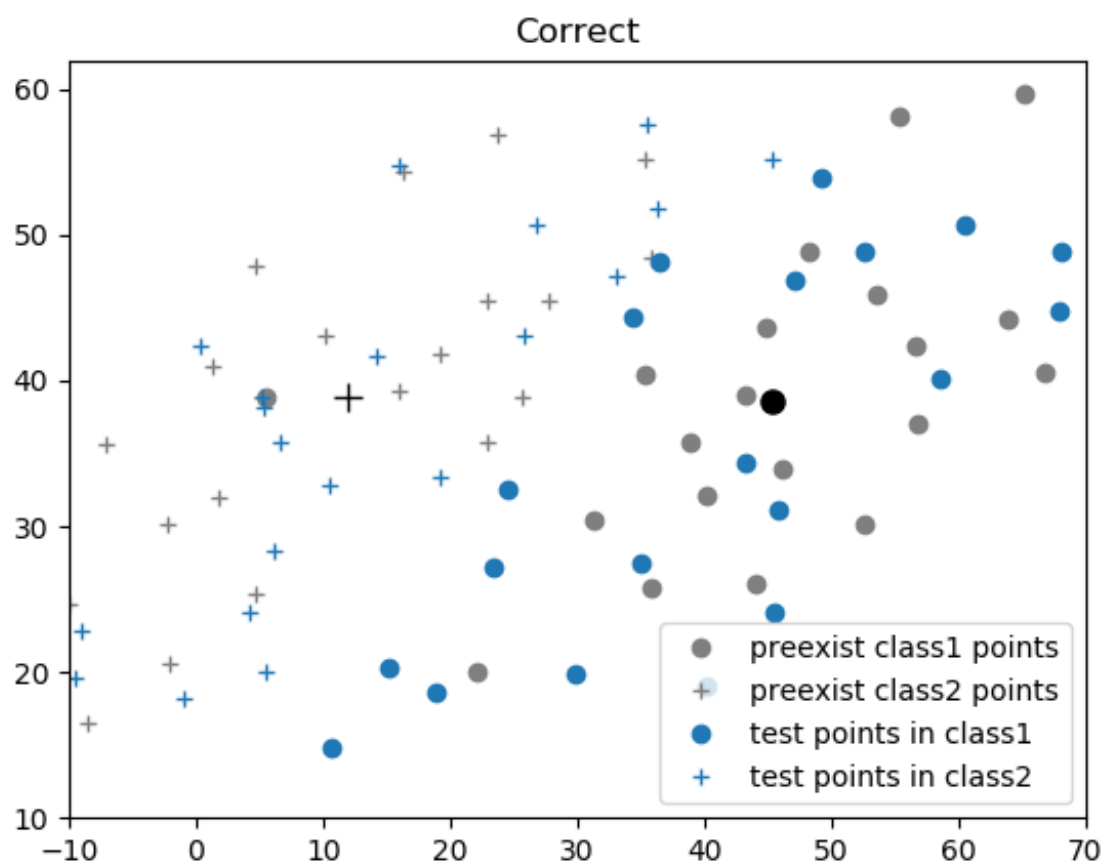
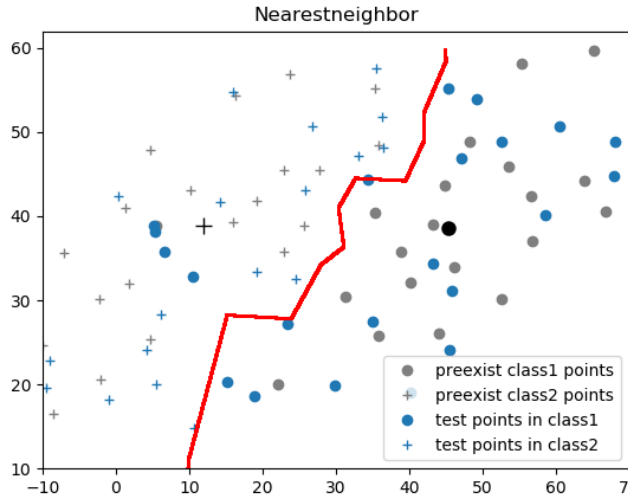
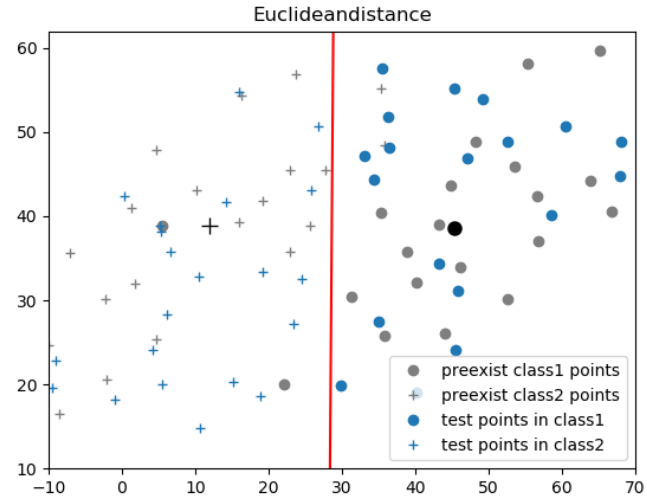


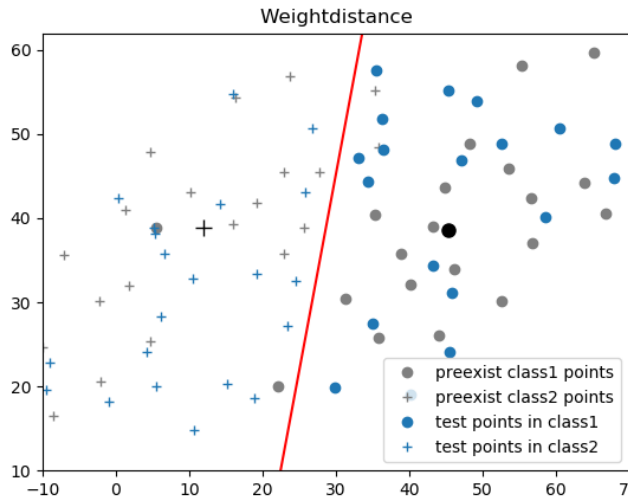
図 3 (a) Correct distribution of test set C.



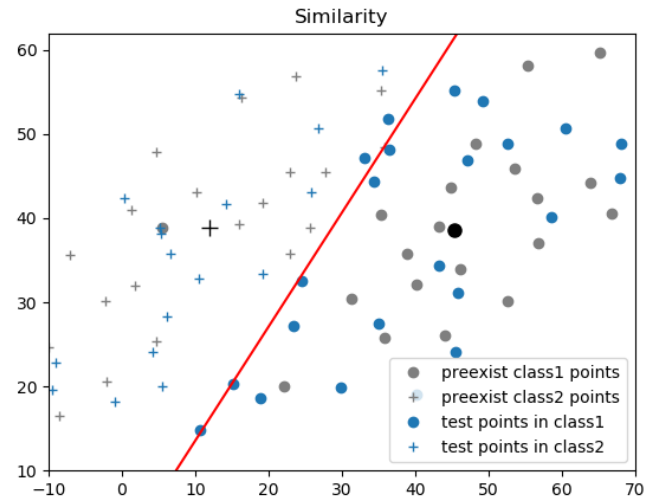
(b) The result of using Nearest Neighbor method. (C)



(b) The result of using EuclideanDistance method.



(d) The result of using Weight Euclidean method. (C)



(e) The result of using Similarity method. (C)

1.2 データセット D

表 4 方法別認識正誤表 (D)

方法	正誤	認識率 [%]
Nearest	TTTTTTTTTTTTFTTTTTTFTFTTTTTTTTTTTTTTTTTFTT	90.0
ユークリッド距離	TFTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTFTT	77.5
重み付きユークリッド	TFTTTTTTTTTTTTTTTFTFTTTTTTTTTTTTTTTTTFTT	90.0
類似度	TFTTTTTTTTTTTFTTFFFTTTTTTTTTTFTFTTFTTFTT	77.5

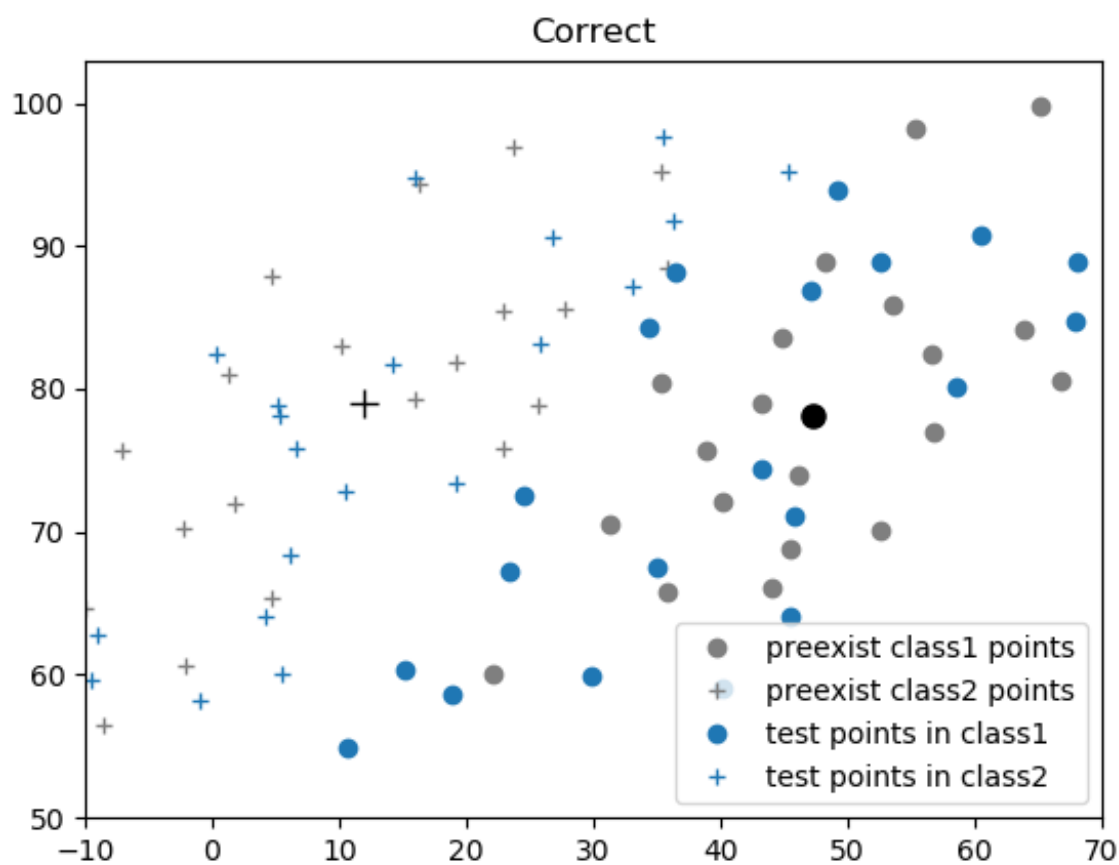
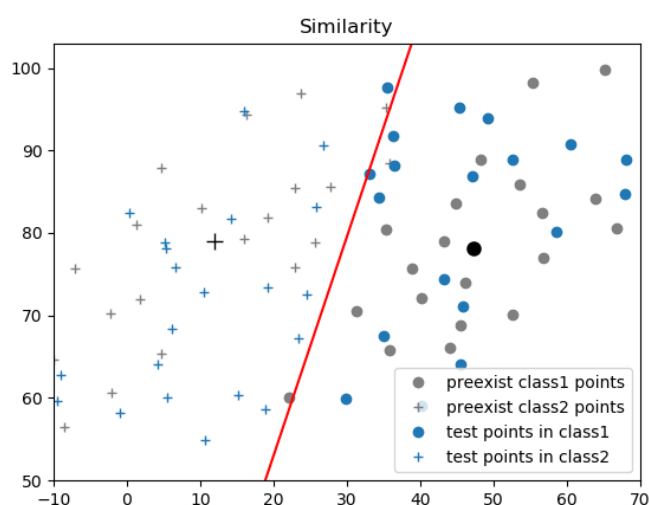
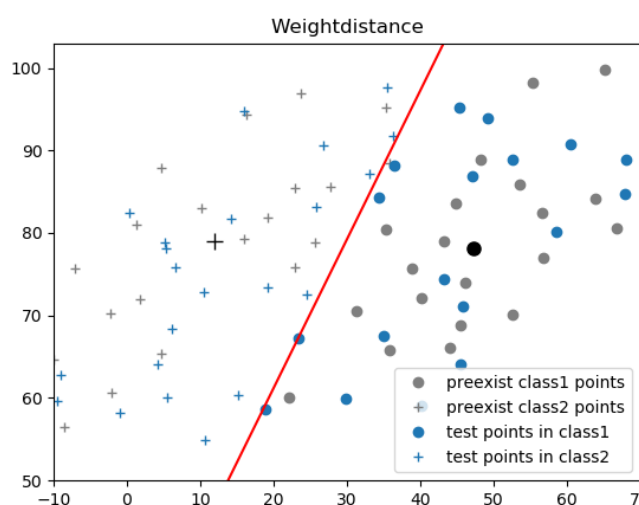
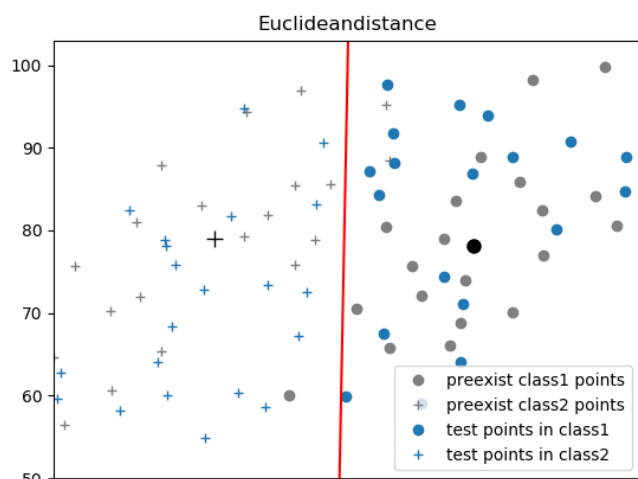
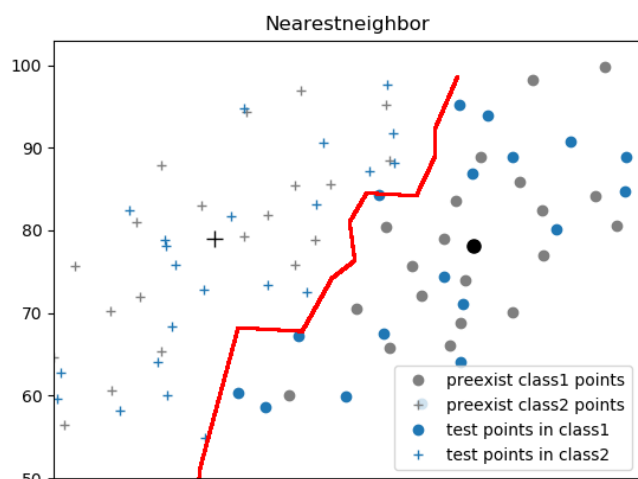


図 4 (a) Correct distribution of test set D.



(d) The result of using Weight Euclidean method. (D)

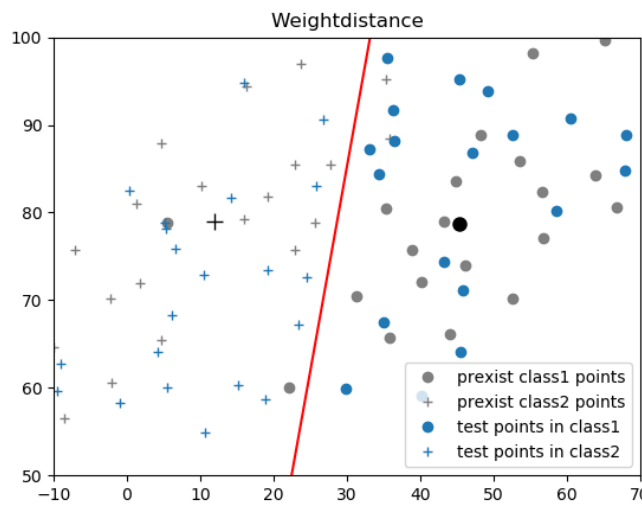
(e) The result of using Similarity method. (D)

1.3 (d)

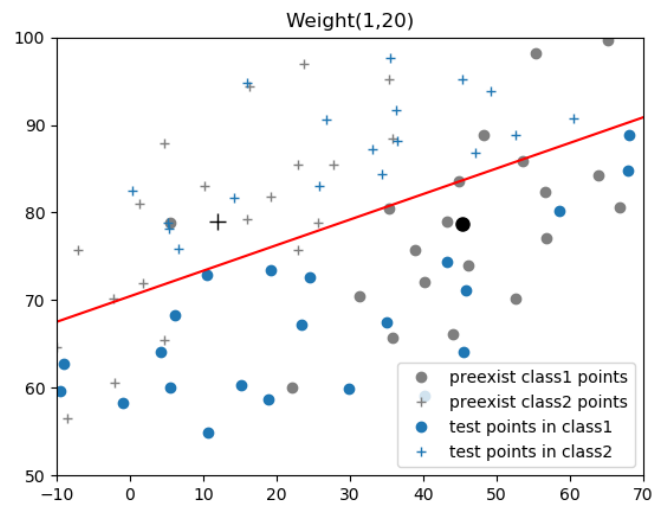
B, C, D データセットの判別結果もセット A と同じような結果が得られた。結論から言うと、今回の学習データはクラス間が適度に空白地帯が空いており、お互い交り合っている部分も少ない（ノイズが少ない）ため nearest neighbor 法はいずれのデータセットにおいても高い認識率を示した。その他の方法ではクラスデータを無理やり一次関数で分断しようとしているため、認識率をあげたくても限界があると予想できる。この困難を克服ため高次方程式で識別境界線を描けるアルゴリズムを作る必要があると考えられる。

1.4 (e)

重み付きユークリッド距離法において重みを変えるというのは各軸の重みを変え、識別線の傾きを変えるということを意味する。データセット A において重みを (1,5) と (1,20) にした結果を下図 5 に示す。明らかに、y 座標がクラス判別に与える影響は (1,20) の方が遥かに大きいことがわかった。



(a) Weight (1,5)



(b) Weight (1,20)