

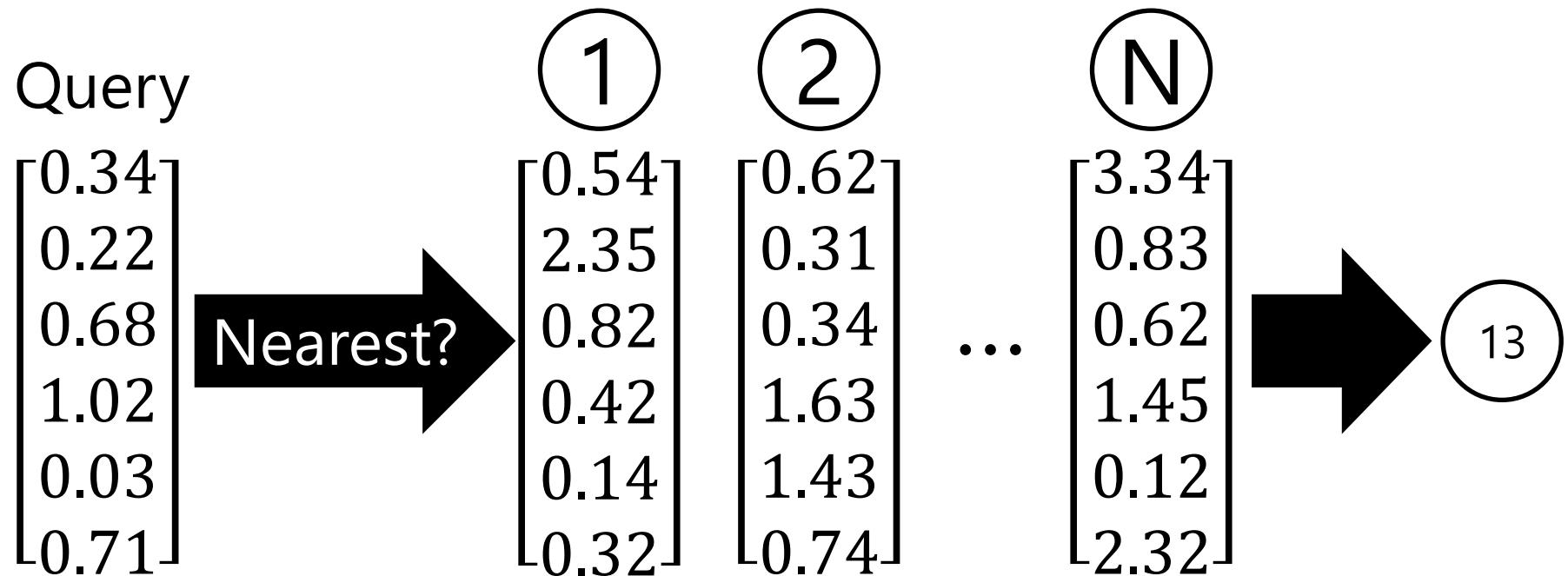
【招待ショートサーベイ】 直積量子化を用いた近似最近傍探索

2016/9/5, PRMU研究会@富山

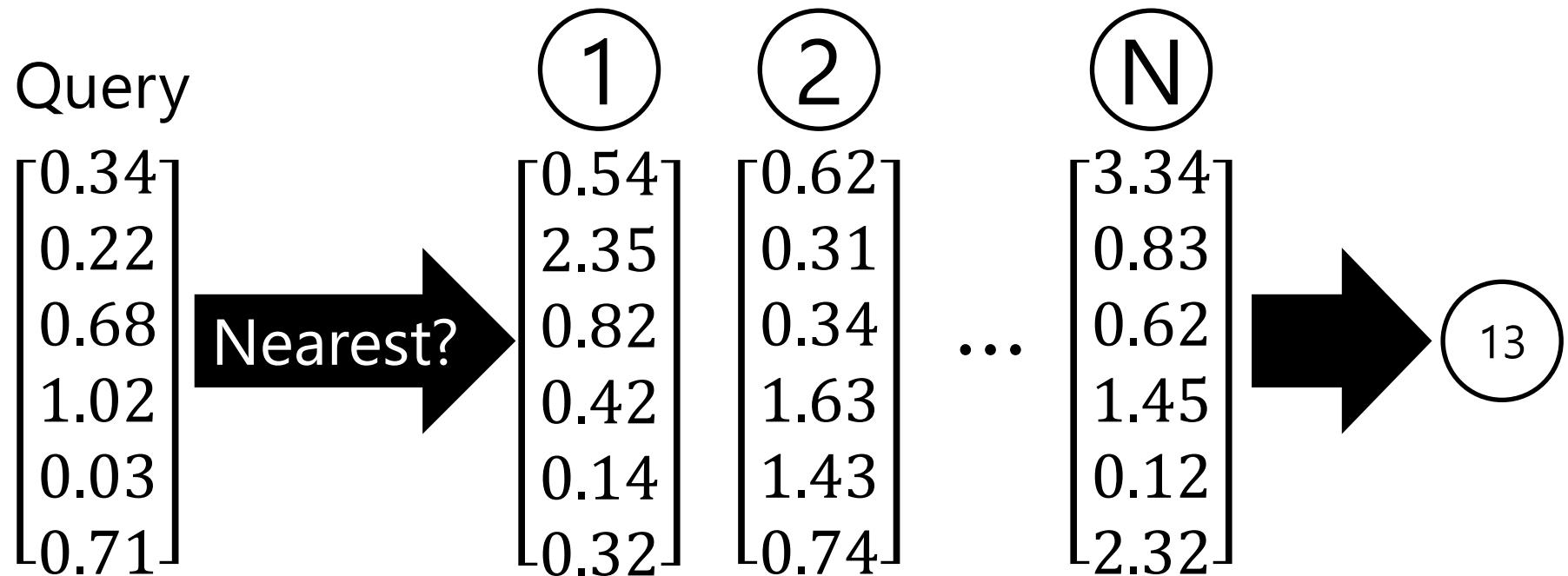
松井勇佑
国立情報学研究所



近似最近傍探索

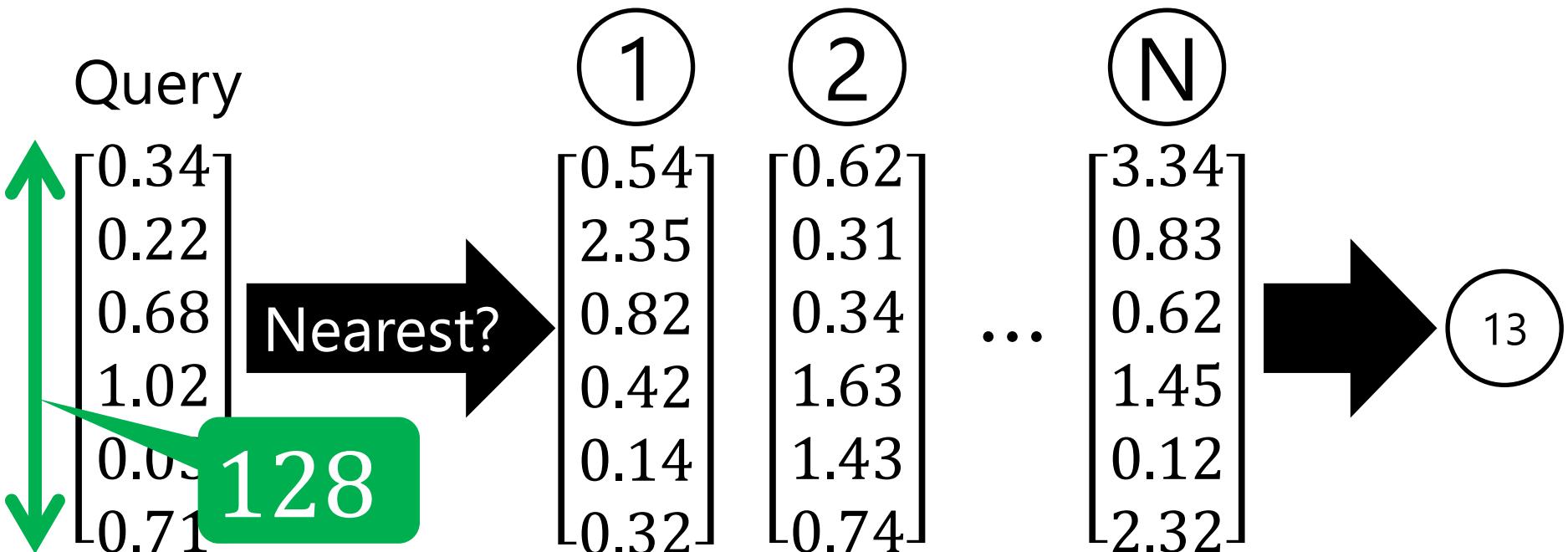


近似最近傍探索



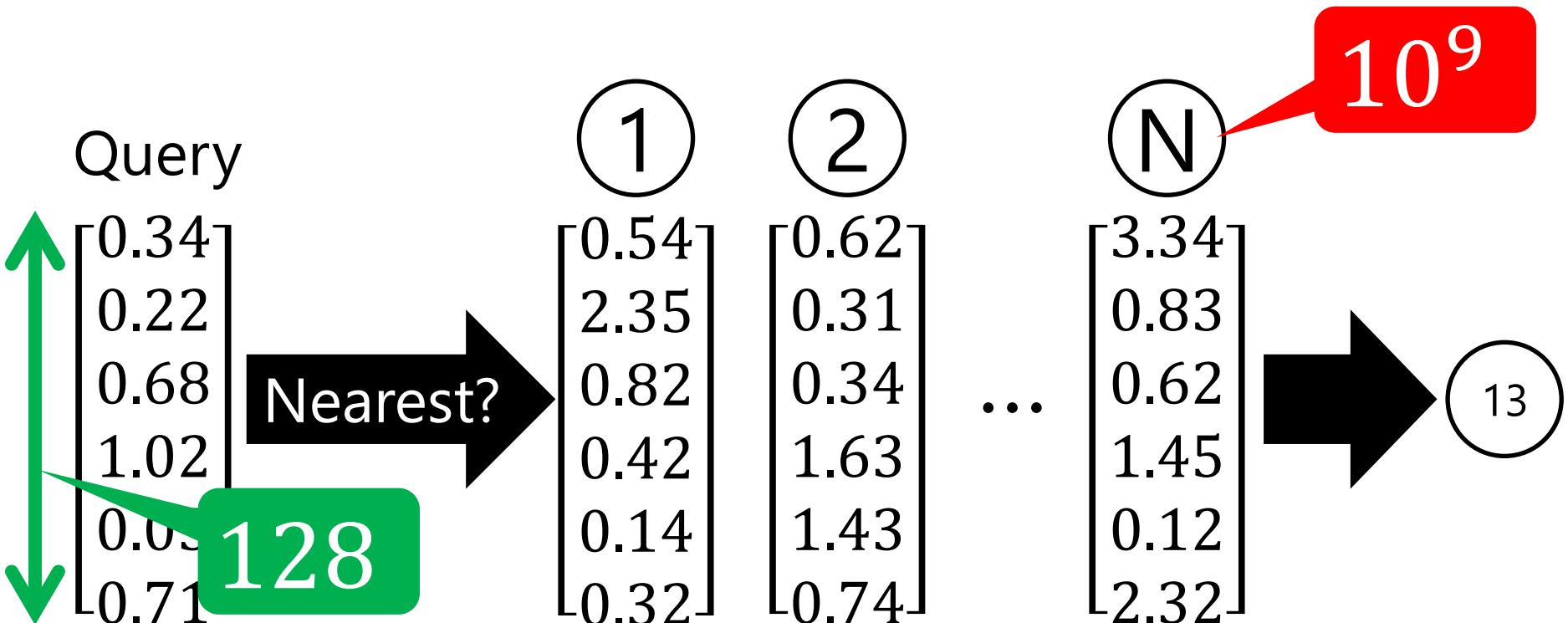
$$n^* = \arg \min_{1 \leq n \leq N} \|q - x_n\|^2$$

近似最近傍探索



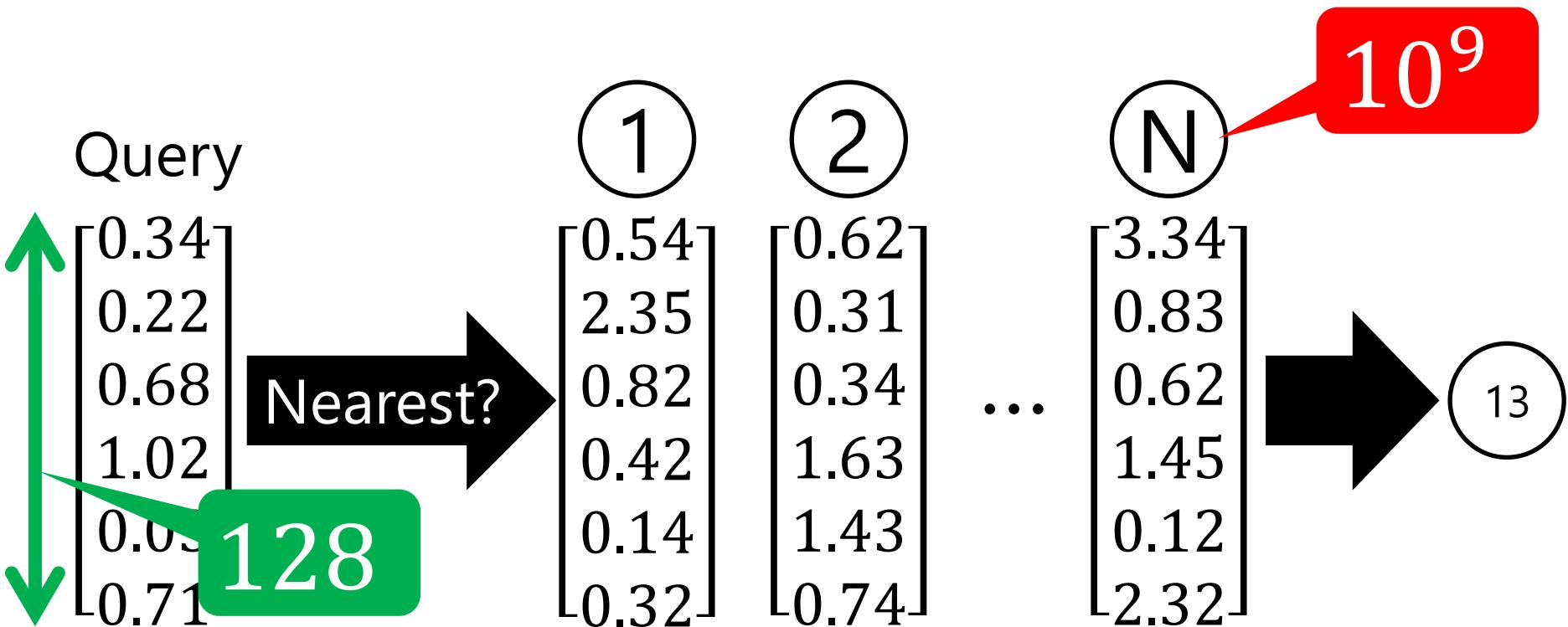
$$n^* = \arg \min_{1 \leq n \leq N} \|q - x_n\|^2$$

近似最近傍探索



$$n^* = \arg \min_{1 \leq n \leq N} \|q - x_n\|^2$$

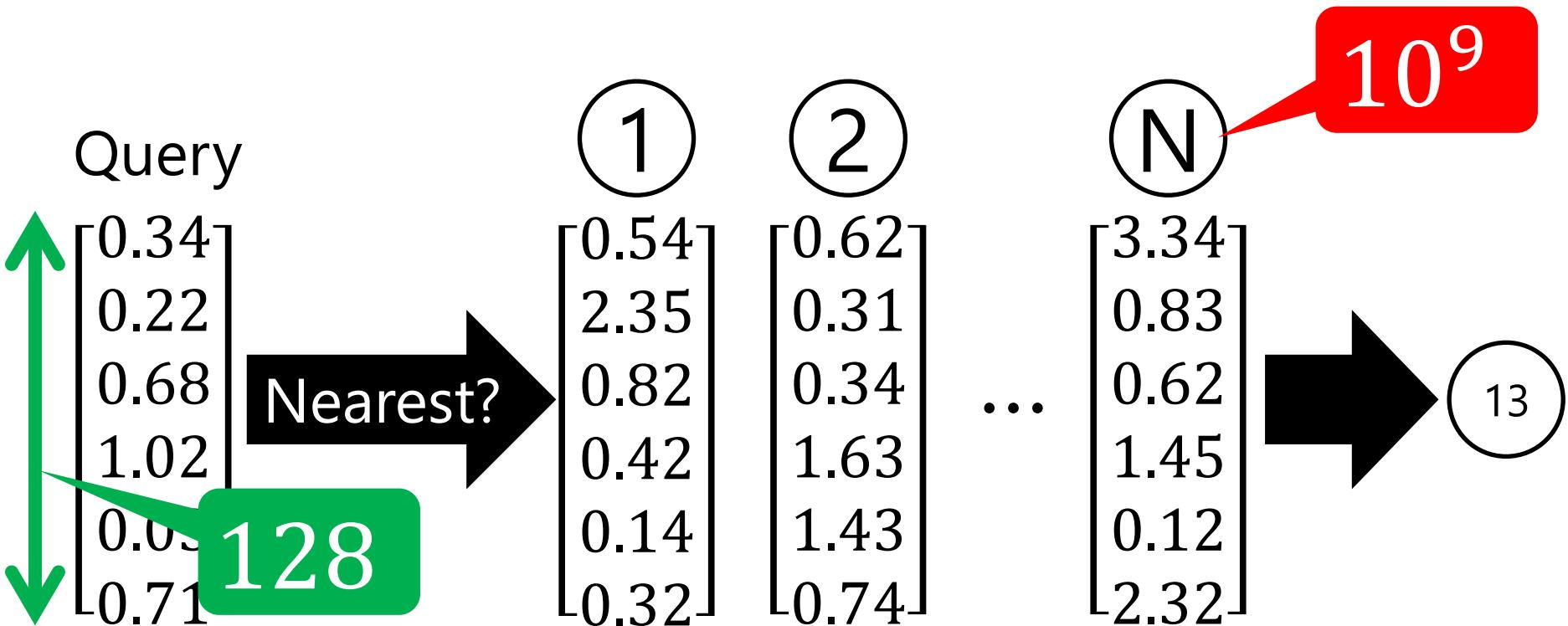
近似最近傍探索



$$n^* = \arg \min_{1 \leq n \leq N} \|q - x_n\|^2$$

10 ms

近似最近傍探索



$$n^* = \arg \min_{1 \leq n \leq N} \|q - x_n\|^2$$

10 ms

32GB RAM

近似最近傍探索

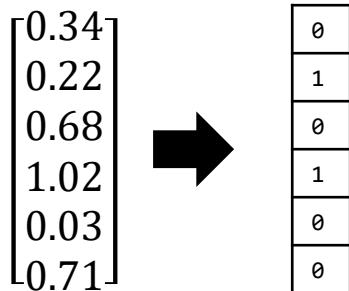
Locality Sensitive
Hashing (LSH)系

Tree系
(e.g., FLANN [Muja, PAMI 14])

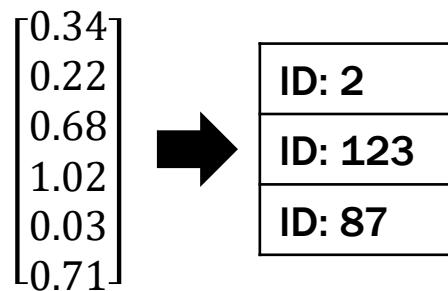
etc.

ショートコード系

バイナリハッシュ



直積量子化 (PQ)

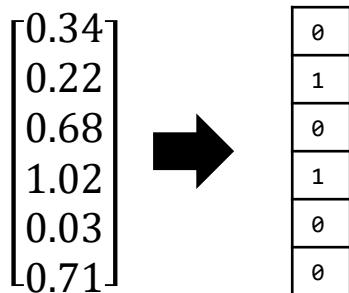


近似最近傍探索

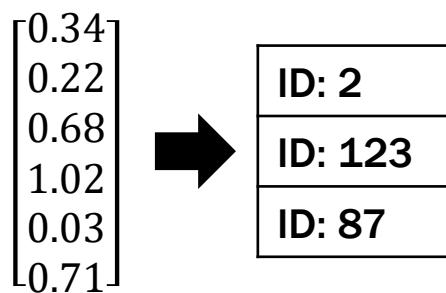
Locality Sensitive Hashing (LSH) 系
メモリ消費大
(e.g., FLANN)
etc.

ショートコード系

バイナリハッシュ



直積量子化 (PQ)



近似最近傍探索

Locality Sensitive Hashing (LSH) 系
メモリ消費大
(e.g., FLANN)
etc.

ショートコード系

シンプル
探索が高速

精度劣る

直積量子化 (PQ)

$$\begin{bmatrix} 0.34 \\ 0.22 \\ 0.68 \\ 1.02 \\ 0.03 \\ 0.71 \end{bmatrix}$$

A diagram illustrating the mapping process. On the left, a vector represented by a column of numbers is shown. An arrow points from this vector to a table on the right. The table has four rows, each containing an ID number: ID: 2, ID: 123, ID: 87, and ID: 123 again.

ID: 2
ID: 123
ID: 87
ID: 123

近似最近傍探索

Locality Sensitive Hashing (LSH) 系
メモリ消費大
(e.g., FLANN)
etc.

ショートコード系

☺ シンプル
探索が高速

☹ 精度劣る

直積量子化 (PQ)

$$\begin{bmatrix} 0.34 \\ 0.22 \\ 0.68 \\ 1.02 \\ 0.03 \\ 0.71 \end{bmatrix}$$

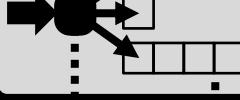


ID: 2
ID: 123
ID: 87

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの の発展

一般化

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った 探索システムの発展

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

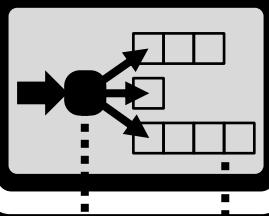
GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

PQそのもの

PQを使った探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの 発展

一般化

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

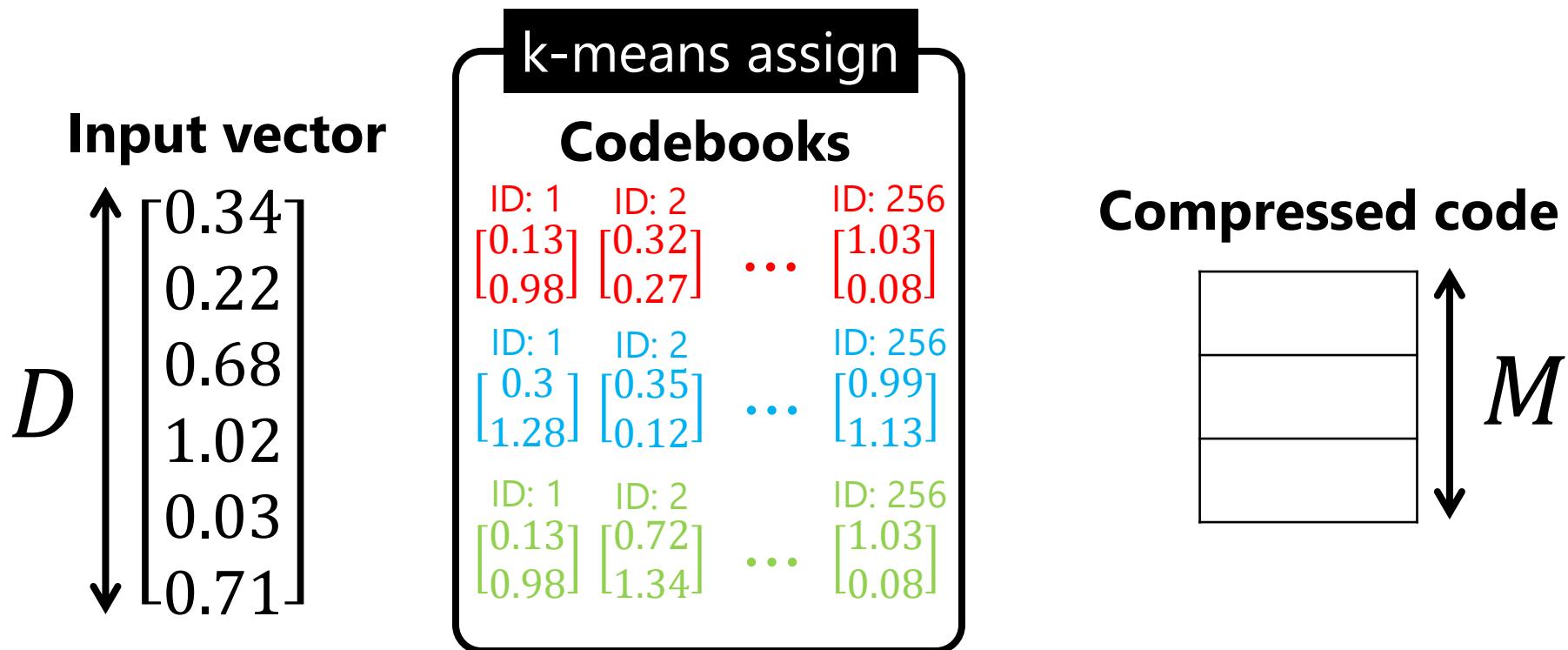
local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った 探索システムの発展

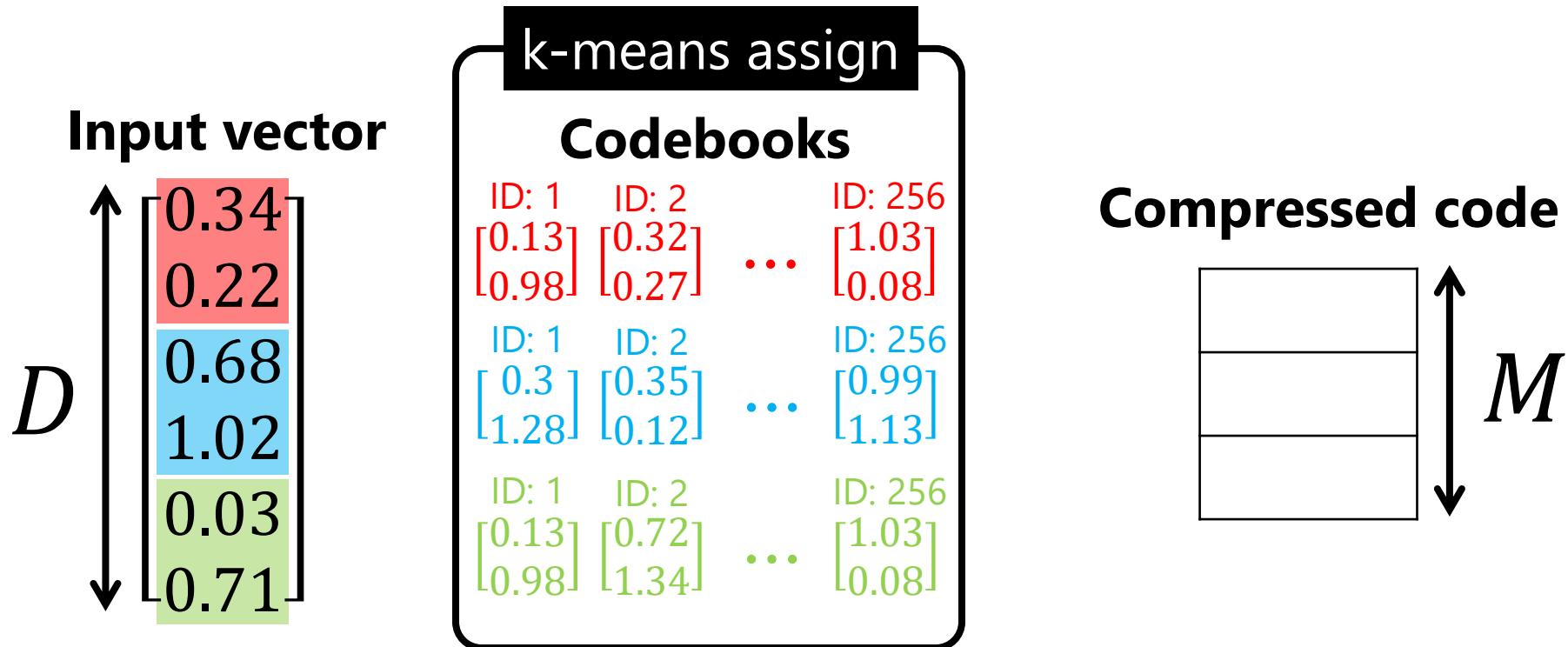
Product Quantization [Jégou, TPAMI 2011]

➤ ベクトルを分割して k-means する



Product Quantization [Jégou, TPAMI 2011]

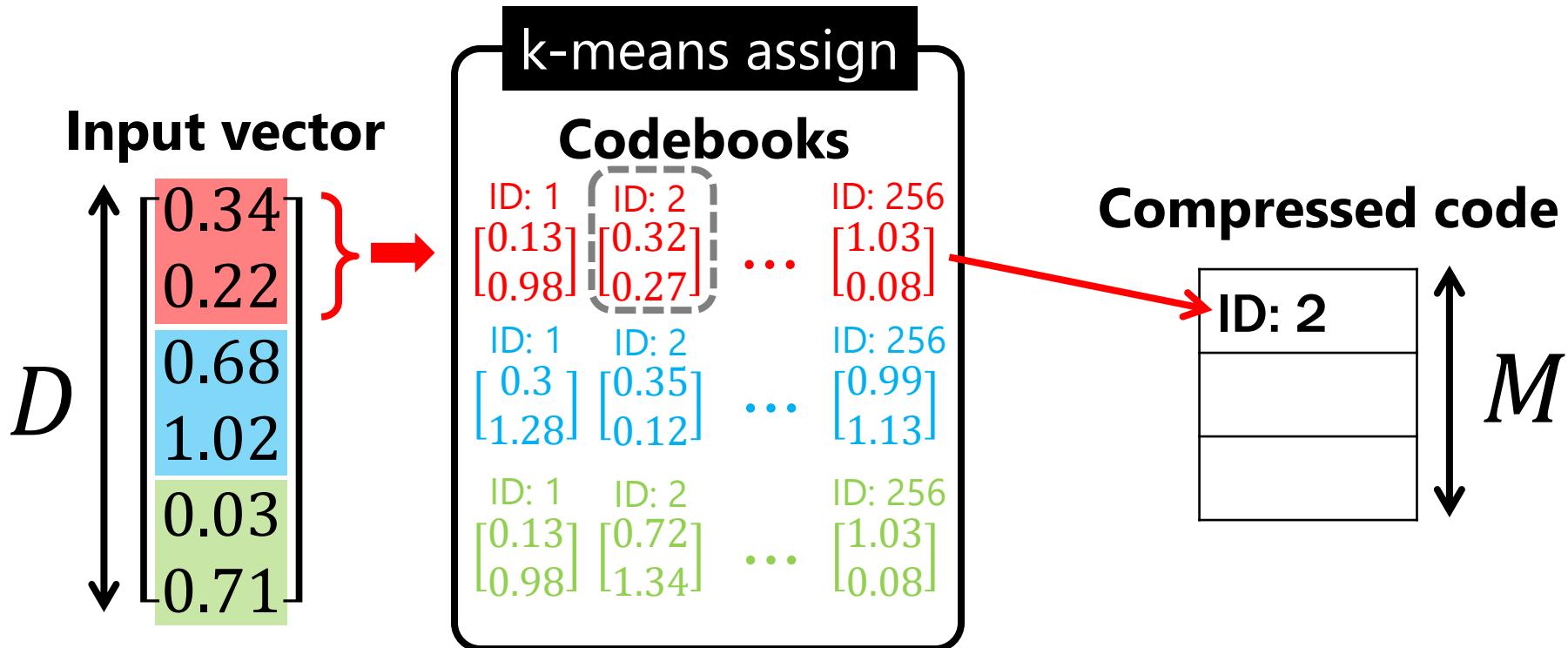
➤ ベクトルを分割して k-means する



Product Quantization

[Jégou, TPAMI 2011]

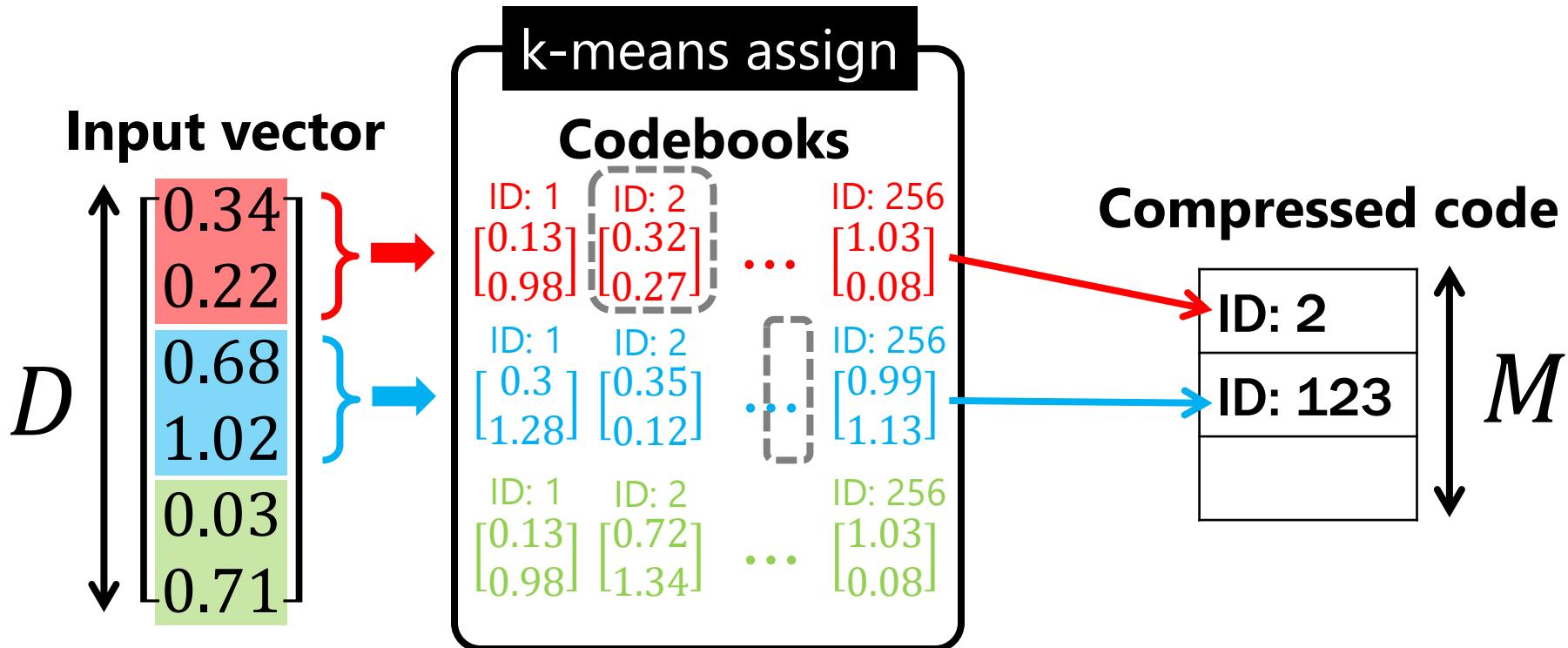
➤ ベクトルを分割して k-means する



Product Quantization

[Jégou, TPAMI 2011]

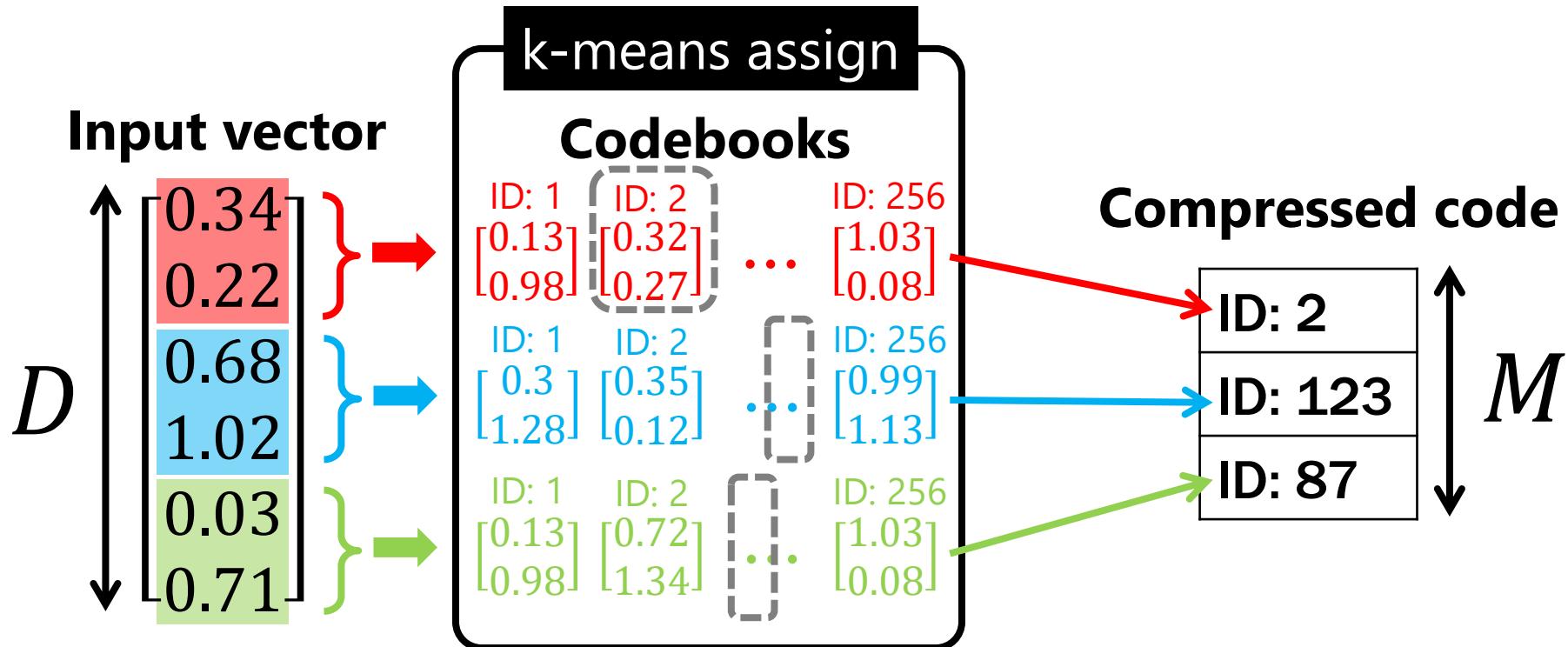
➤ ベクトルを分割して k-means する



Product Quantization

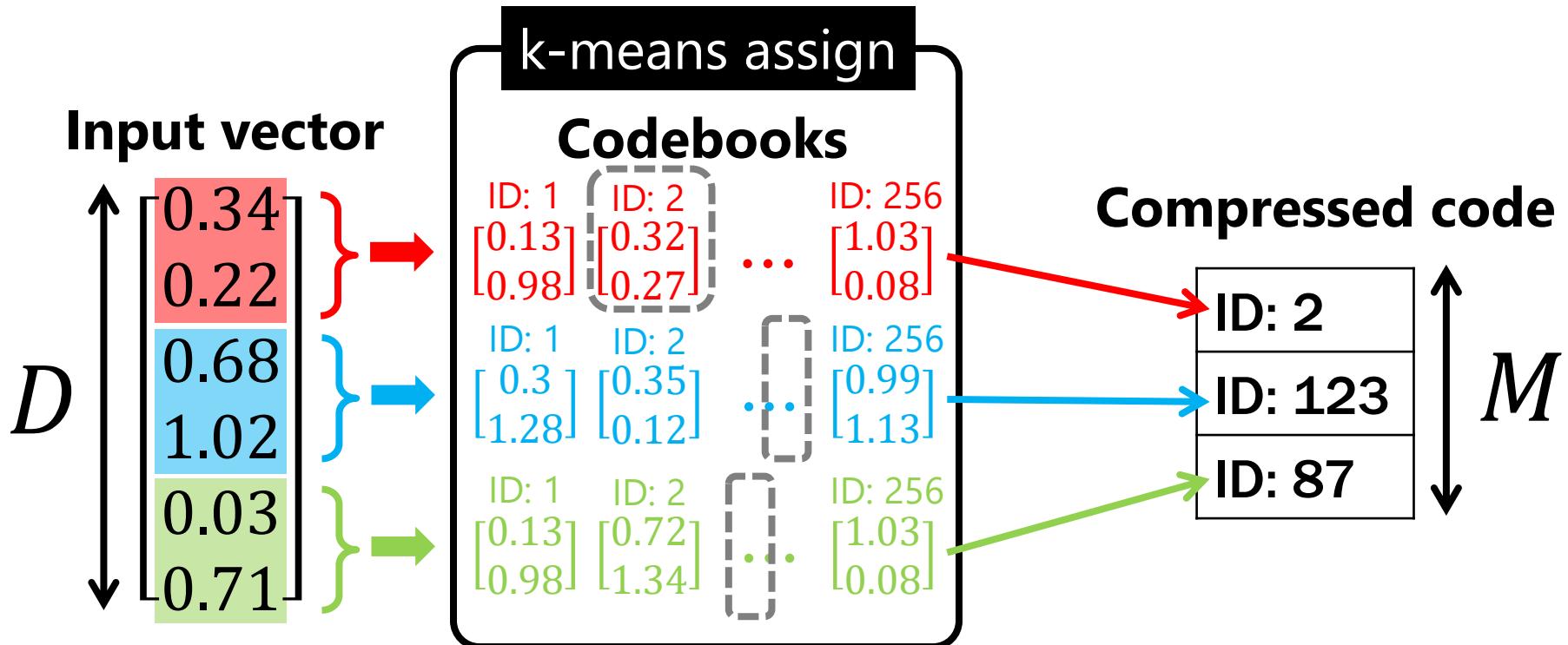
[Jégou, TPAMI 2011]

➤ ベクトルを分割して k-means する



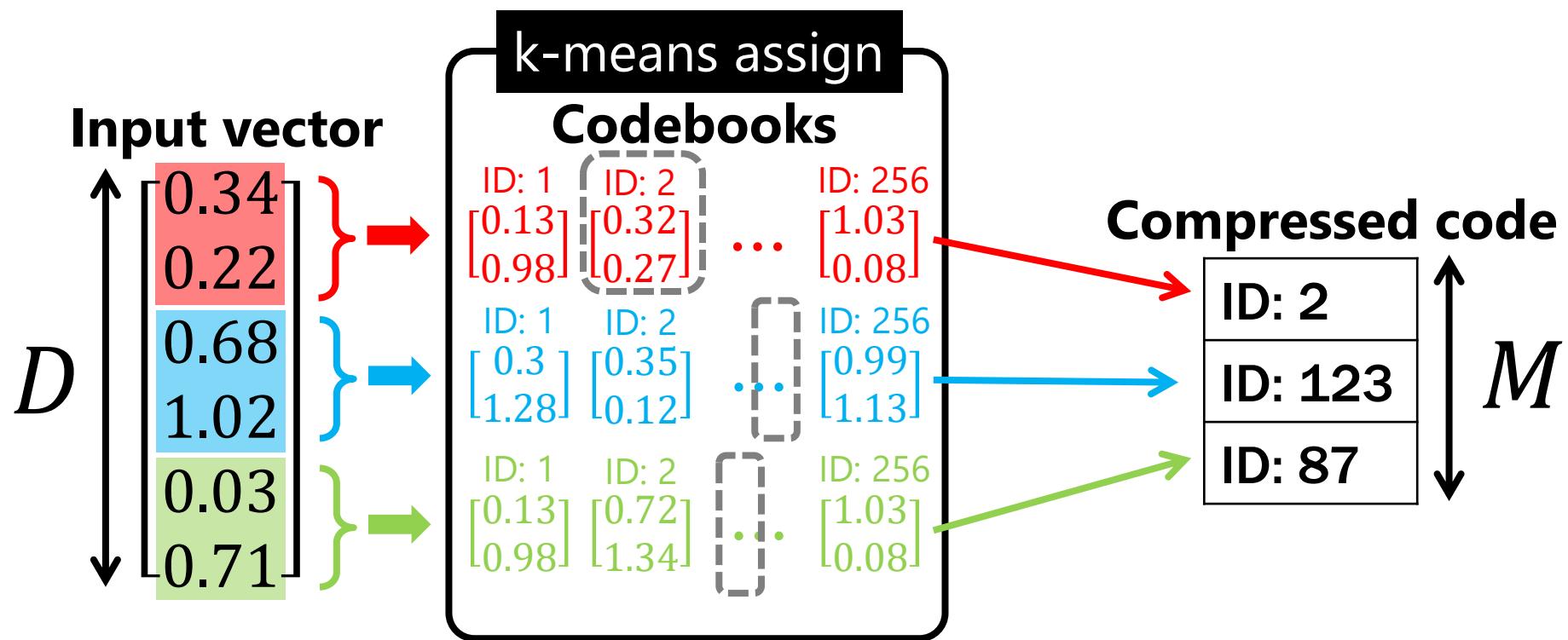
Product Quantization [Jégou, TPAMI 2011]

- ベクトルを分割して k-means する

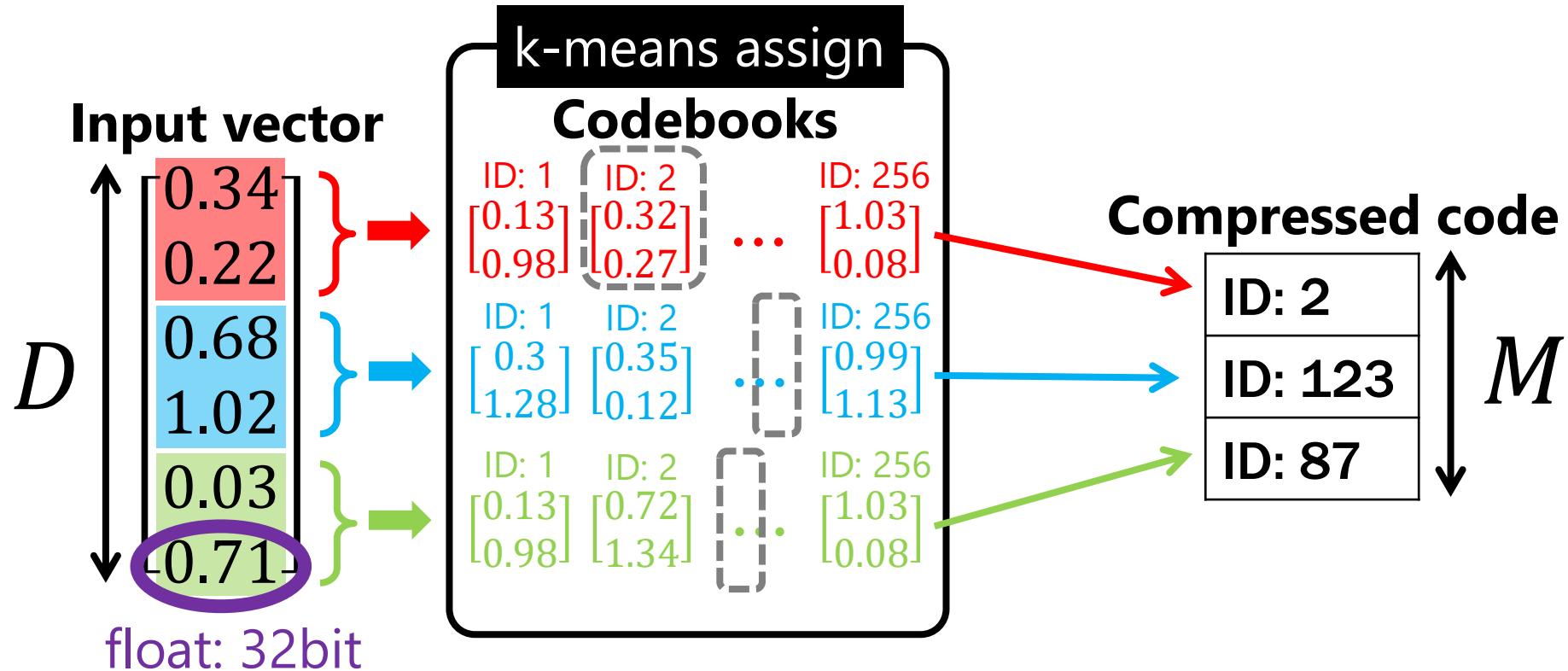


- 単純
- メモリ効率良い
- 距離 $d(\text{input}, \text{code})^2$ を近似計算可能

Product Quantization: メモリ効率良



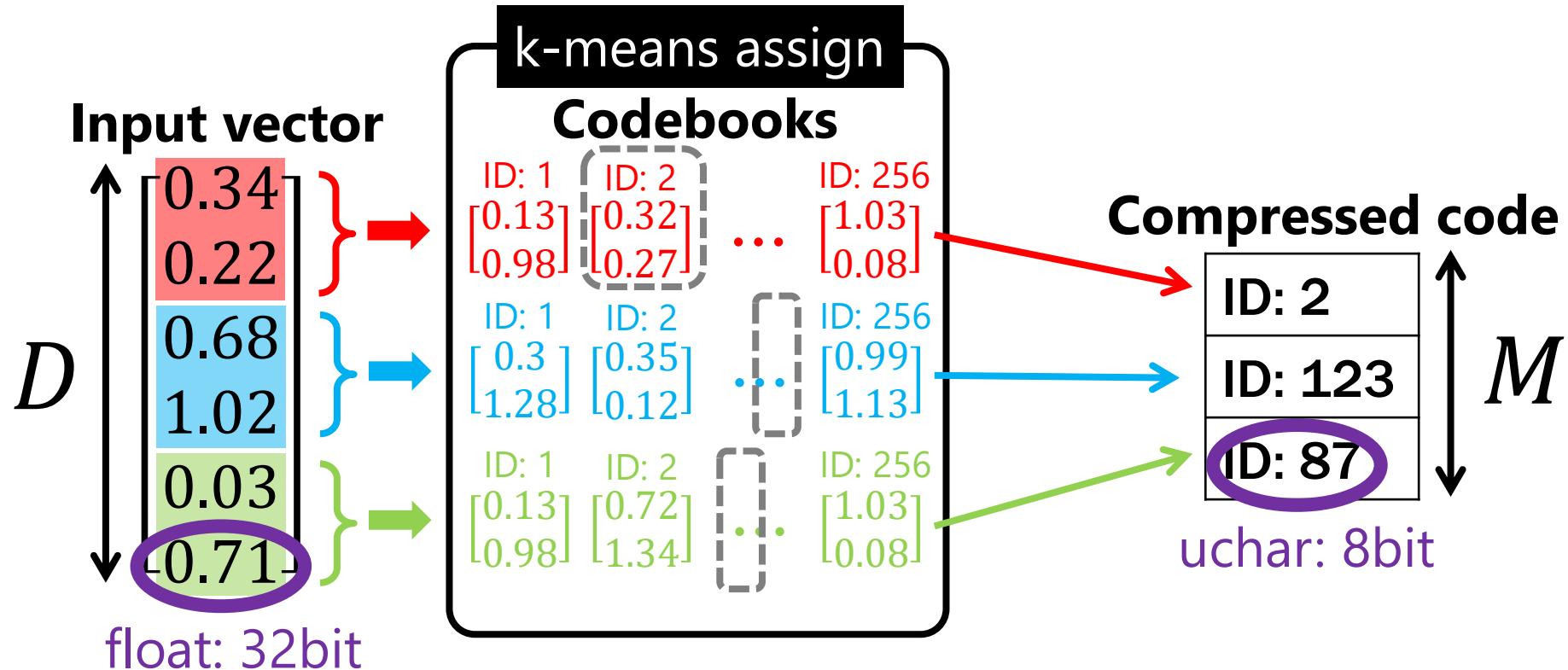
Product Quantization: メモリ効率良



e.g., $D = 128$

$128 \times 32 = 4096$ [bit]

Product Quantization: メモリ効率良



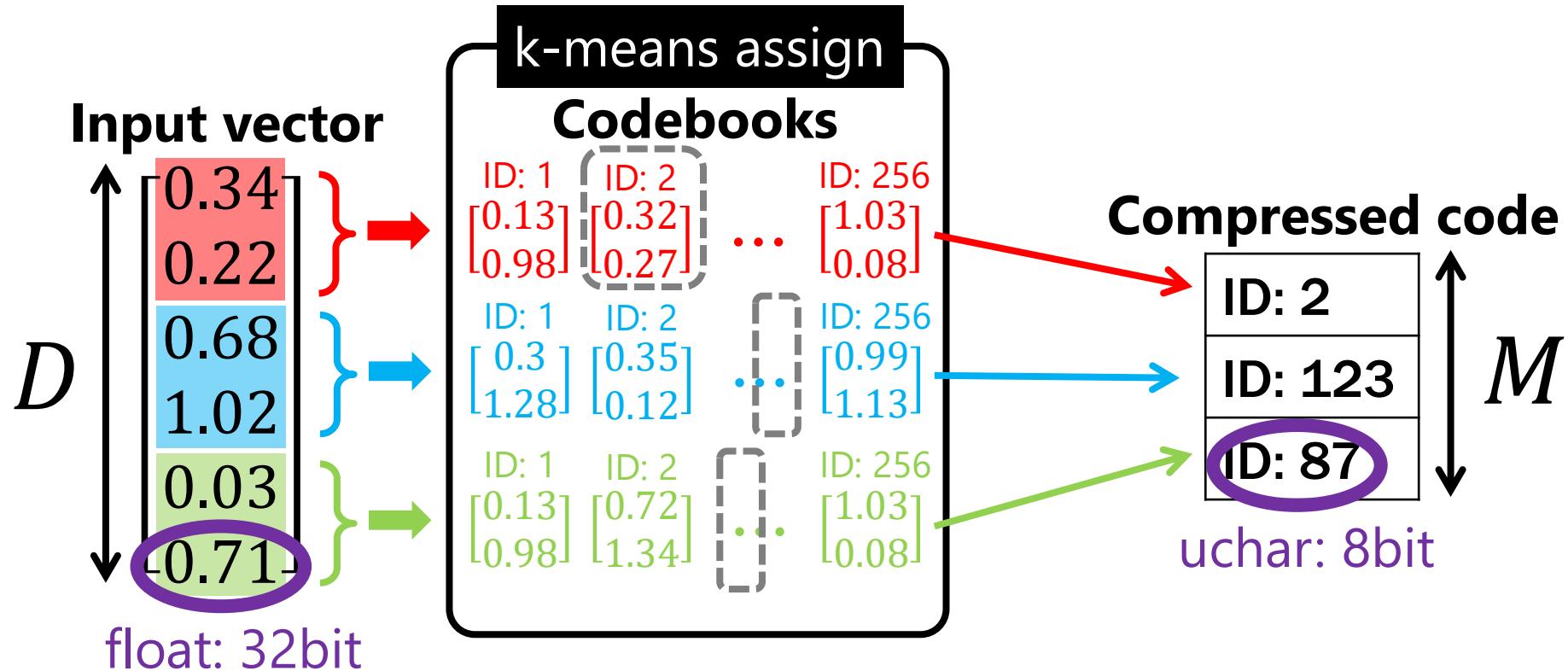
e.g., $D = 128$

$$128 \times 32 = 4096 \text{ [bit]}$$

e.g., $M = 8$

$$8 \times 8 = 64 \text{ [bit]}$$

Product Quantization: メモリ効率良



e.g., $D = 128$

$$128 \times 32 = 4096 \text{ [bit]}$$

e.g., $M = 8$

$$8 \times 8 = 64 \text{ [bit]}$$

1/64に圧縮

Product Quantization: 距離近似

Query

$$\begin{bmatrix} 0.34 \\ 0.22 \\ 0.68 \\ 1.02 \\ 0.03 \\ 0.71 \end{bmatrix}$$

1

$$\begin{bmatrix} 0.54 \\ 2.35 \\ 0.82 \\ 0.42 \\ 0.14 \\ 0.32 \end{bmatrix}$$

2

$$\begin{bmatrix} 0.62 \\ 0.31 \\ 0.34 \\ 1.63 \\ 1.43 \\ 0.74 \end{bmatrix}$$

...

N

$$\begin{bmatrix} 3.34 \\ 0.83 \\ 0.62 \\ 1.45 \\ 0.12 \\ 2.32 \end{bmatrix}$$

Product Quantization: 距離近似

Query

$$\begin{bmatrix} 0.34 \\ 0.22 \\ 0.68 \\ 1.02 \\ 0.03 \\ 0.71 \end{bmatrix}$$

1

$$\begin{bmatrix} 0.54 \\ 2.35 \\ 0.82 \\ 0.42 \\ 0.14 \\ 0.32 \end{bmatrix}$$

2

$$\begin{bmatrix} 0.62 \\ 0.31 \\ 0.34 \\ 1.63 \\ 1.43 \\ 0.74 \end{bmatrix}$$

...

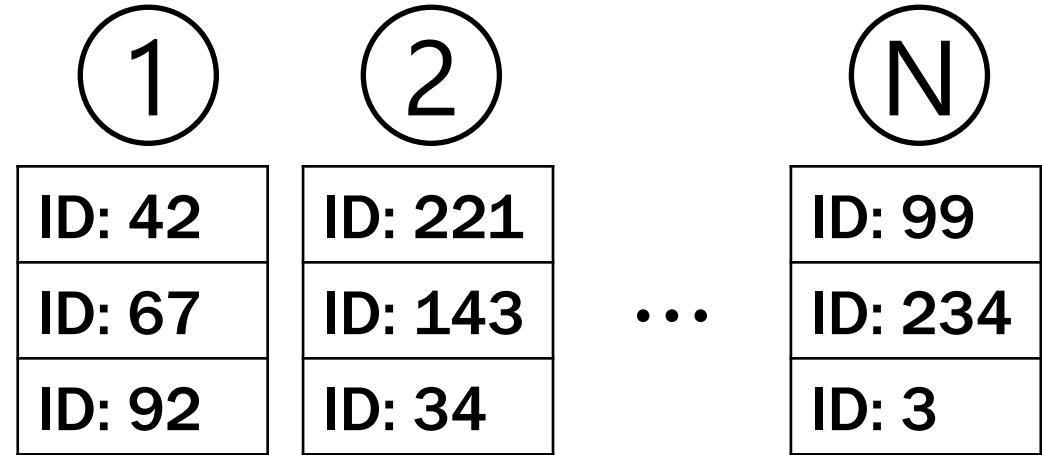
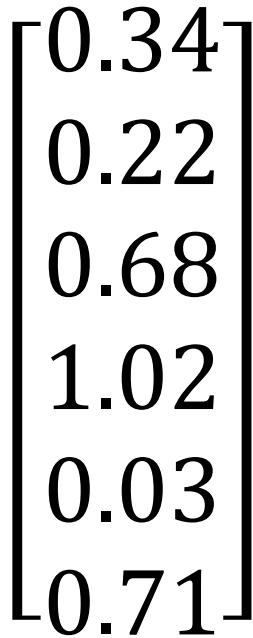
N

$$\begin{bmatrix} 3.34 \\ 0.83 \\ 0.62 \\ 1.45 \\ 0.12 \\ 2.32 \end{bmatrix}$$

Product
quantization

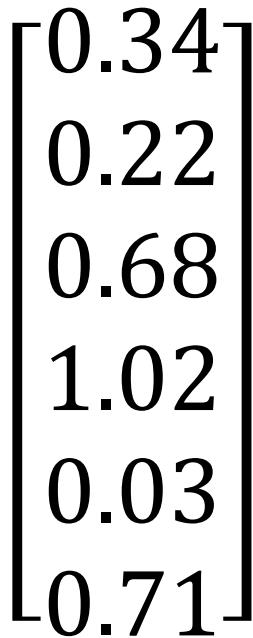
Product Quantization: 距離近似

Query

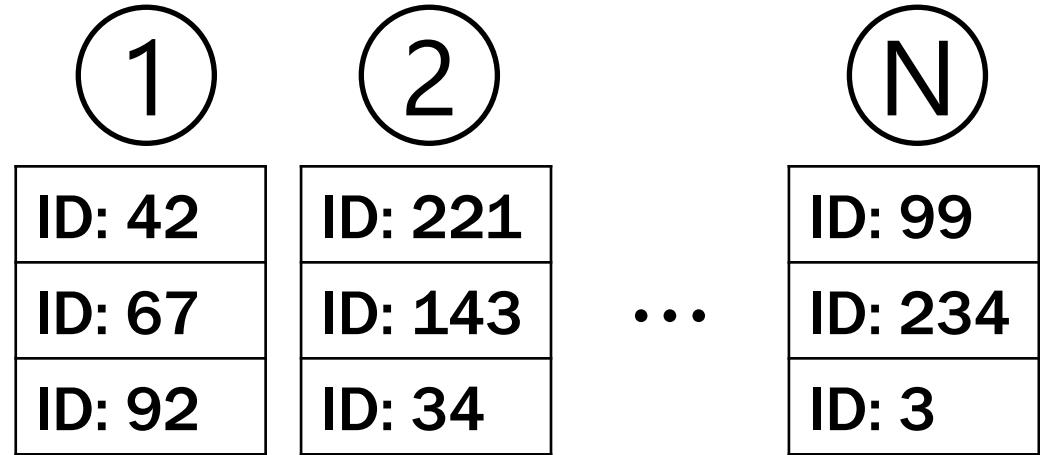


Product Quantization: 距離近似

Query

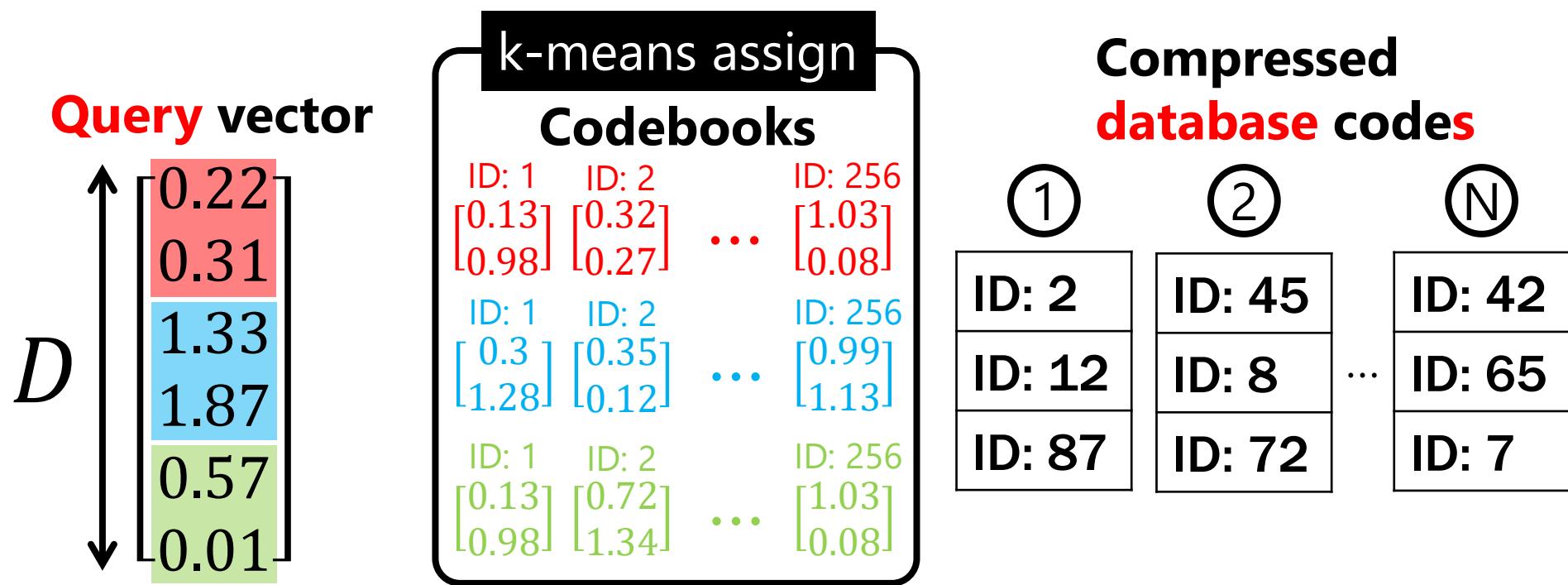


線形
探索

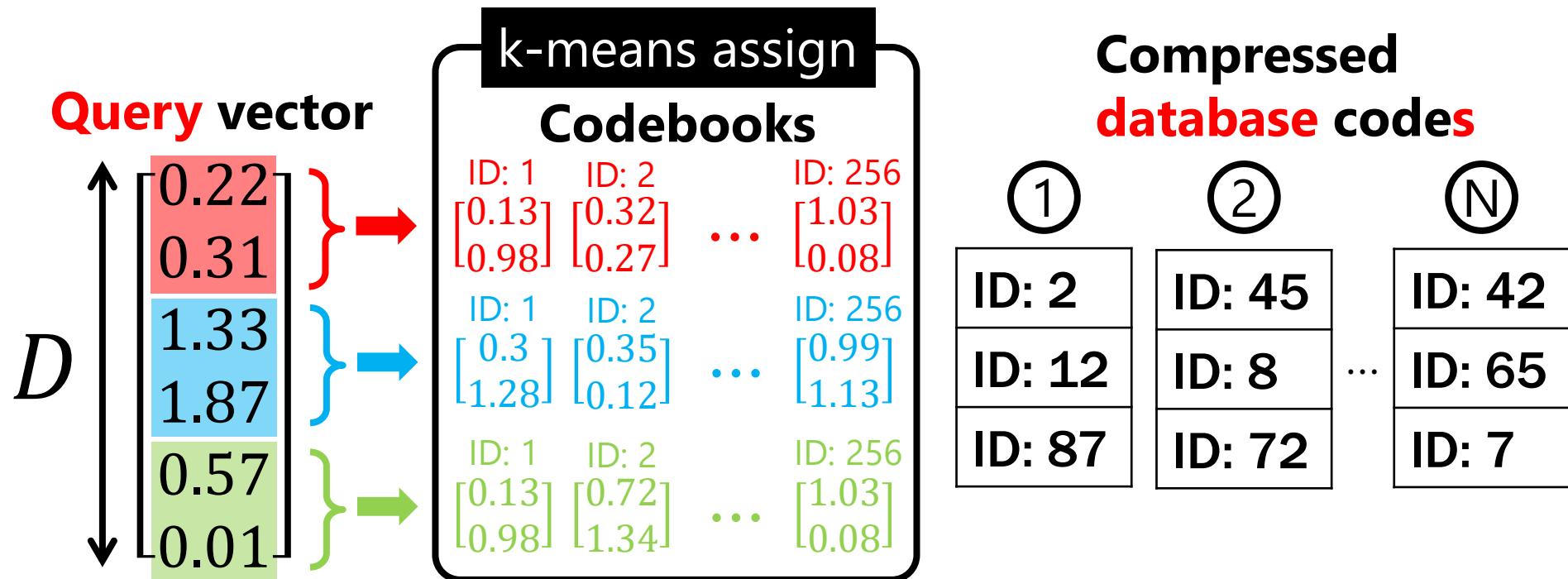


(近似的距離で) 線形探索が出来る

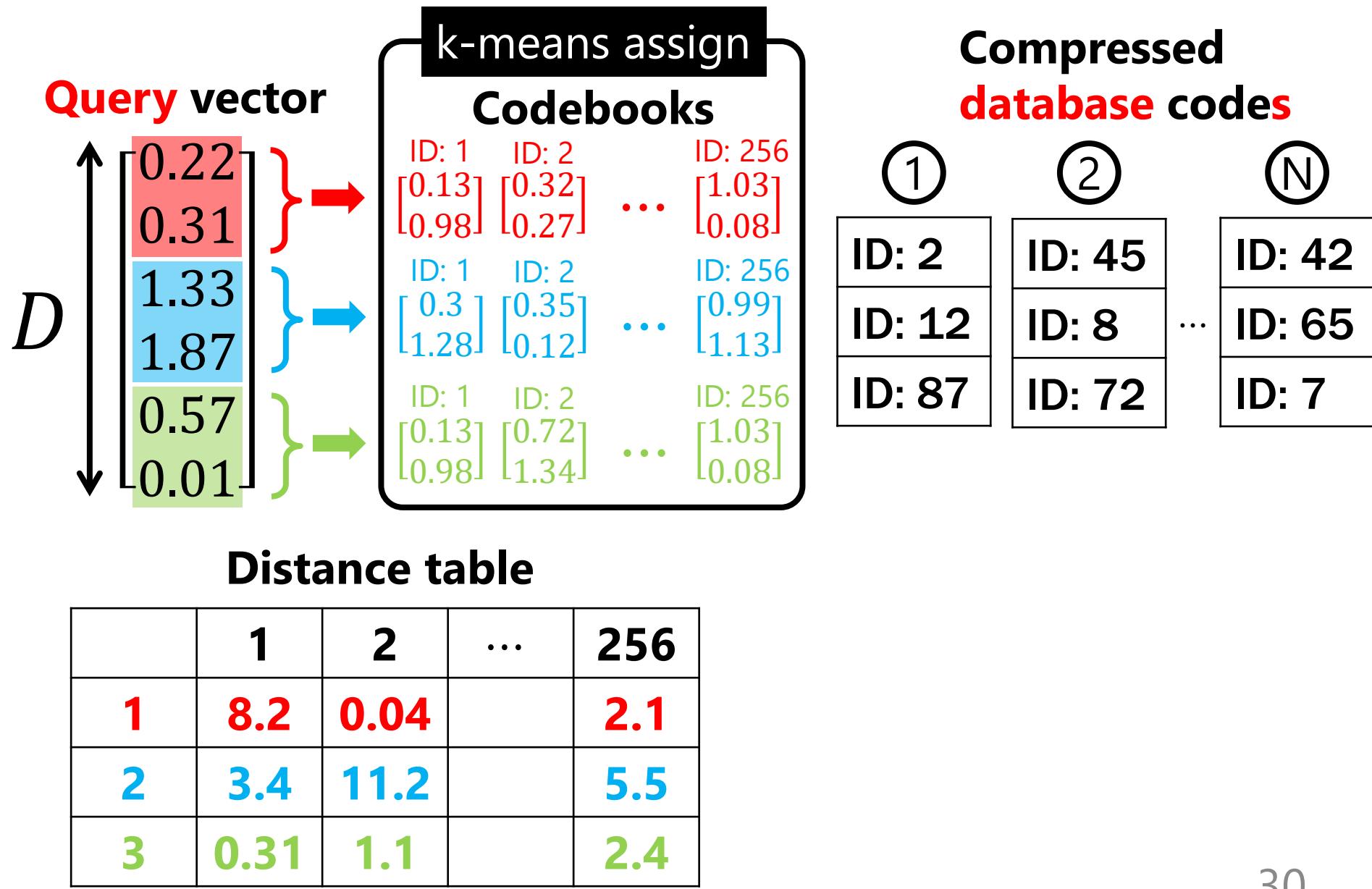
Product Quantization: 距離近似



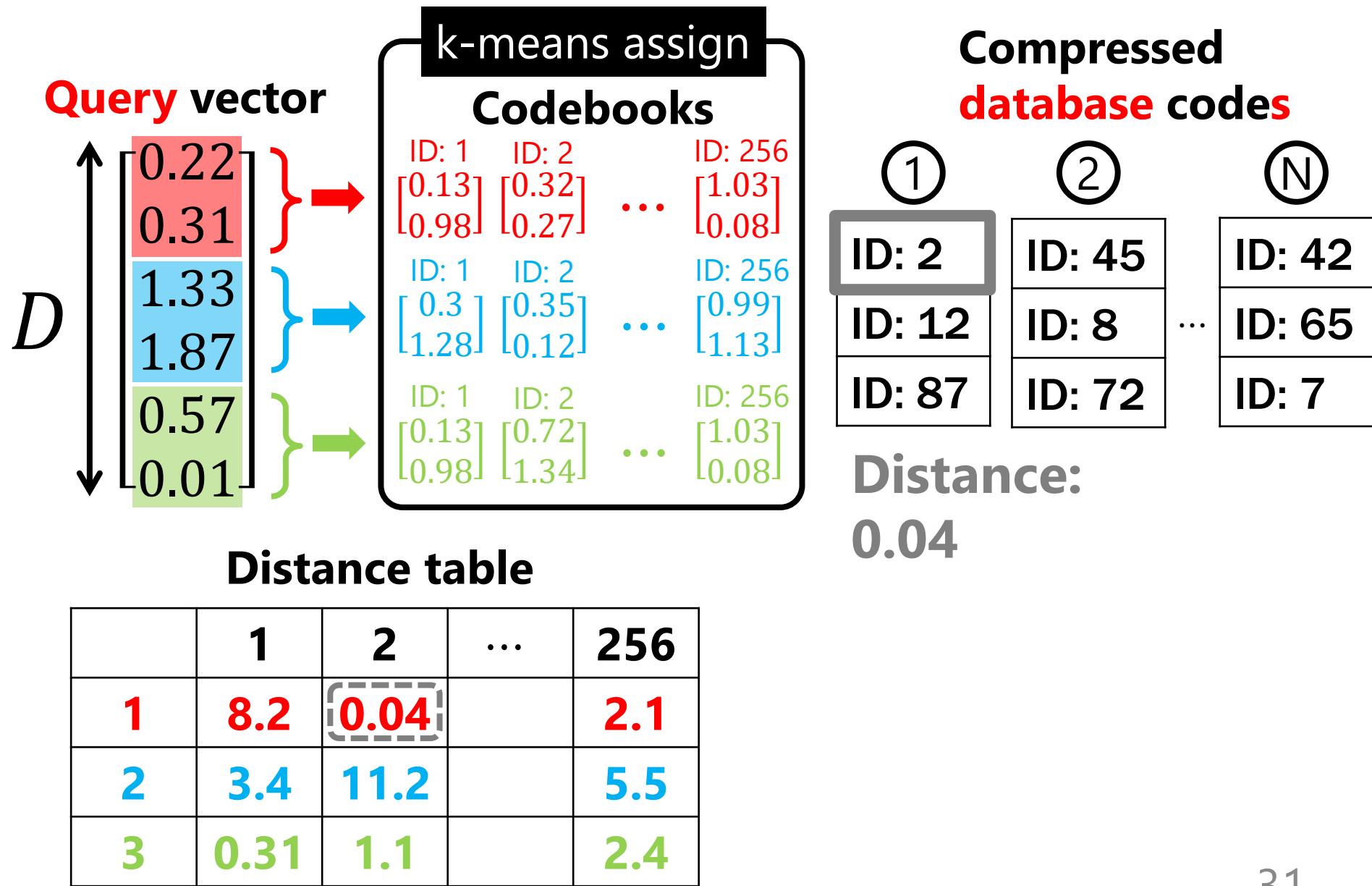
Product Quantization: 距離近似



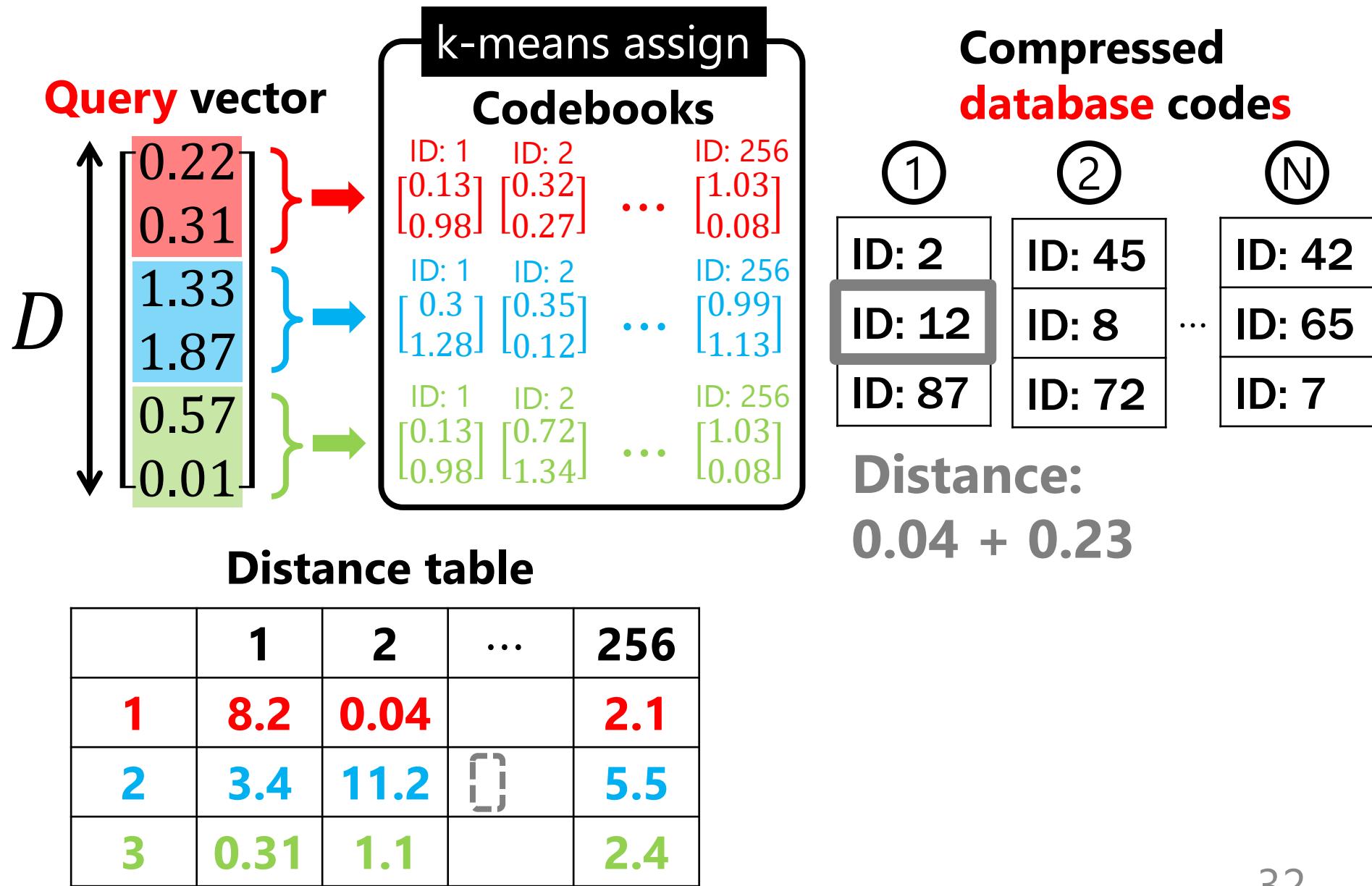
Product Quantization: 距離近似



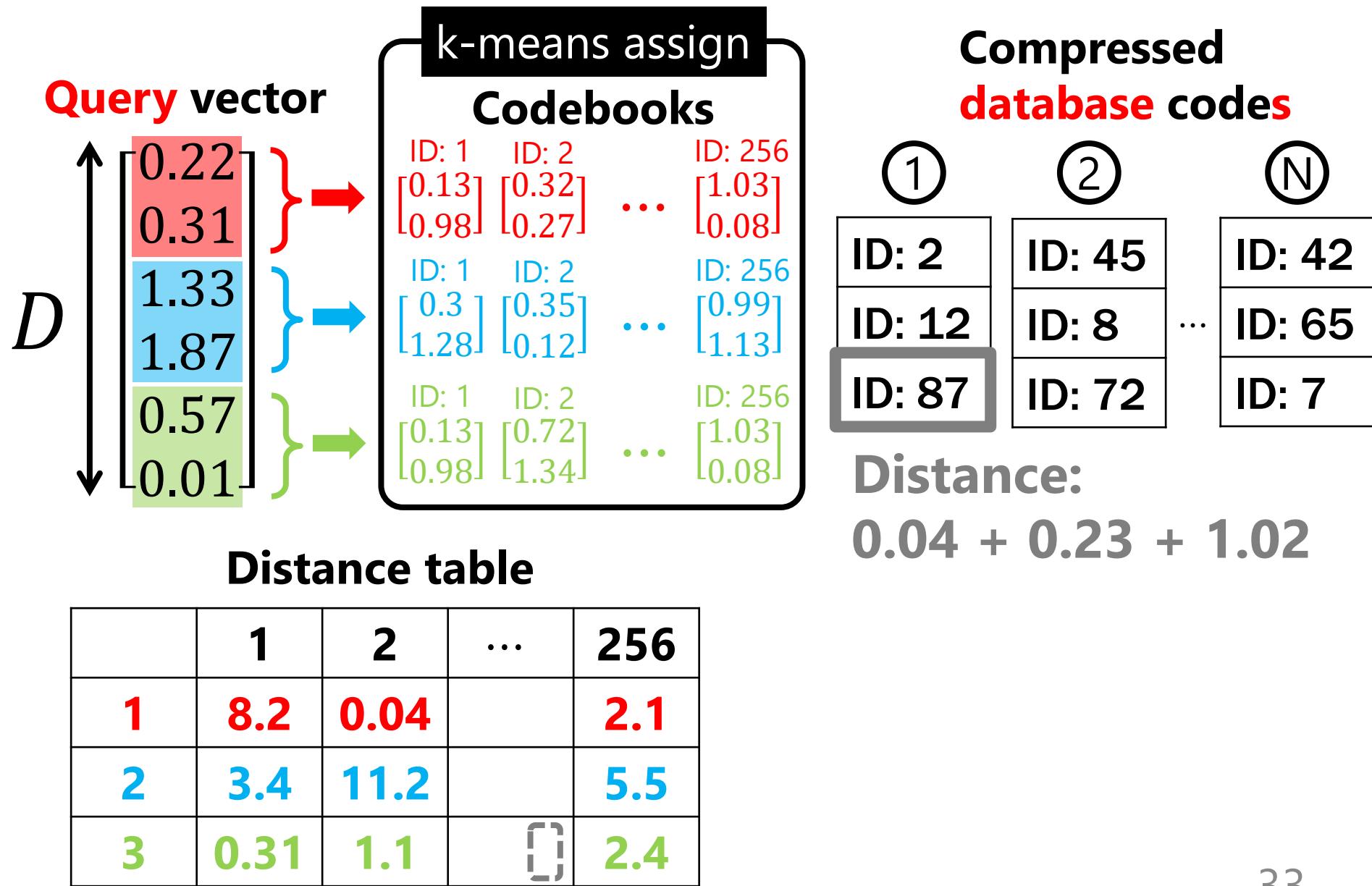
Product Quantization: 距離近似



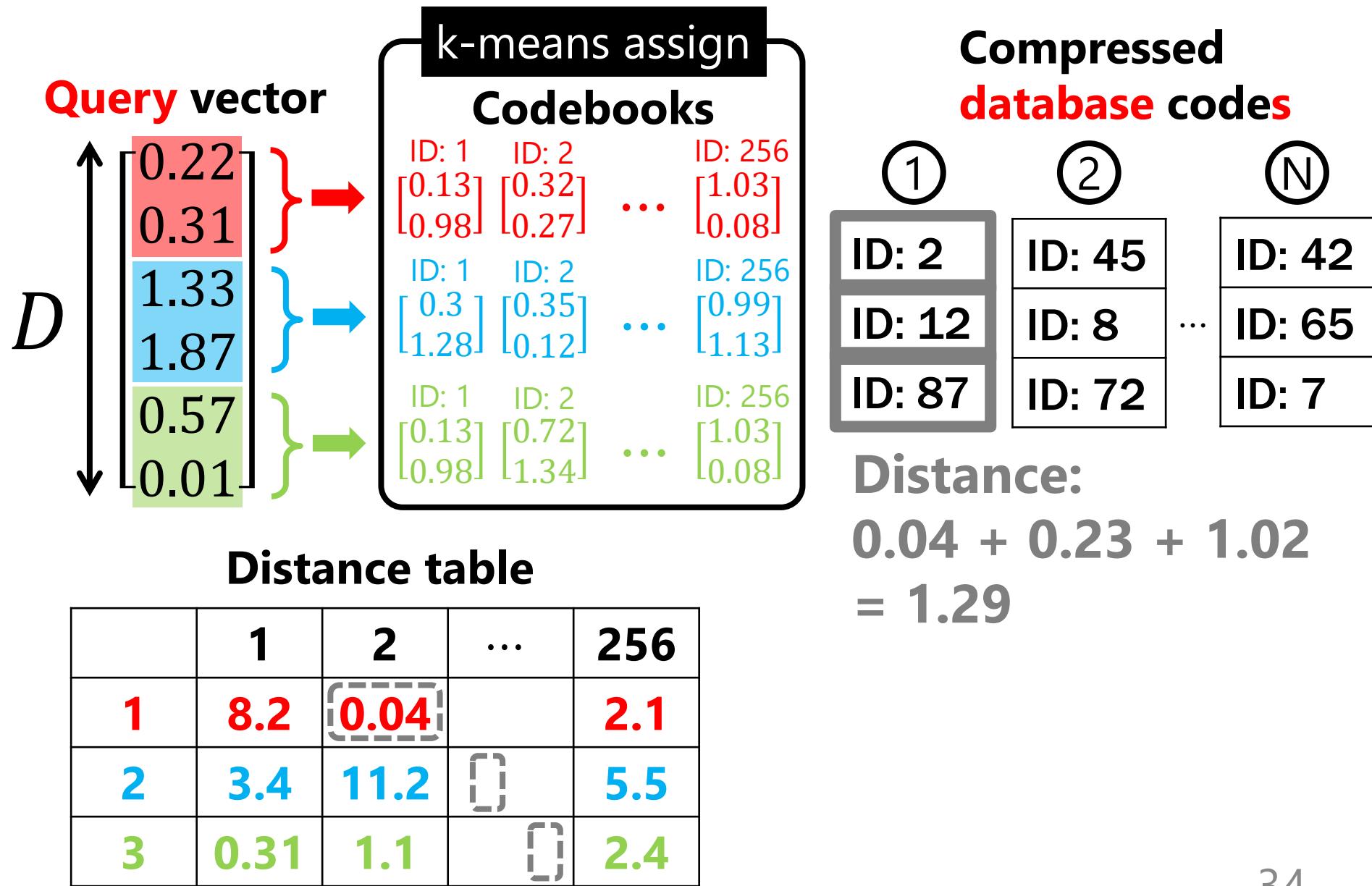
Product Quantization: 距離近似



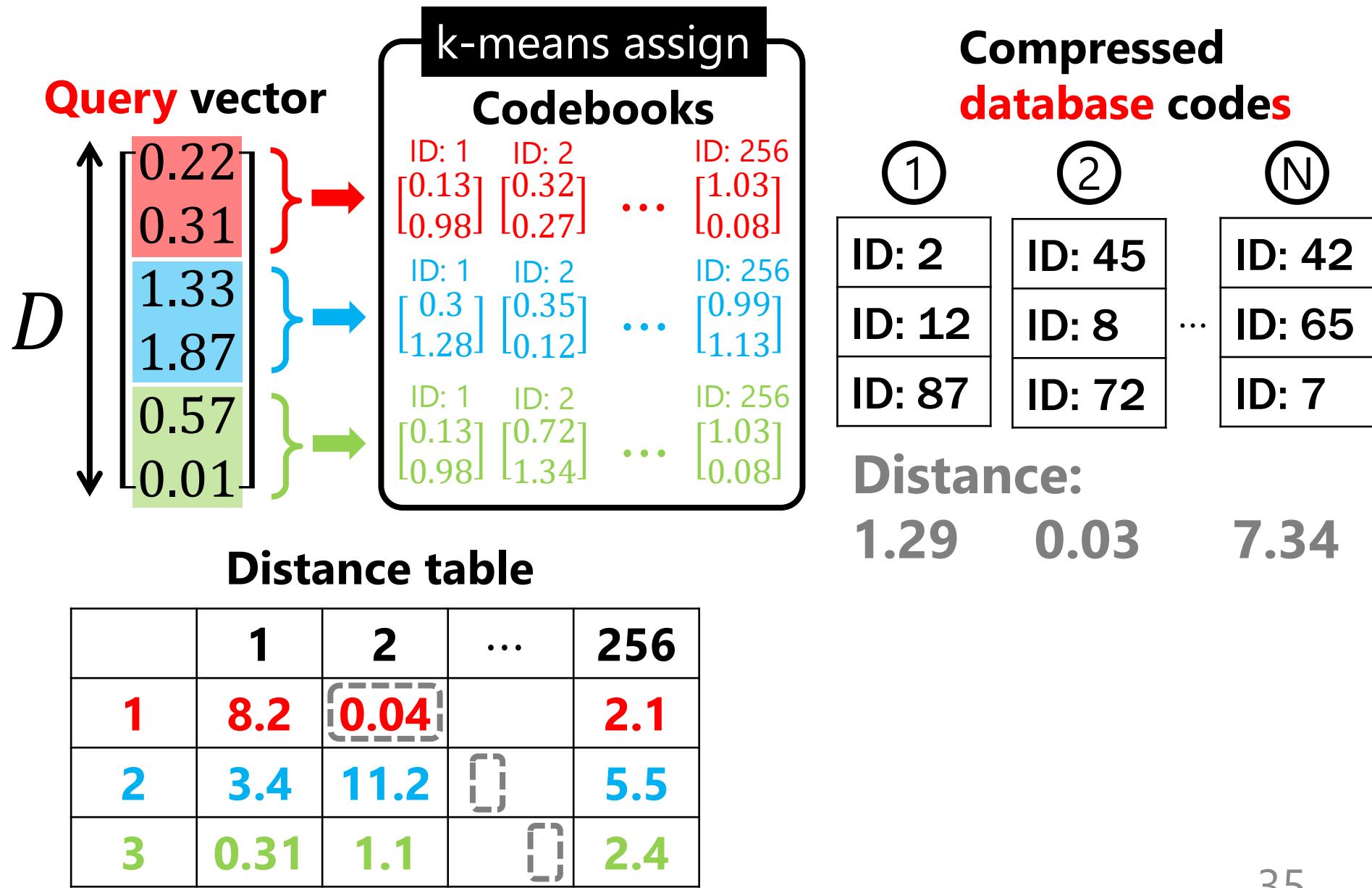
Product Quantization: 距離近似



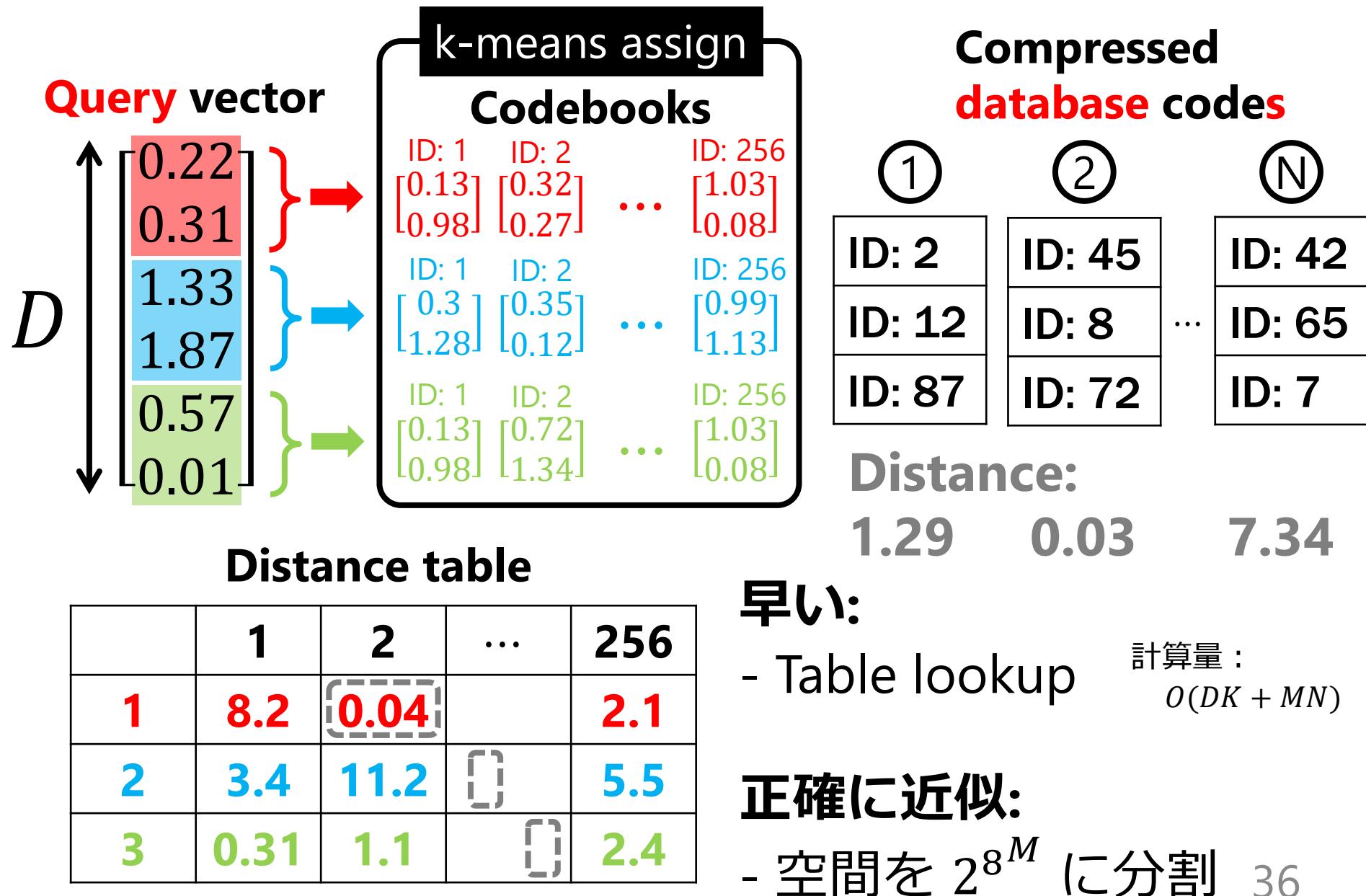
Product Quantization: 距離近似



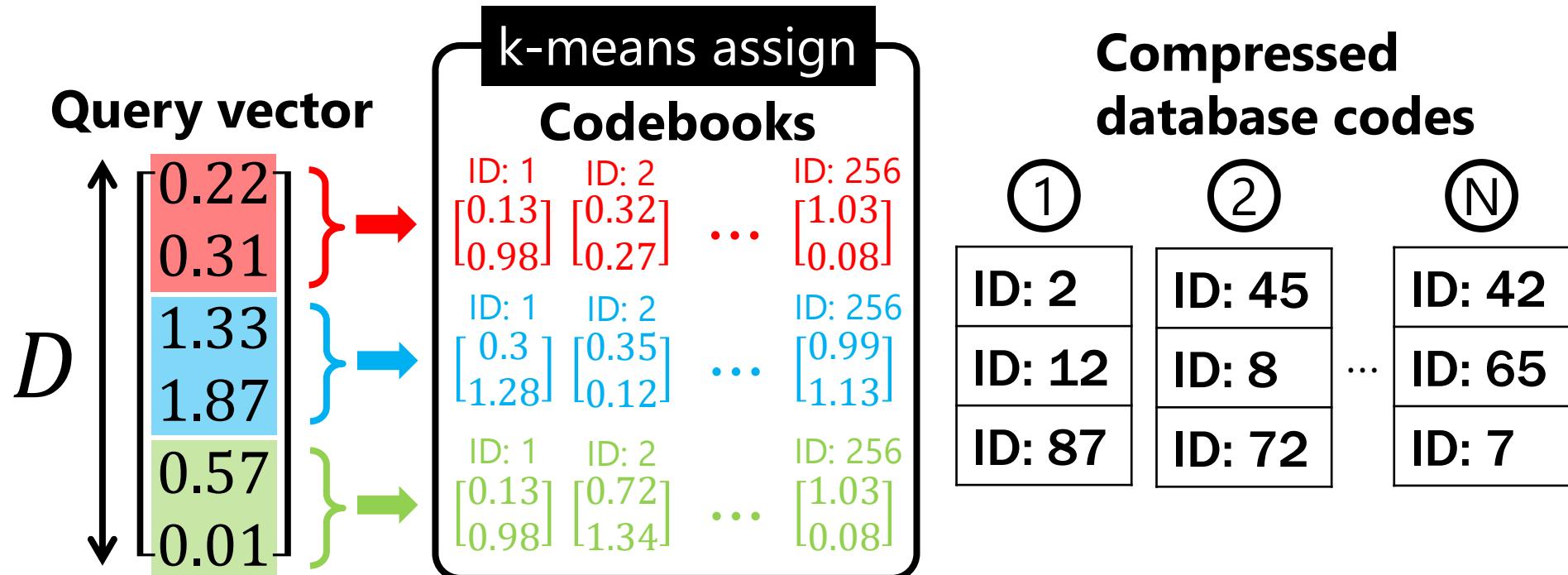
Product Quantization: 距離近似



Product Quantization: 距離近似



Product Quantization



- 単純
- メモリ効率良い
- 距離 $d(\text{input}, \text{code})^2$ を近似計算可能

Product Quantization

```
import numpy as np
from scipy.cluster.vq import vq, kmeans2
from scipy.spatial.distance import cdist

def train(vec, M):
    Ds = int(vec.shape[1] / M) #  $D_s = D / M$ 
    # codeword[m][k] =  $c_k^m$ 
    codeword = np.empty((M, 256, Ds), np.float32)

    for m in range(M):
        vec_sub = vec[:, m * Ds : (m + 1) * Ds]
        codeword[m], label = kmeans2(vec_sub, 256)

    return codeword

def encode(codeword, vec): #  $\text{vec} = \{\mathbf{x}_n\}_{n=1}^N$ 
    M, _K, Ds = codeword.shape
    # pqcode[n] =  $i(\mathbf{x}_n)$ , pqcode[n][m] =  $i^m(\mathbf{x}_n)$ 
    pqcode = np.empty((vec.shape[0], M), np.uint8)

    for m in range(M): # Eq. (2) and Eq. (3)
        vec_sub = vec[:, m * Ds : (m + 1) * Ds]
        pqcode[:, m], dist = vq(vec_sub, codeword[m])

    return pqcode
```

```
def search(codeword, pqcode, query):
    M, _K, Ds = codeword.shape
    # dist_table =  $D(m, k)$ 
    dist_table = np.empty((M, 256), np.float32)

    for m in range(M):
        query_sub = query[m * Ds: (m + 1) * Ds]
        dist_table[m, :] = cdist([query_sub],
                                codeword[m], 'sqeuclidean')[0] # Eq. (5)

    # Eq. (6)
    dist = np.sum(dist_table[range(M), pqcode], axis=1)

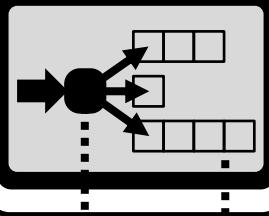
    return dist

if __name__ == "__main__":
    # Read vec_train, vec ( $\{\mathbf{x}_n\}_{n=1}^N$ ), and query (y)
    M = 4
    codeword = train(vec_train, M)
    pqcode = encode(codeword, vec)
    dist = search(codeword, pqcode, query)
    print(dist)
```

➤ 非常に単純

PQそのもの

PQを使った探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの
の発展

一般化

Additive quantization
[Babenko, CVPR 14]

立式が同じ
Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った
探索システムの発展

PQそのもの

PQを使った探索システム

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの
の発展

一般化

Additive quantization
[Babenko, CVPR 14]

立式が同じ、

新しい問題設定

高速符号化 [Zhang, CVPR 15]
教師付き [Wang, CVPR16]

➤ 「ベクトルの分割」が単純すぎないか？

➤ 事前にベクトルを“回転”させ、
分割したときのエラーを最小にする

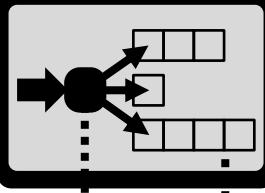
- 1) $x \leftarrow Rx$
- 2) $x \in PQ$

このときの誤差が
最小となるように
回転行列 R を求める

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの 発展

一般化

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

立式が同じ

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

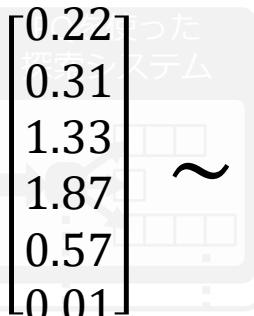
GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

PQを使った 探索システムの発展

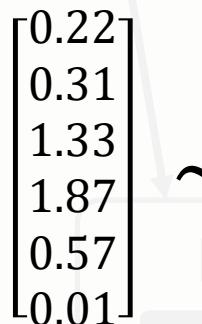
➤ PQは x をM個のD/M次元コードワードで表した



$$\begin{bmatrix} 0.22 \\ 0.31 \\ 1.33 \\ 1.87 \\ 0.57 \\ 0.01 \end{bmatrix} \sim \begin{bmatrix} 0.32 \\ 0.27 \\ 0 \\ 1.19 \\ 1.66 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.32 \\ 1.10 \end{bmatrix}$$

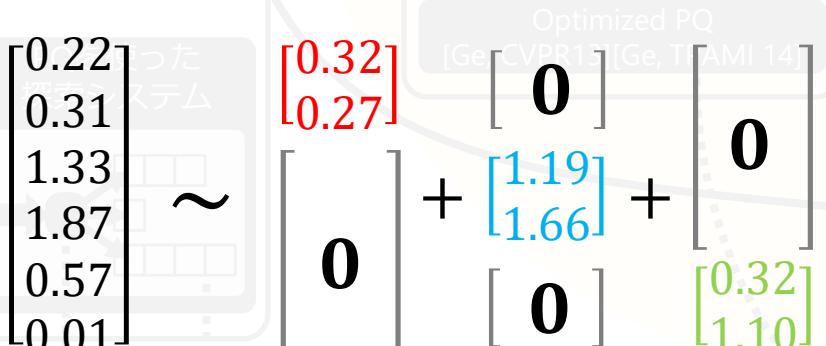
疎量子化：
k-means

➤ AQ・CQはM個のD次元コードワードで表す



$$\begin{bmatrix} 0.22 \\ 0.31 \\ 1.33 \\ 1.87 \\ 0.57 \\ 0.01 \end{bmatrix} \sim \begin{bmatrix} 0.02 \\ 0.13 \\ 0.42 \\ 0.33 \\ 0.12 \\ 0.00 \end{bmatrix} + \begin{bmatrix} 0.19 \\ 0.02 \\ 0.31 \\ 0.01 \\ 0.23 \\ 0.01 \end{bmatrix} + \begin{bmatrix} 0.02 \\ 0.20 \\ 0.56 \\ 1.02 \\ 0.66 \\ 0.10 \end{bmatrix}$$

PQの一般化。表現能力高い
ランクイン側残差考慮 [Heo, CVPR 16]



Optimized PQ
[Ge, CVPR 15] [Ge, TAMI 14]

各次元を複数
コードブックで表す

- Optimized Cartesian k-means [Wang, TKDE 14]
- Tree quantization [Babenko, CVPR 15]

PQそのものの の発展

新しい問題設定

- 高速符号化 [Zhang, CVPR 15]
- 教師付き [Wang, CVPR 16]
- マルチモーダル [Zhang, CVPR 16]

一般化

- Additive quantization [Babenko, CVPR 14]
- Composite quantization [Zhang, ICML 14]

立式が同じ

その他トピック

- 余分ビット [Heo, CVPR 14]
- ハッシュテーブル [Matsui, ICCV 15]
- GPU [Wieschollek, CVPR 16]
- キャッシュ [André, VLDB 15]
- 画像検索システム [Jégou, PAMI 12] [Spyromitros-Xioufis, TMM 14]

PQを使った 探索システムの発展



OPQ+local codebook
[Babenko, TPAMI 15]

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの の発展

一般化

Additive quantization
[Babenko, CVPR 14]

立式が同じ
Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った 探索システムの発展

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

元論文
[Jedra, CVPR 11]

PQそのものの
Given トレーニングデータ,

POを使った
min
コードブック
CQコード

CQの
コスト項

CQ+別の問題:

Given トレーニングデータ,
k-means
リランキン
残差PQコード
min
コードブック
CQコード
その他変数
OPQ+local codebook
[Babenko, TPAMI 15]

- 辞書を疎ベクトルに→量子化を高速化
- ラベルを当てる線形分離変換→教師付き探索
- 画像と文章を同じ空間に変換→マルチモーダル

交互最適化

各次元を複数
コードブックで表す

Optimized Cartesian k-means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの
の発展

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

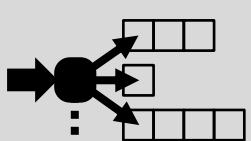
GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

PQそのもの

PQを使った探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの
の発展

一般化

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

立式が同じ

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

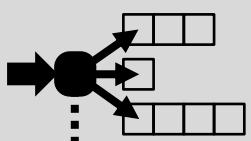
local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った
探索システムの発展

PQそのもの

PQを使った探索システム



疎量子化 :
k-means

リランキン
グ :
残差PQコード

リラン
キン

二段階

local codebook [Kalantidis, CVPR 14]

リランキン
グ側残差考慮 [He et al., TMM 14]

オリジナル論文では
疎量子化 : k-means
**リランキン
グ** :
バケット中心ベクトル
との残差をPQ

- 転置インデックスとPQを組合わせた探索システム
- 疎量子化でざっくりバケットに分け,
PQコードにリランキン
グ
- 高速高精度省メモリ

$\begin{bmatrix} 0.22 \\ 0.31 \\ 1.33 \\ 1.87 \\ 0.57 \\ 0.01 \end{bmatrix}$

疎量子化

ジャーナル版

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, CVPR 14]

245

ID: 42

ID: 37

ID: 9

12

ID: 25

ID: 47

ID: 32

1932

ID: 38

ID: 49

ID: 72

1721

ID: 16

ID: 72

ID: 95

1721

ID: 24

ID: 87

ID: 49

3721

ID: 18

ID: 4

ID: 96

:

8621

ID: 24

ID: 54

ID: 23

145

ID: 77

ID: 21

ID: 5

1721

ID: 32

ID: 11

ID: 85

リランキン
グ

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

オーバーラップ
[Andre, VLDB 15]

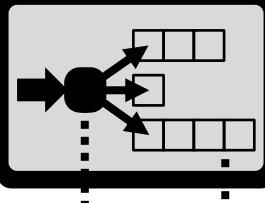
探索システム
[Kang, PAMI 12]

GPU
[Wieschollek, CVPR 16]

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの 発展

一般化

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

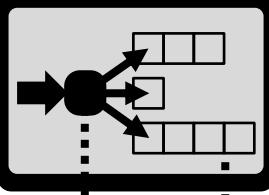
local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った 探索システムの発展

PQそのもの

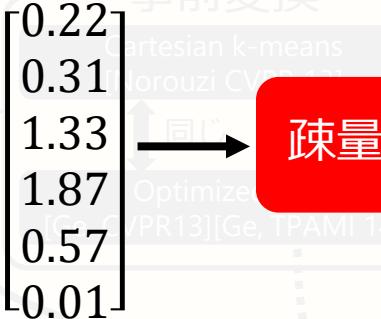
PQを使った探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換



疎量子化

- 疎量子化でもPQを使う
- 近いバケットを「列挙」する
- 速度・精度のバランスが良い

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 15]

Tree quantization
[Babenko, CVPR 14]

ID: 24
ID: 18

ID: 87
ID: 4

ID: 49
ID: 96

1721
3721

PQそのもの
の発展

リランキング

新しい向量改定
高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

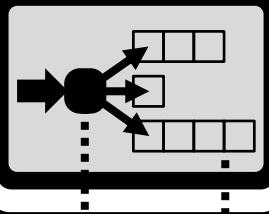
リランキング側残差考慮 [Heo, CVPR 16]

PQを使った 探索システムの発展

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

一般化

Additive quantization
[Babenko, CVPR 14]

立式が同じ
Composite quantization
[Zhang, ICML 14]

PQそのものの 発展

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

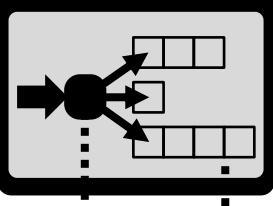
PQを使った 探索システムの発展

元論文

[Jégou, TPAMI 11]

PQそのもの

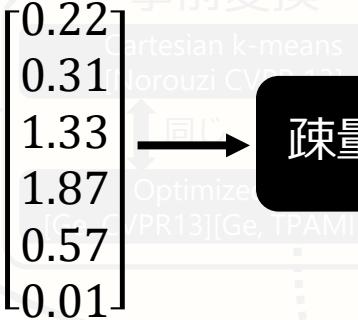
PQを使った探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

事前変換



疎量子化

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 13]
Morouzi CVPR 13

Tree quantization
[Babenko, CVPR 14]

Optimized
[Ge, TPAMI 14]

Cartesian k-means
[Ge, TPAMI 14]

Optimized
[Ge, TPAMI 14]

➤ 全てのバケットで同じコードワードを使う
のではなく、バケットごとに学習する

PQそのものの発展

リランキング

高速符号化 [Zhang, CVPR 15]

新規問題設定 [Zhang, CVPR 16]

高速符号化 [Zhang, CVPR 15]

新規問題設定 [Zhang, CVPR 16]

高速符号化 [Zhang, CVPR 16]

新規問題設定 [Zhang, CVPR 16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

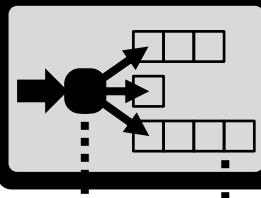
画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

PQを使った探索システムの発展

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

PQそのものの 発展

一般化

Additive quantization
[Babenko, CVPR 14]

Composite quantization
[Zhang, ICML 14]

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

疎量子化の工夫

複数k-means [Xia, ICCV 13]

PQ [Babenko, CVPR 12]

列挙の高速化 [Iwamura, ICCV 13]

残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

[0.22
0.31
1.33
1.87
0.57
0.01]

疎量子化

➤ OPQ

1721

ID: 24
ID: 87
ID: 4
ID: 49

3721

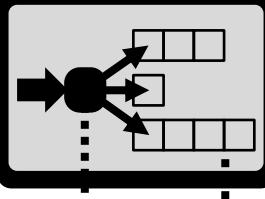
ID: 18
ID: 4
ID: 96

➤ OPQ
➤ local
codebook

元論文 [Jégou, TPAMI 11]

PQそのもの

PQを使った
探索システム



疎量子化 :
k-means

リランキング :
残差PQコード

疎量子化の工夫

- 複数k-means [Xia, ICCV 13]
- PQ [Babenko, CVPR 12]
- 列挙の高速化 [Iwamura, ICCV 13]
- 残差量子化 [Babenko, CVPR 16]

ジャーナル版

OPQ+local codebook
[Babenko, TPAMI 15]

リランキングの工夫

二段階 [Jégou, ICASSP 11]

local codebook [Kalantidis, CVPR 14]

リランキング側残差考慮 [Heo, CVPR 16]

PQを使った 探索システムの発展

事前変換

Cartesian k-means
[Norouzi CVPR 13]

同じ

Optimized PQ
[Ge, CVPR13][Ge, TPAMI 14]

各次元を複数
コードブックで表す

Optimized Cartesian k-
means [Wang, TKDE 14]

Tree quantization
[Babenko, CVPR 15]

一般化

Additive quantization
[Babenko, CVPR 14]

立式が同じ
Composite quantization
[Zhang, ICML 14]

PQそのものの の発展

新しい問題設定

高速符号化 [Zhang, CVPR 15]

教師付き [Wang, CVPR16]

マルチモーダル [Zhang, CVPR16]

その他トピック

余分ビット [Heo, CVPR 14]

ハッシュテーブル
[Matsui, ICCV 15]

GPU
[Wieschollek, CVPR 16]

キャッシュ
[André, VLDB 15]

画像検索システム
[Jégou, PAMI 12]
[Spyromitros-Xioufis, TMM 14]

参考文献

- [Muja, TPAMI 14] M. Muja and D. G. Lowe: "Scalable nearest neighbour algorithms for high dimensional data", IEEE TPAMI (2014).
- [Jégou, TPAMI 11] H. Jégou, M. Douze and C. Schmid: "Product quantization for nearest neighbor search", IEEE TPAMI (2011).
- [Norouzi, CVPR 13] M. Norouzi and D. J. Fleet: "Cartesian k-means", CVPR (2013).
- [Ge, CVPR 13] T. Ge, K. He, Q. Ke and J. Sun: "Optimized product quantization for approximate nearest neighbor search", CVPR (2013).
- [Ge, TPAMI 14] T. Ge, K. He, Q. Ke and J. Sun: "Optimized product quantization", IEEE TPAMI, (2014).
- [Wang, TKDE 14] J. Wang, J. Wang, J. Song, X.-S. Xu, H. T. Shen and S. Li: "Optimized cartesian k-means", IEEE TKDE (2014).
- [Babenko, CVPR 15] A. Babenko and V. Lempitsky: "Tree quantization for large-scale similarity search and classification", CVPR (2015).
- [Babenko, CVPR 14] A. Babenko and V. Lempitsky: "Additive quantization for extreme vector compression", CVPR (2014).
- [Zhang, ICML 14] T. Zhang, C. Du and J. Wang: "Composite quantization for approximate nearest neighbor search", ICML (2014).
- [Zhang, CVPR 15] T. Zhang, G.-J. Qi, J. Tang and J. Wang: "Sparse composite quantization", CVPR (2015).
- [Wang, CVPR16] X. Wang, T. Zhang, G.-J. Qi, J. Tang and J. Wang: "Supervised quantization for similarity search", CVPR (2016).
- [Zhang, CVPR 16] T. Zhang and J. Wang: "Collaborative quantization for cross-modal similarity search", CVPR (2016).
- [Xia, ICCV 13] Y. Xia, K. He, F. Wen and J. Sun: "Joint inverted indexing", ICCV (2013).
- [Babenko, CVPR 12] A. Babenko and V. Lempitsky: "The inverted multi-index", CVPR (2012).
- [Iwamura, ICCV 13] M. Iwamura, T. Sato and K. Kise: "What is the most efficient way to select nearest neighbor candidates for fast approximate nearest neighbor search?", ICCV (2013).
- [Babenko, CVPR 16] A. Babenko and V. Lempitsky: "Efficient indexing of billion-scale datasets of deep descriptors", CVPR (2016).
- [Babenko, TPAMI 15] A. Babenko and V. Lempitsky: "The inverted multi-index", IEEE TPAMI, (2015).
- [Jégou, ICASSP 11] H. Jégou, R. Tavenard, M. Douze and L. Amsaleg: "Searching in one billion vectors: Re-rank with souce coding", ICASSP (2011).
- [Kalantidis, CVPR 14] Y. Kalantidis and Y. Avrithis: "Locally optimized product quantization for approximate nearest neighbor search", CVPR (2014).
- [Heo, CVPR 16] J.-P. Heo, Z. Lin, X. Shen, J. Brandt and S. eui Yoon: "Shortlist selection with residual aware distance estimator for k-nearest neighbor search", CVPR (2016).
- [Heo, CVPR 14] J.-P. Heo, Z. Lin and S.-E. Yoon: "Distance encoded product quantization", CVPR (2014).
- [Matsui, ICCV 15] Y. Matsui, T. Yamasaki and K. Aizawa: "Pqtable: Fast exact asymmetric distance neighbor search for product quantization using hash tables", ICCV (2015).
- [Wieschollek, CVPR 16] P. Wieschollek, O. Wang, A. Sorkine-Hornung and H. P. A. Lensch: "Efficient large-scale approximate nearest neighbor search on the gpu", CVPR (2016).
- [André, VLDB 15] F. André, A.-M. Kermarrec and N. L. Scouarnec: "Cache locality is not enough: High-performance nearest neighbor search with product quantization fast scan", VLDB (2015).
- [Jégou, PAMI 12] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez and C. Schmid: "Aggregating local image descriptors into compact codes", IEEE TPAMI, (2012).
- [Spyromitros-Xioufis, TMM 14] E. Spyromitros-Xioufis, S. Papadopoulos, I. Y. Kompatsiaris, G. Tsoumakas and I. Vlahavas: "A comprehensive study over vlad and product quantization in large-scale image retrieval", IEEE TMM, (2014).