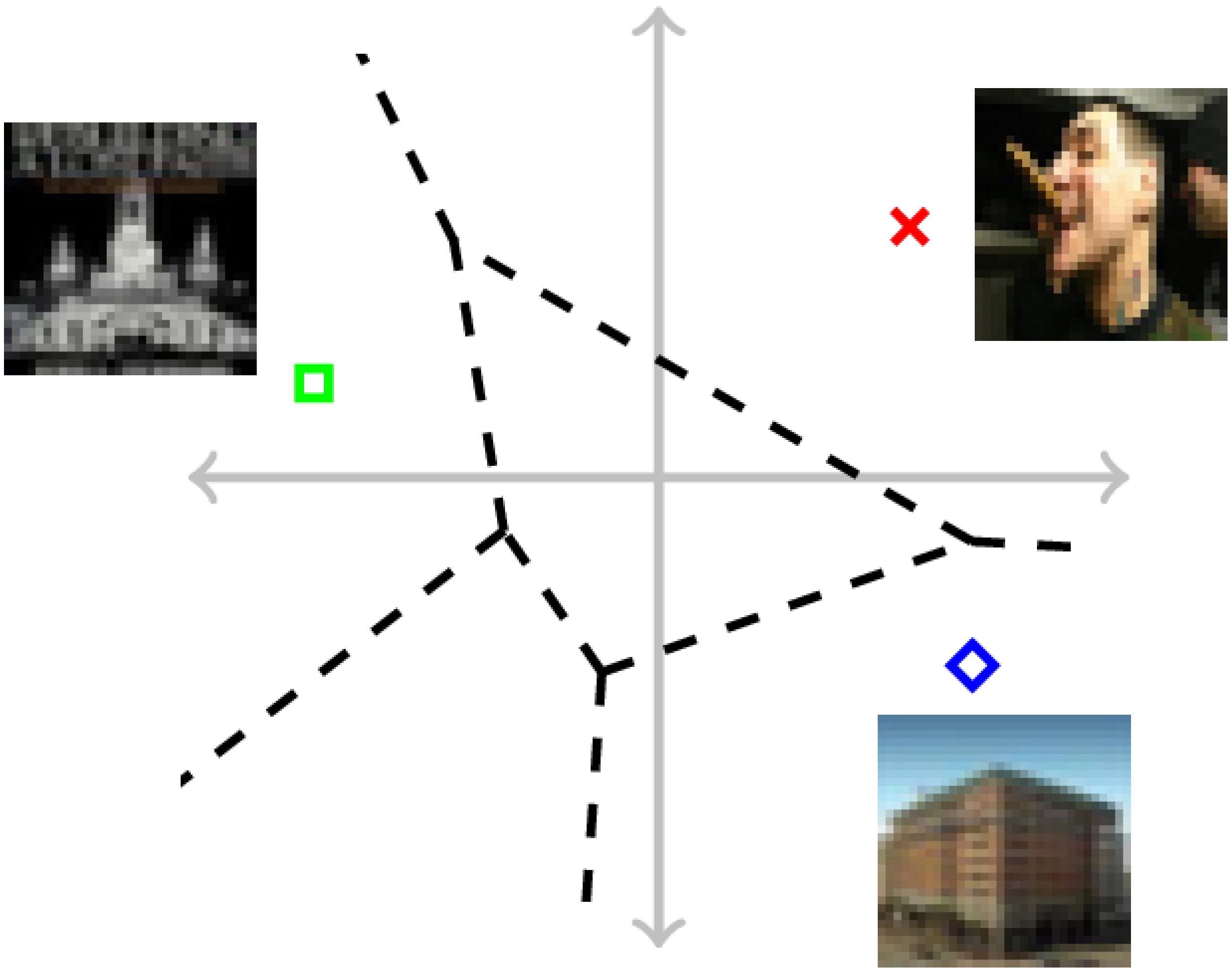
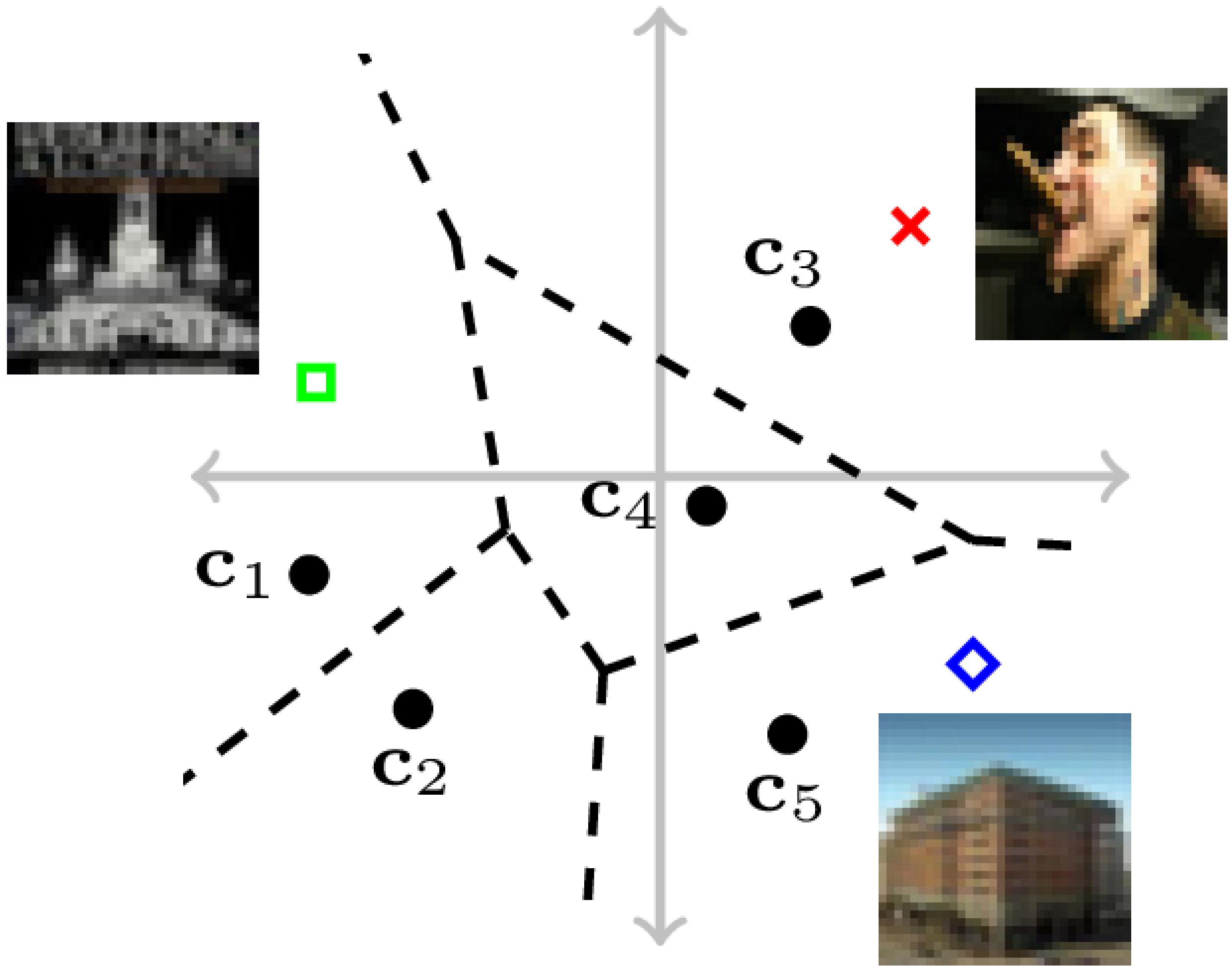
Cartesian k-means

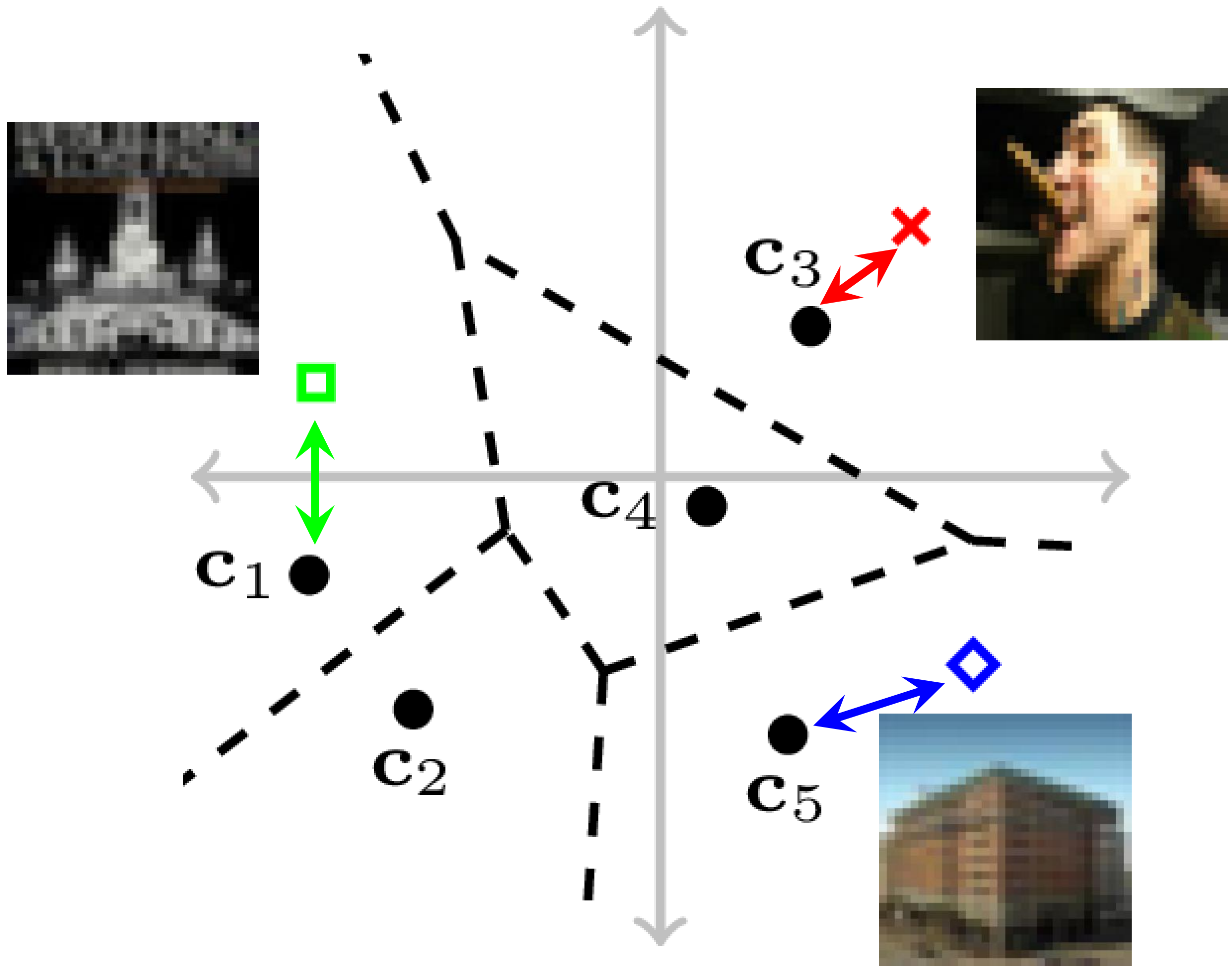
Mohammad Norouzi

David Fleet



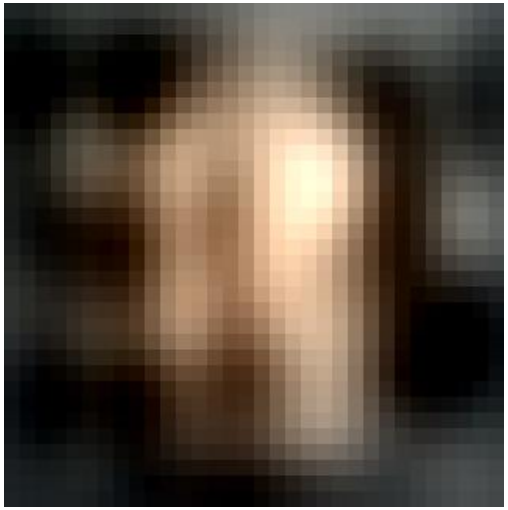


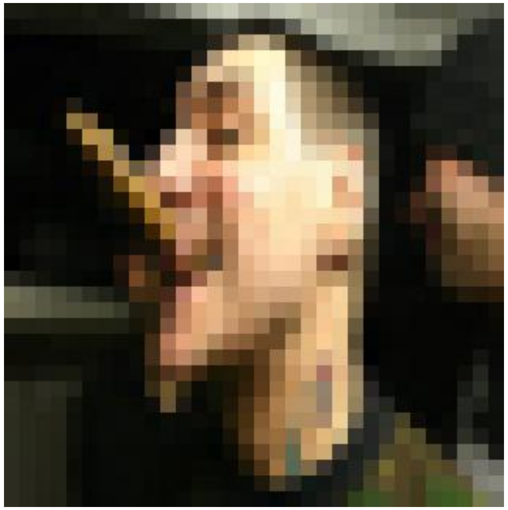
$\mathbf{c}_3$
$\mathbf{c}_4$
$\mathbf{c}_1$
$\mathbf{c}_2$
$\mathbf{c}_5$



$\mathbf{c}_1$
$\mathbf{c}_2$
$\mathbf{c}_3$
$\mathbf{c}_4$
$\mathbf{c}_5$

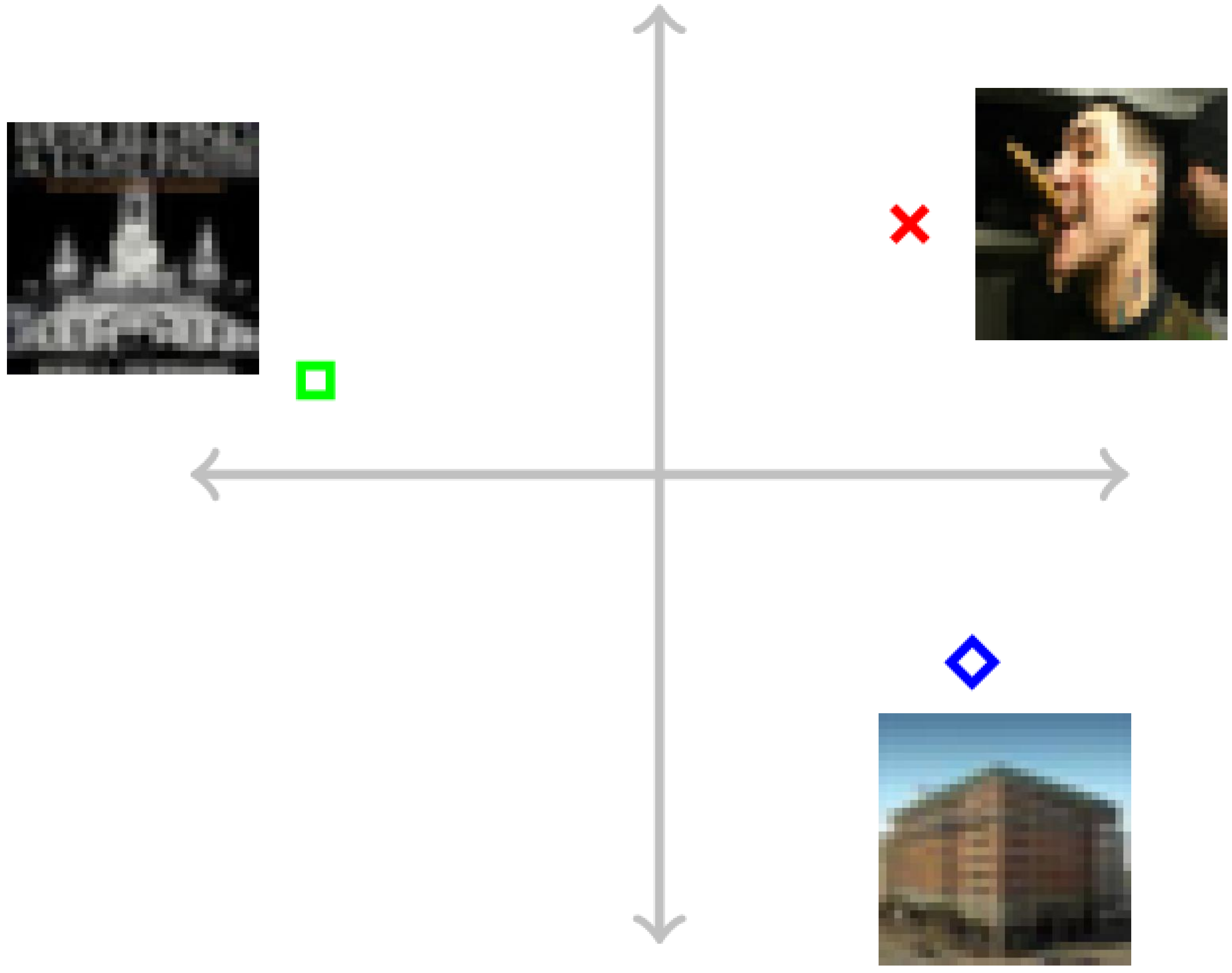
We need many clusters

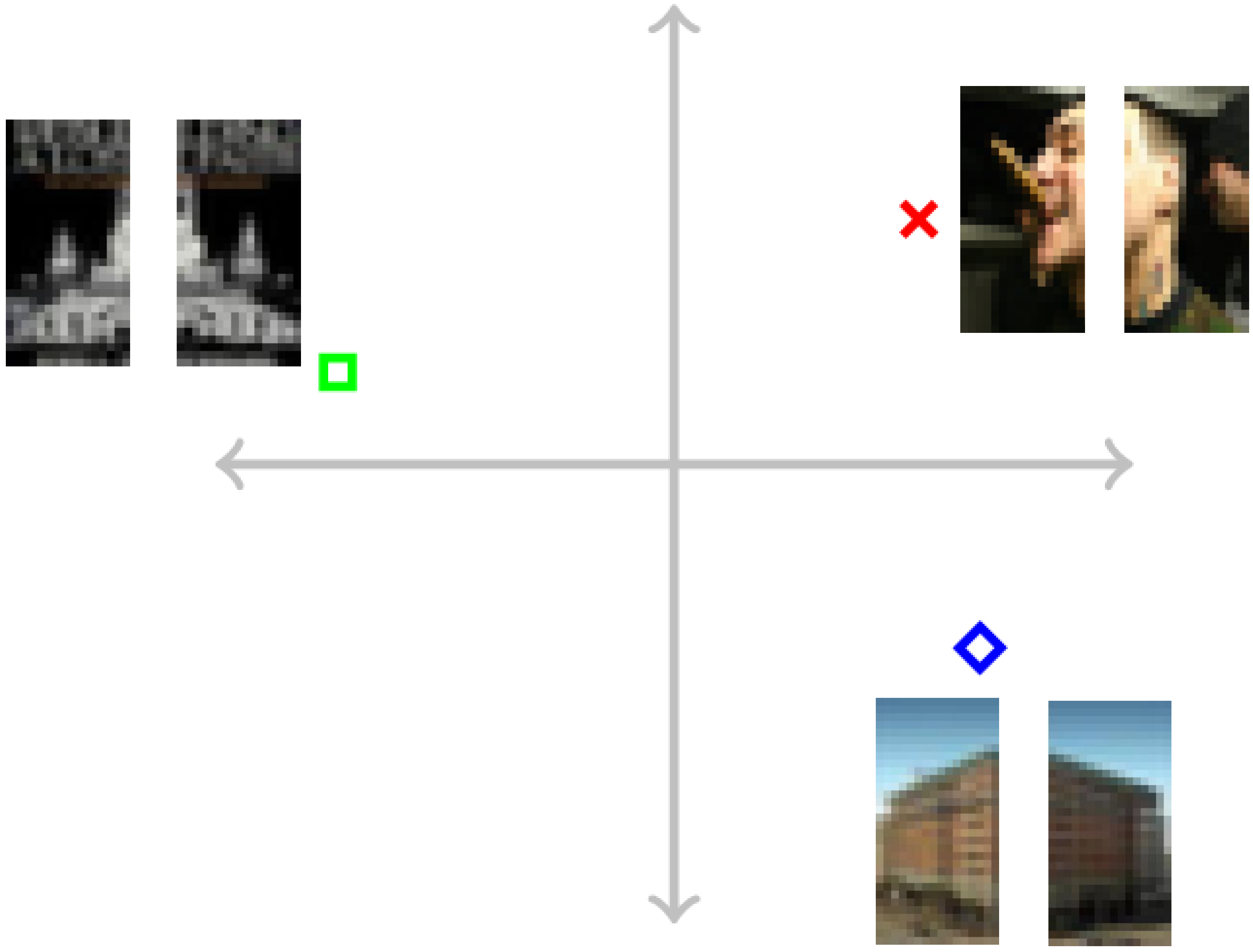


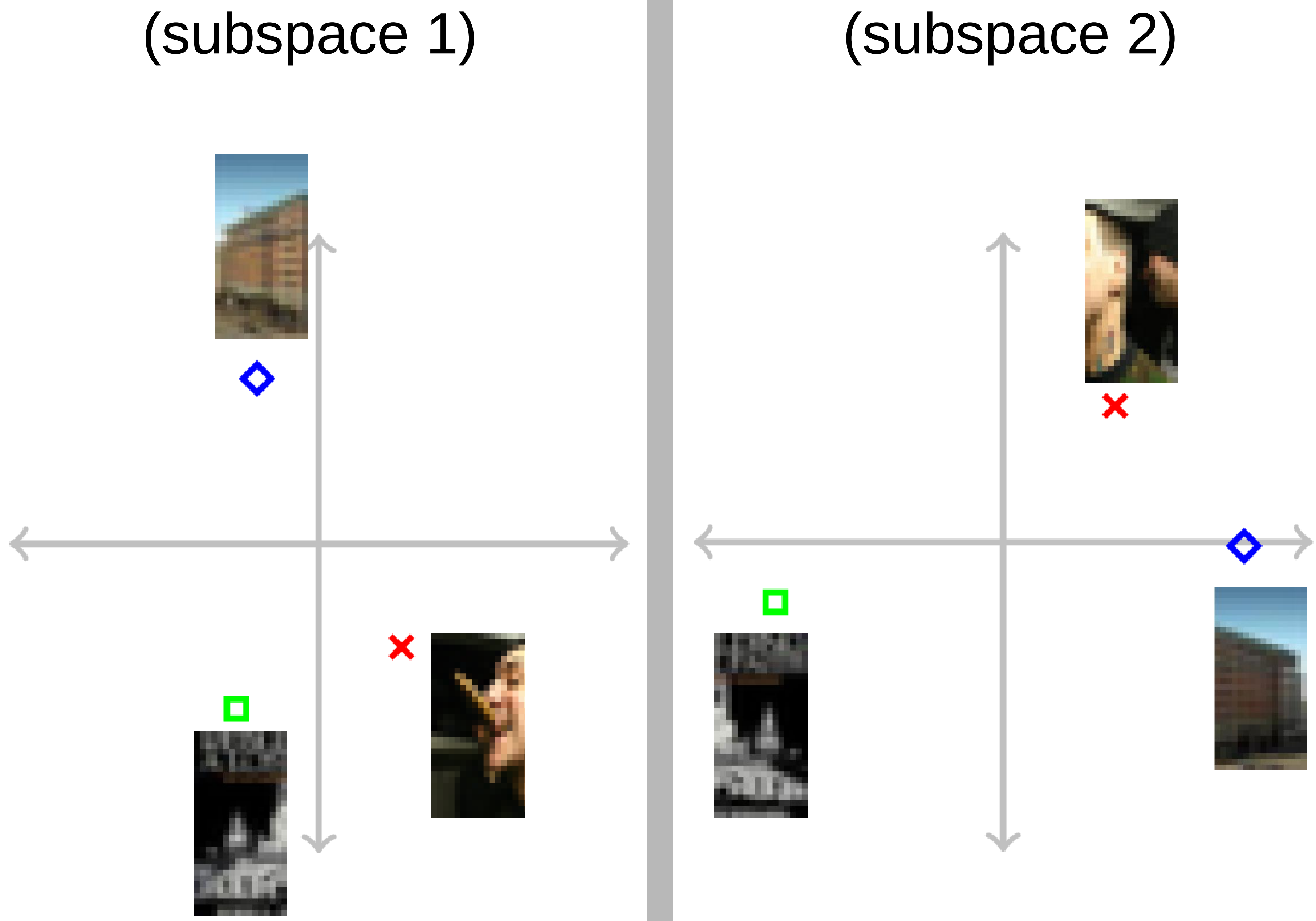


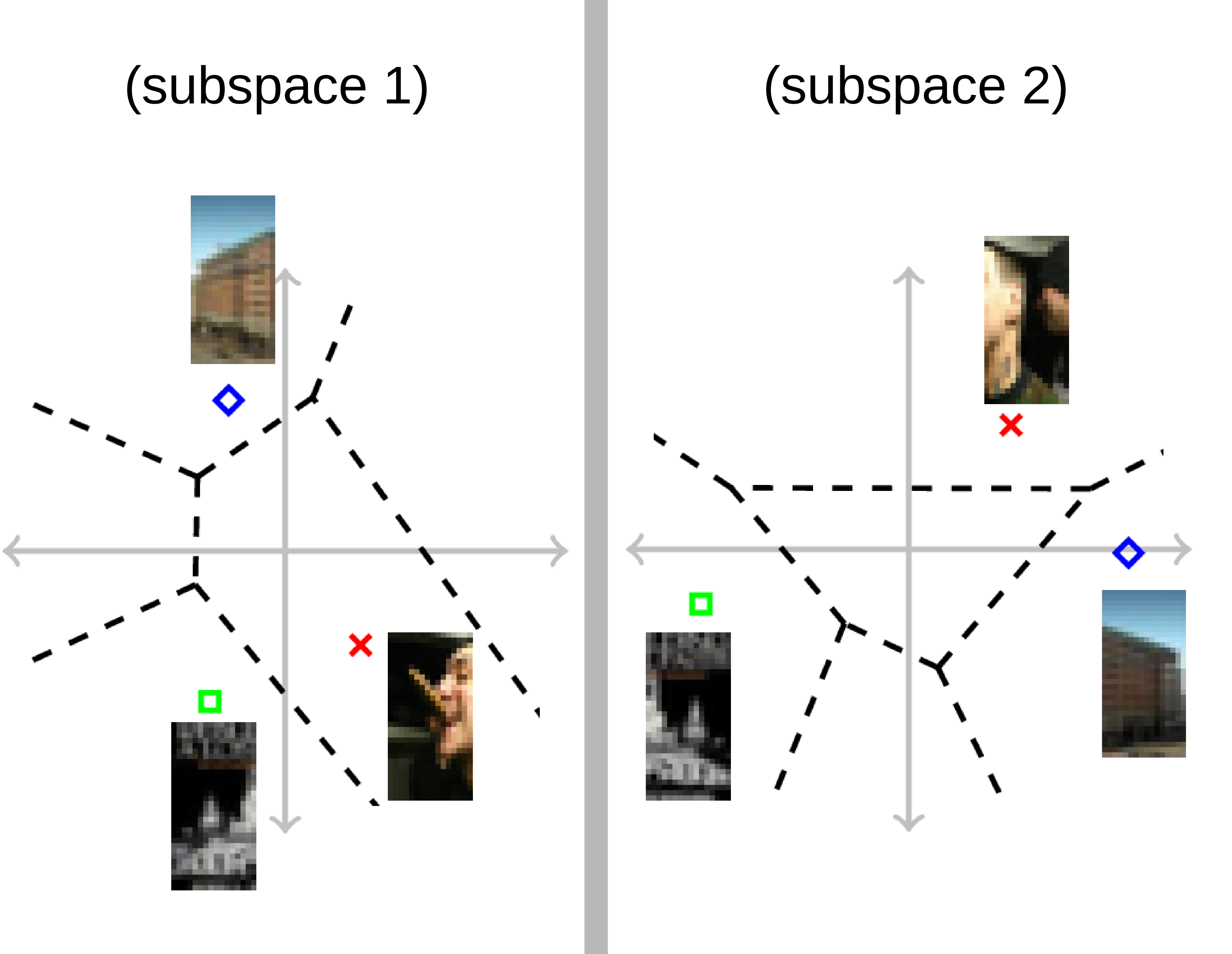
Increasing number of clusters

Problem: Search time, storage cost

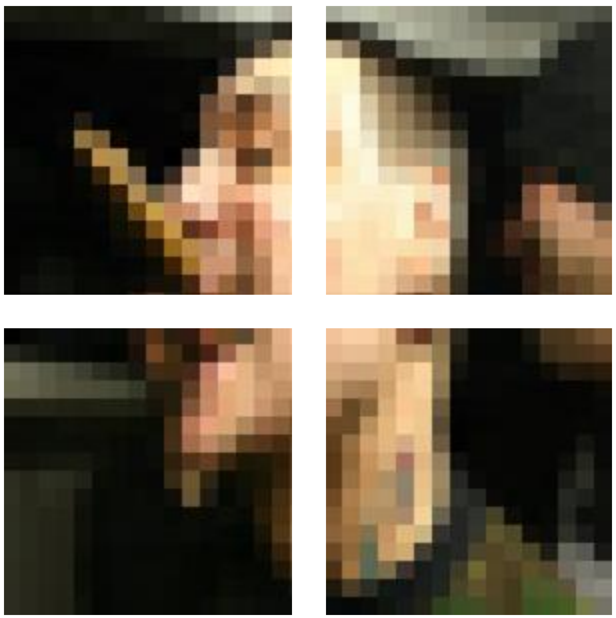





(subspace 1)
(subspace 2)


(subspace 1)
(subspace 2)

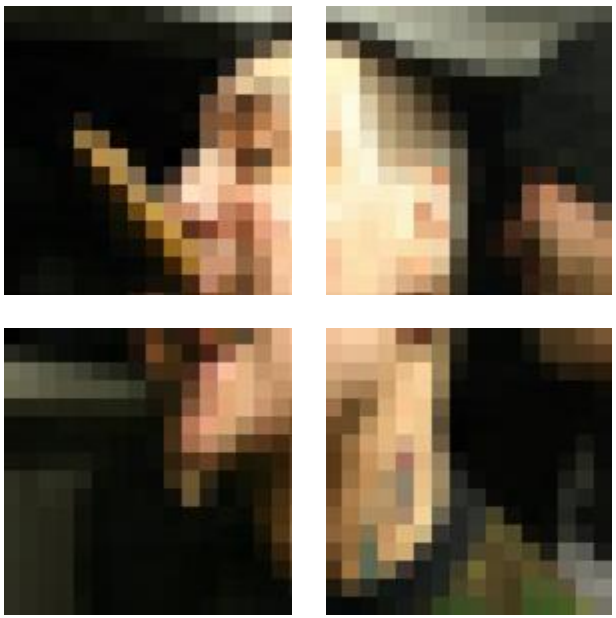
Compositional representation



m subspaces

h regions per subspace

Compositional representation



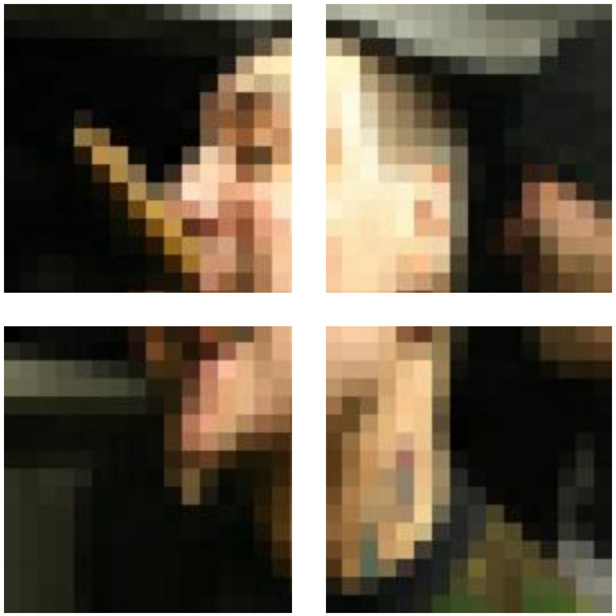
m subspaces

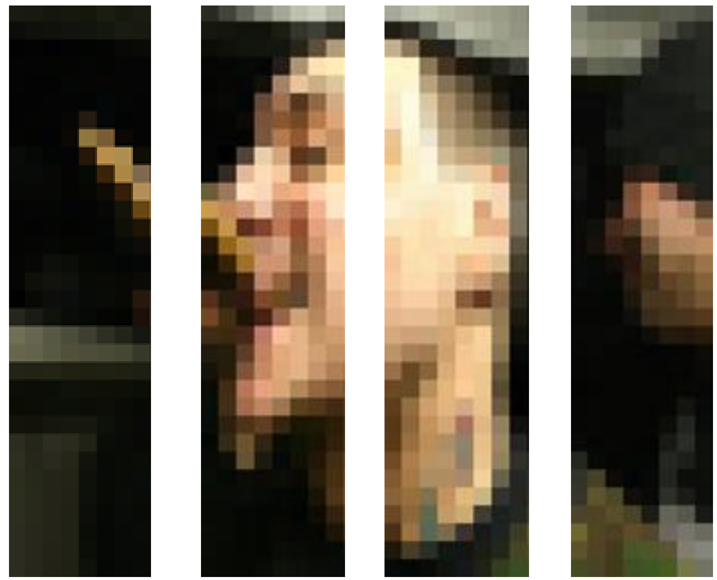
h regions per subspace

$k = h^m$ centers

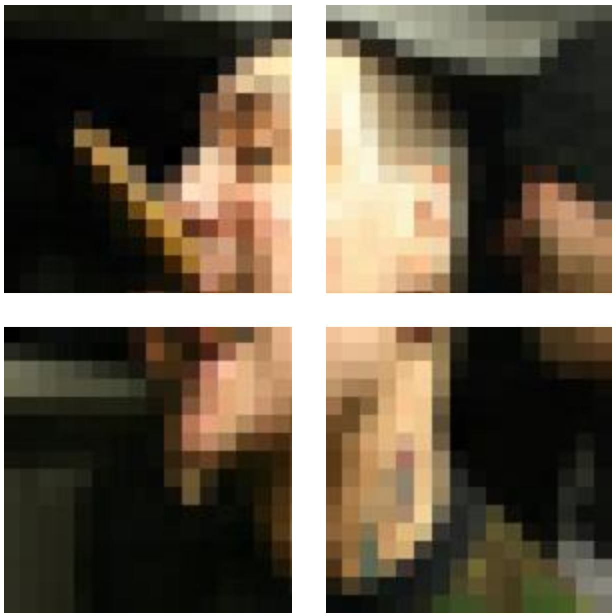
$O(mh)$ parameters

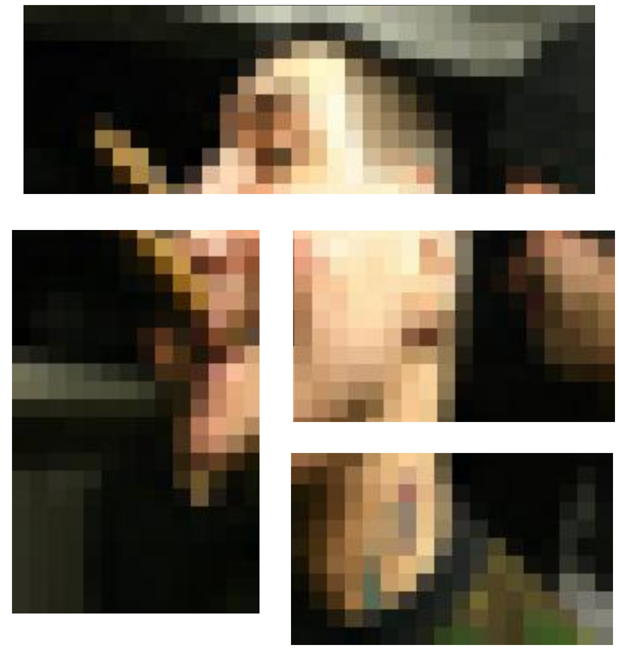
Which subspaces?





Which subspaces? Learning





k-means

k cluster centers: $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k]$

k-means

k cluster centers: $C = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_k]$

$$\ell_{k-means}(C) = \sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b} \in H_1^k} \|\boldsymbol{x} - C\boldsymbol{b}\|^2$$

k-means

k cluster centers: $C = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_k]$

$$\ell_{k-means}(C) = \sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b} \in H_1^k} \|\boldsymbol{x} - C\boldsymbol{b}\|^2$$

$H_1^k \equiv \{\boldsymbol{b} \in \{0,1\}^k \,\big|\, \|\boldsymbol{b}\| = 1\}$ is **one-of-k encoding**

Orthogonal k-means

m center basis vecotrs: $C = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_m]$

$$\ell_{ok-means}(C) = \sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b} \in B^m} \|\boldsymbol{x} - C\boldsymbol{b}\|^2$$

$B^m \equiv \{-1,1\}^m$ is **arbitrary m-bit encoding**

Orthogonal k-means

m center basis vecotrs: $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m]$

$$\ell_{ok-means}(C) = \sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b} \in B^m} \|\boldsymbol{x} - C\boldsymbol{b}\|^2$$

$B^m \equiv \{-1,1\}^m$ is **arbitrary m-bit encoding**

#centers: $k = 2^m$

Orthogonal k-means

m center basis vecotrs: $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m]$

$$\ell_{ok-means}(C) = \sum_{\boldsymbol{x} \in D} \boxed{\min_{\boldsymbol{b} \in B^m} \|\boldsymbol{x} - C\boldsymbol{b}\|^2}$$

$B^m \equiv \{-1,1\}^m$ is **arbitrary m-bit encoding**

#centers: $k = 2^m$

Orthogonal k-means

m center basis vecotrs: $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m]$

$$\ell_{ok-means}(C) = \sum_{\boldsymbol{x} \in D} \boxed{\min_{\boldsymbol{b} \in B^m} \|\boldsymbol{x} - C\boldsymbol{b}\|^2}$$

Additional constraints: $\forall\, i \neq j, c_i \perp c_j$

$\widehat{\boldsymbol{b}} = sgn(C^T \boldsymbol{x})$

Orthogonal k-means

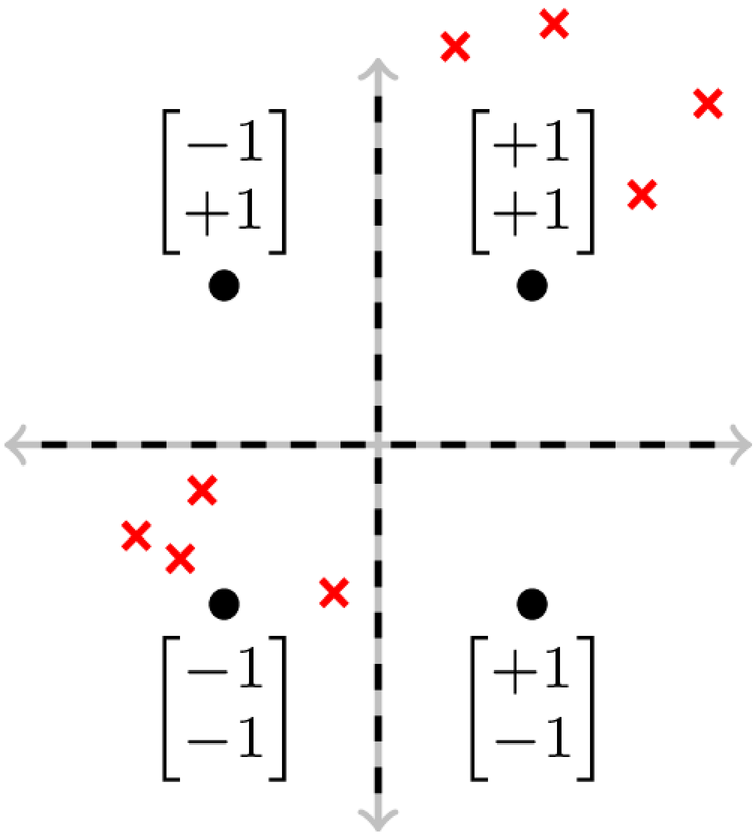
m center basis vecotrs: $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m]$

$$\ell_{ok-means}(C) = \sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b} \in B^m} \|\boldsymbol{x} - C\boldsymbol{b}\|^2$$

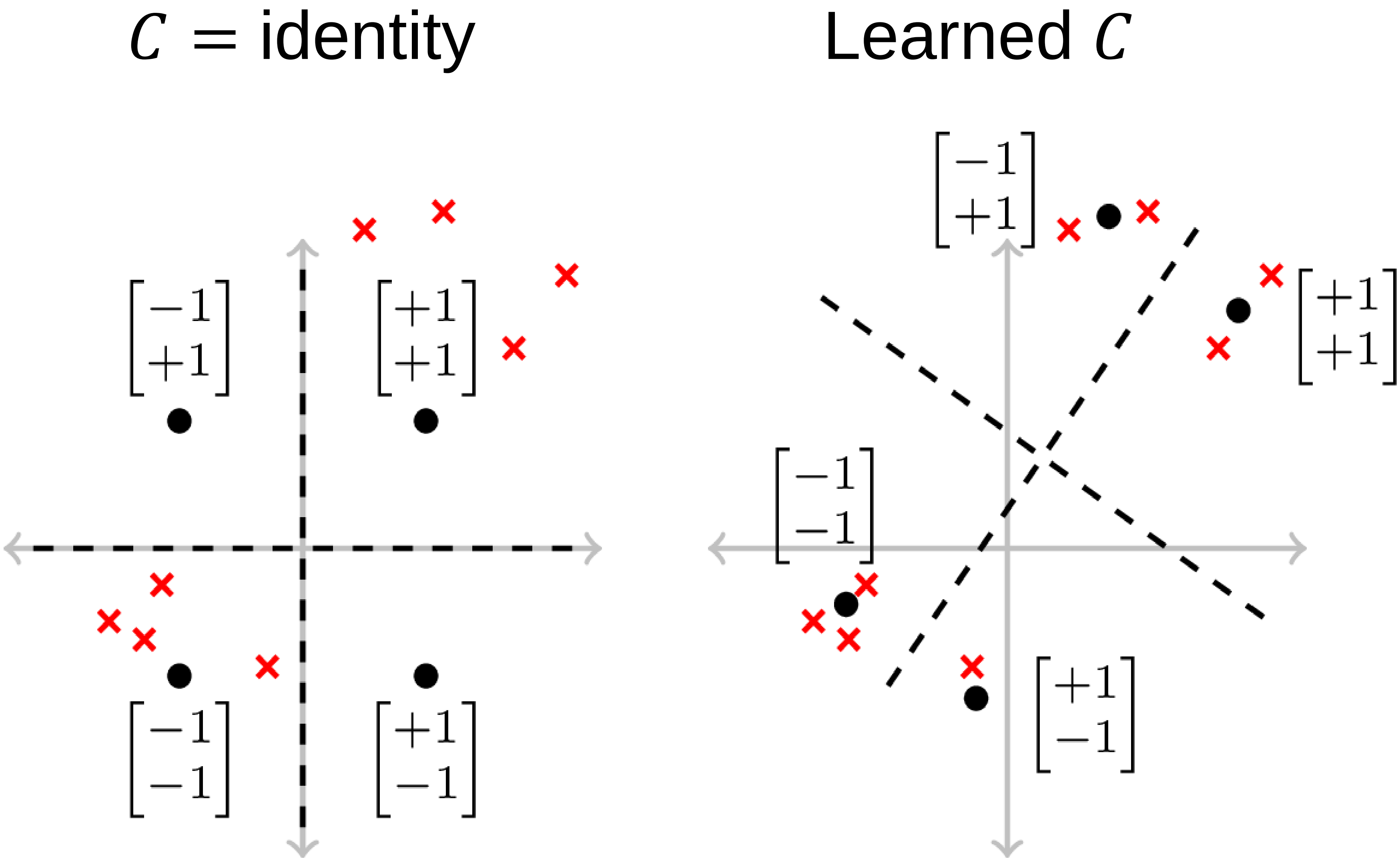
Additional constraints: $\forall\, i \neq j, c_i \perp c_j$

$\widehat{\boldsymbol{b}} = sgn(C^T \boldsymbol{x})$

C = identity

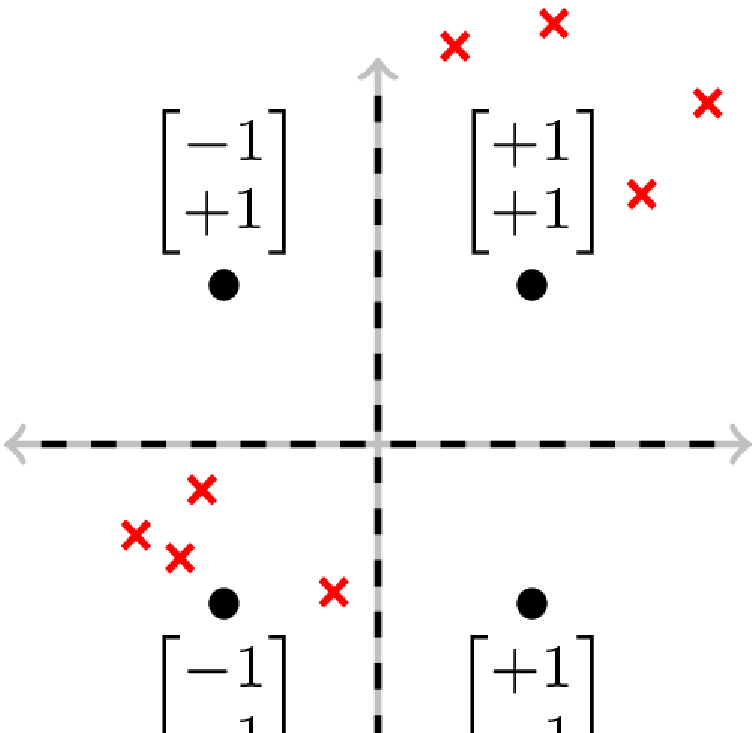


$\begin{bmatrix}-1\\+1\end{bmatrix}$
$\begin{bmatrix}+1\\+1\end{bmatrix}$
$\begin{bmatrix}-1\\-1\end{bmatrix}$
$\begin{bmatrix}+1\\-1\end{bmatrix}$

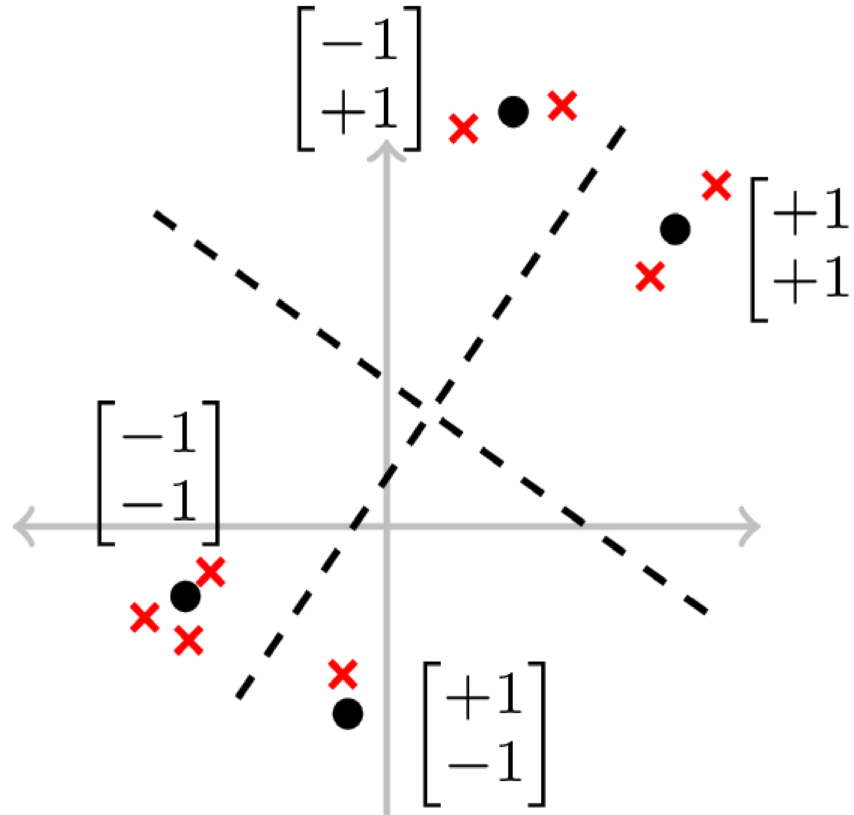


C = identity
Learned C
$\begin{bmatrix}-1\\+1\end{bmatrix}$
$\begin{bmatrix}+1\\+1\end{bmatrix}$
$\begin{bmatrix}-1\\-1\end{bmatrix}$
$\begin{bmatrix}+1\\-1\end{bmatrix}$
$\begin{bmatrix}-1\\+1\end{bmatrix}$
$\begin{bmatrix}+1\\+1\end{bmatrix}$
$\begin{bmatrix}-1\\-1\end{bmatrix}$
$\begin{bmatrix}+1\\-1\end{bmatrix}$

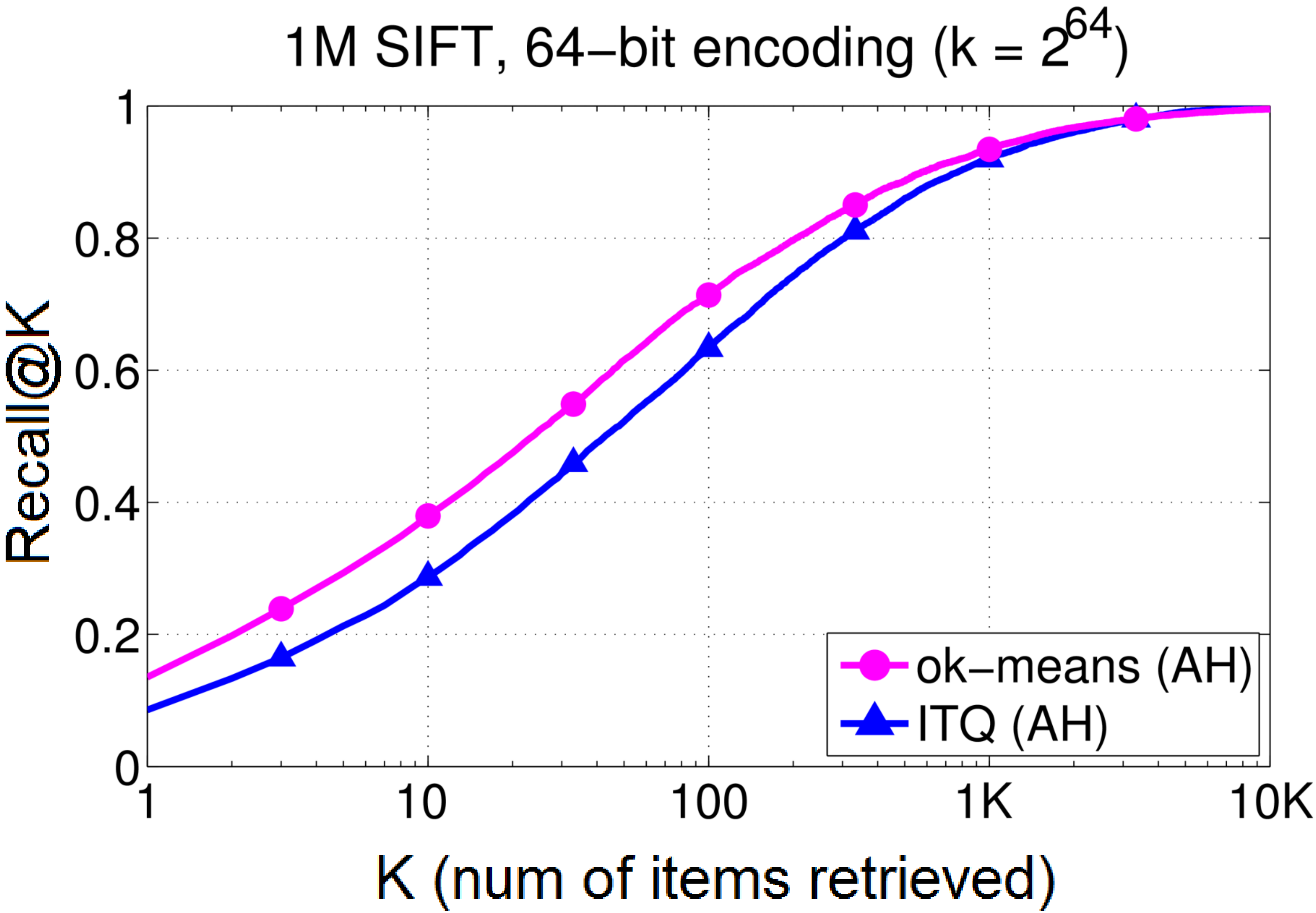
C = identity

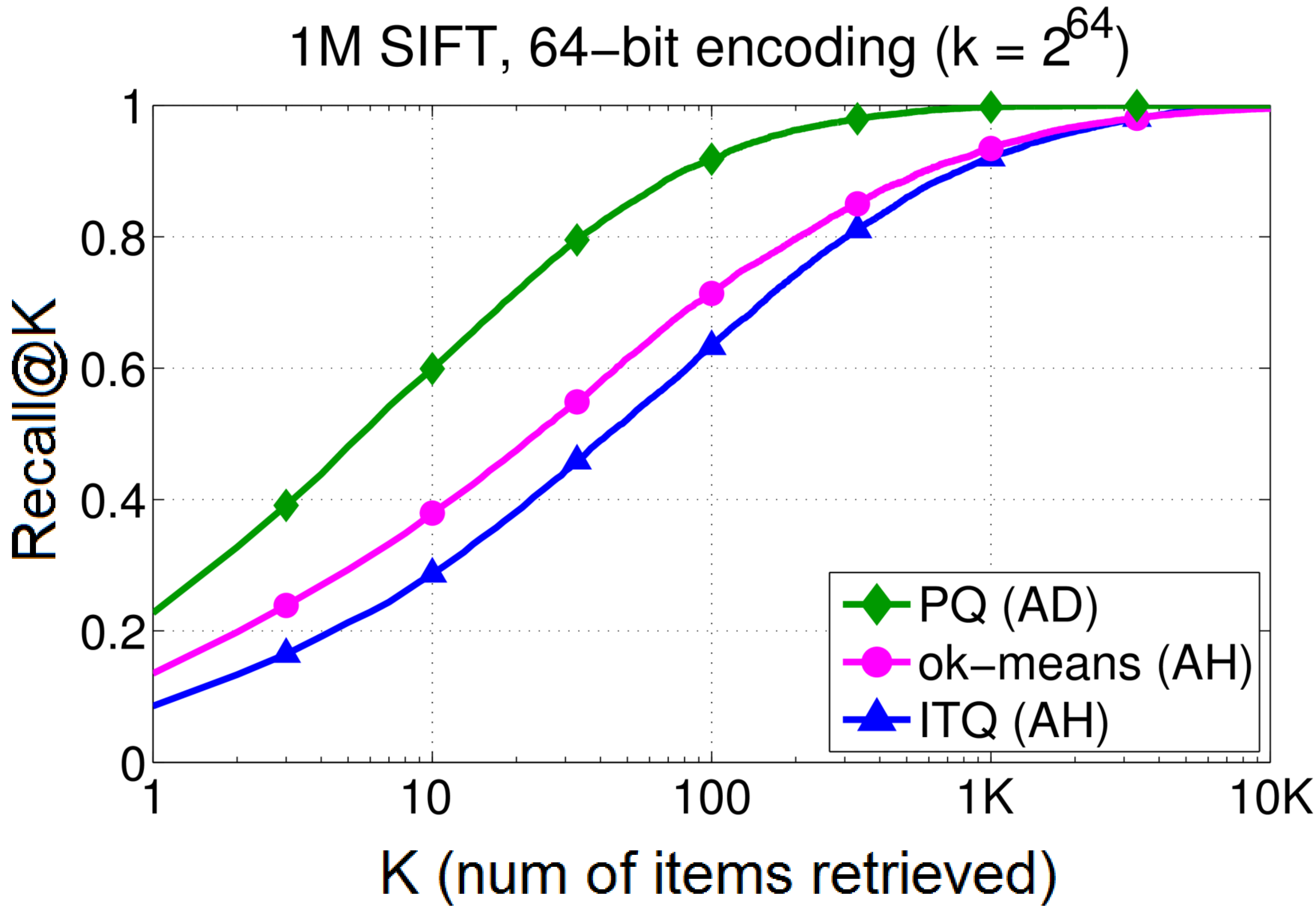


Learned C

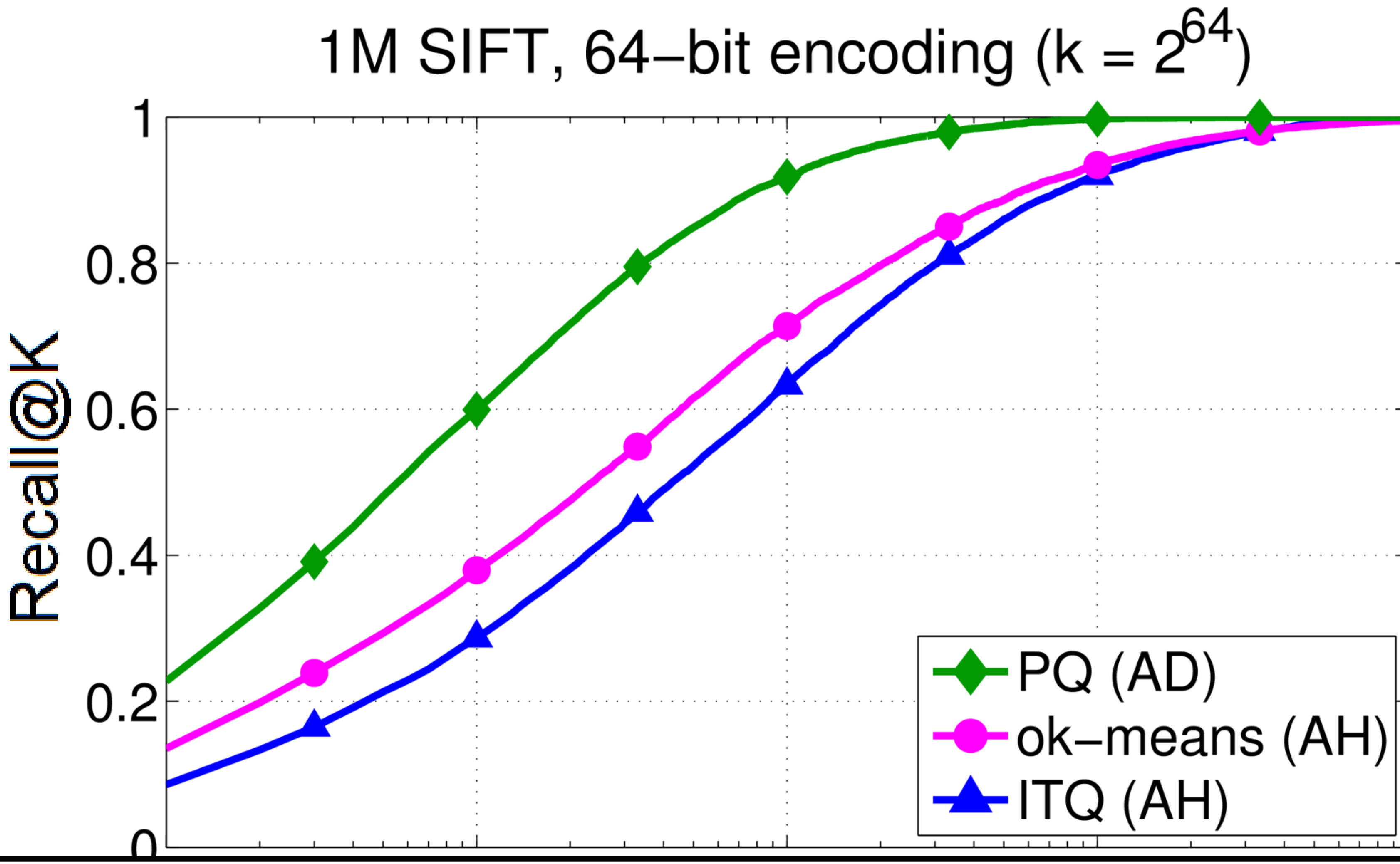


Iterative Quantization *[Gong & Lazebnik, CVPR'11]*


1M SIFT, 64−bit encoding (k = 2^64)
Recall@K
K (num of items retrieved)
0
0.2
0.4
0.6
0.8
1
1
10
100
1K
10K
ok−means (AH)
ITQ (AH)



1M SIFT, 64−bit encoding (k = 2^{64})
Recall@K
K (num of items retrieved)
0
0.2
0.4
0.6
0.8
1
1
10
100
1K
10K
PQ (AD)
ok−means (AH)
ITQ (AH)



Product Quantization *[Jegou, Douze, Schmid, PAMI'11]*

Cartesian k-means

$$\begin{bmatrix} \\ x \\ \\ \end{bmatrix} \approx \begin{bmatrix} \\ C^{(1)} & C^{(2)} \\ \\ \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \\ \boldsymbol{b}^{(2)} \end{bmatrix}$$

$$C^{(1)} \perp C^{(2)}$$

Cartesian k-means

$$\begin{bmatrix} x \end{bmatrix} \approx \begin{bmatrix} \underbrace{C^{(1)}}_{h} & \underbrace{C^{(2)}}_{h} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix}$$

$$C^{(1)} \perp C^{(2)}$$

Cartesian k-means

$$\begin{bmatrix} x \end{bmatrix} \approx \begin{bmatrix} \underbrace{C^{(1)}}_{h} & \underbrace{C^{(2)}}_{h} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \begin{array}{l} \left.\right\} \text{one-of-}h \\ \left.\right\} \text{one-of-}h \end{array}$$

$$C^{(1)} \perp C^{(2)}$$

$$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$$

Cartesian k-means

$$\begin{bmatrix} x \end{bmatrix} \approx \begin{bmatrix} \underbrace{C^{(1)}}_{h} & \underbrace{C^{(2)}}_{h} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \begin{matrix} \left.\right\} \text{one-of-}h \\ \left.\right\} \text{one-of-}h \end{matrix}$$

$$C^{(1)} \perp C^{(2)}$$

$$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$$

$\boxed{\text{\#centers: } k = h^2}$

Cartesian k-means

$$\begin{bmatrix} x \end{bmatrix} \approx \begin{bmatrix} \underbrace{C^{(1)}}_{h} & \underbrace{C^{(2)}}_{h} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \begin{matrix} \}\ \text{one-of-}h \\ \}\ \text{one-of-}h \end{matrix}$$

→ $C^{(1)} \perp C^{(2)}$

$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$

#centers: $k = h^2$

Storage cost: $O(\sqrt{k})$

Search time: $O(\sqrt{k})$

Learning Cartesian k-means

$$\sum_{\boldsymbol{x}\in D} \min_{\boldsymbol{b}^{(1)},\boldsymbol{b}^{(2)}} \left\| \boldsymbol{x} - \begin{bmatrix} C^{(1)} & C^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \right\|^2$$

$$C^{(1)} \perp C^{(2)}$$

$$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$$

Learning Cartesian k-means

$$\sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}} \left\| \boldsymbol{x} - \begin{bmatrix} R^{(1)} & R^{(2)} \end{bmatrix} \begin{bmatrix} D^{(1)} & 0 \\ 0 & D^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \right\|^2$$

$$R^{(1)} \perp R^{(2)}$$

$$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$$

Learning Cartesian k-means

$$\sum_{\boldsymbol{x}\in D} \min_{\boldsymbol{b}^{(1)},\boldsymbol{b}^{(2)}} \left\| \boldsymbol{x} - \begin{bmatrix} R^{(1)} & R^{(2)} \end{bmatrix} \begin{bmatrix} D^{(1)} & 0 \\ 0 & D^{(2)} \end{bmatrix} \underline{\begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix}} \right\|^2$$

$R^{(1)} \perp R^{(2)}$

$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$

Finding optimal b by two Independent NNS

Learning Cartesian k-means

$$\sum_{\boldsymbol{x} \in D} \min_{\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}} \left\| \boldsymbol{x} - \begin{bmatrix} R^{(1)} & R^{(2)} \end{bmatrix} \underline{\begin{bmatrix} D^{(1)} & 0 \\ 0 & D^{(2)} \end{bmatrix}} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \right\|^2$$

$$R^{(1)} \perp R^{(2)}$$

$$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$$

Update D by one step of k-means in each subspace

Learning Cartesian k-means

$$\sum_{\boldsymbol{x}\in D} \min_{\boldsymbol{b}^{(1)},\boldsymbol{b}^{(2)}} \left\| \boldsymbol{x} - \underline{\begin{bmatrix} R^{(1)} & R^{(2)} \end{bmatrix}} \begin{bmatrix} D^{(1)} & 0 \\ 0 & D^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \boldsymbol{b}^{(2)} \end{bmatrix} \right\|^2$$

$$R^{(1)} \perp R^{(2)}$$

$$\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \in H_1^h$$

Update R by *SVD* to solve Orthogonal procrustes

Cartesian k-means

$$\begin{bmatrix} x \end{bmatrix} \approx \begin{bmatrix} C^{(1)} & \dots & C^{(m)} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(1)} \\ \vdots \\ \boldsymbol{b}^{(m)} \end{bmatrix} \begin{matrix} \left.\vphantom{\boldsymbol{b}^{(1)}}\right\} \text{one-of-}h \\ \\ \left.\vphantom{\boldsymbol{b}^{(m)}}\right\} \text{one-of-}h \end{matrix}$$

$$\forall\, i \neq j \quad C^{(i)} \perp C^{(j)}$$

$$\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(m)} \in H_1^h$$

#centers: $k = h^m$

Storage cost: $O(\sqrt[m]{k})$

Search time: $O(\sqrt[m]{k})$

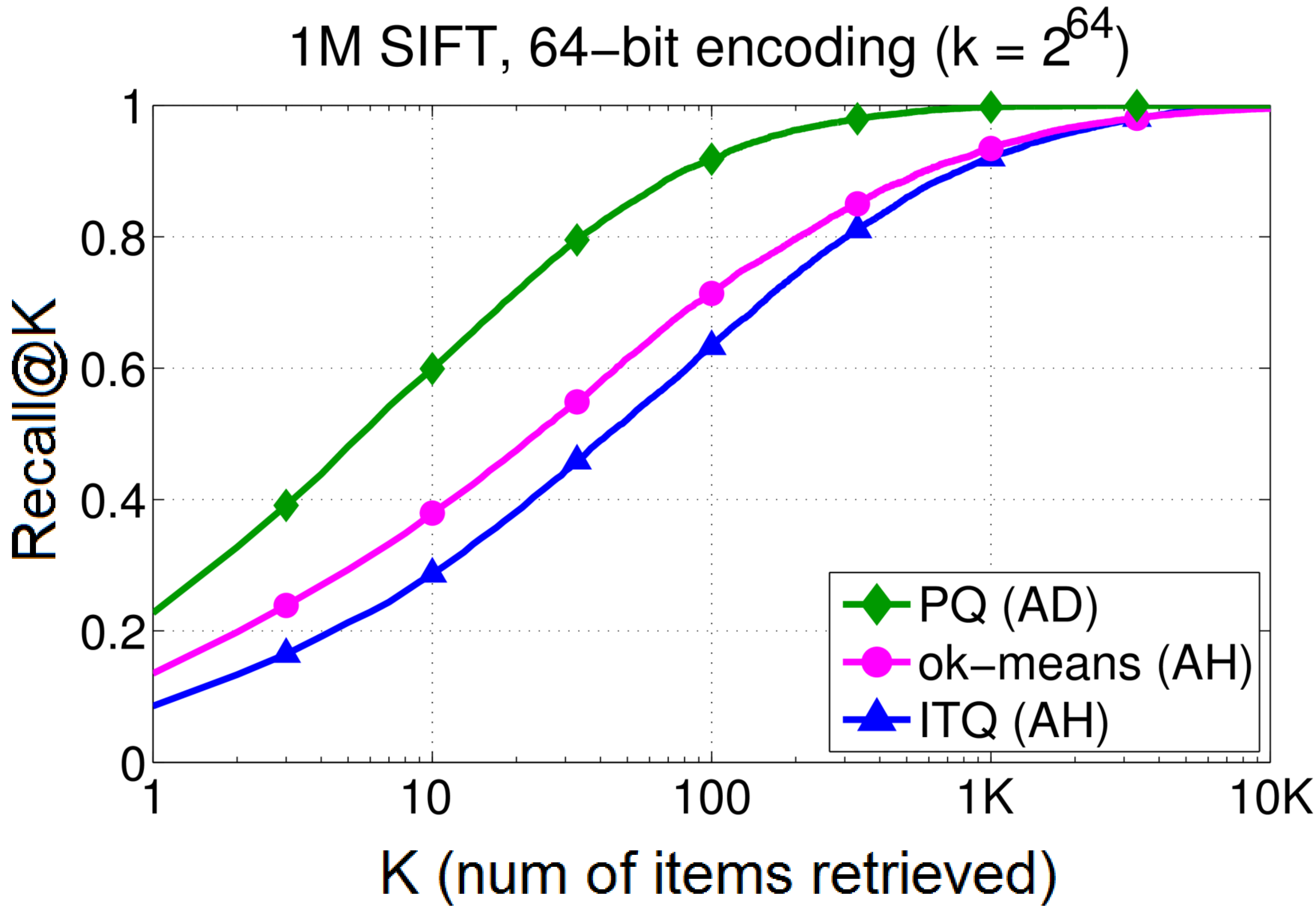
Cartesian k-means

m subspaces, h regions per subspace

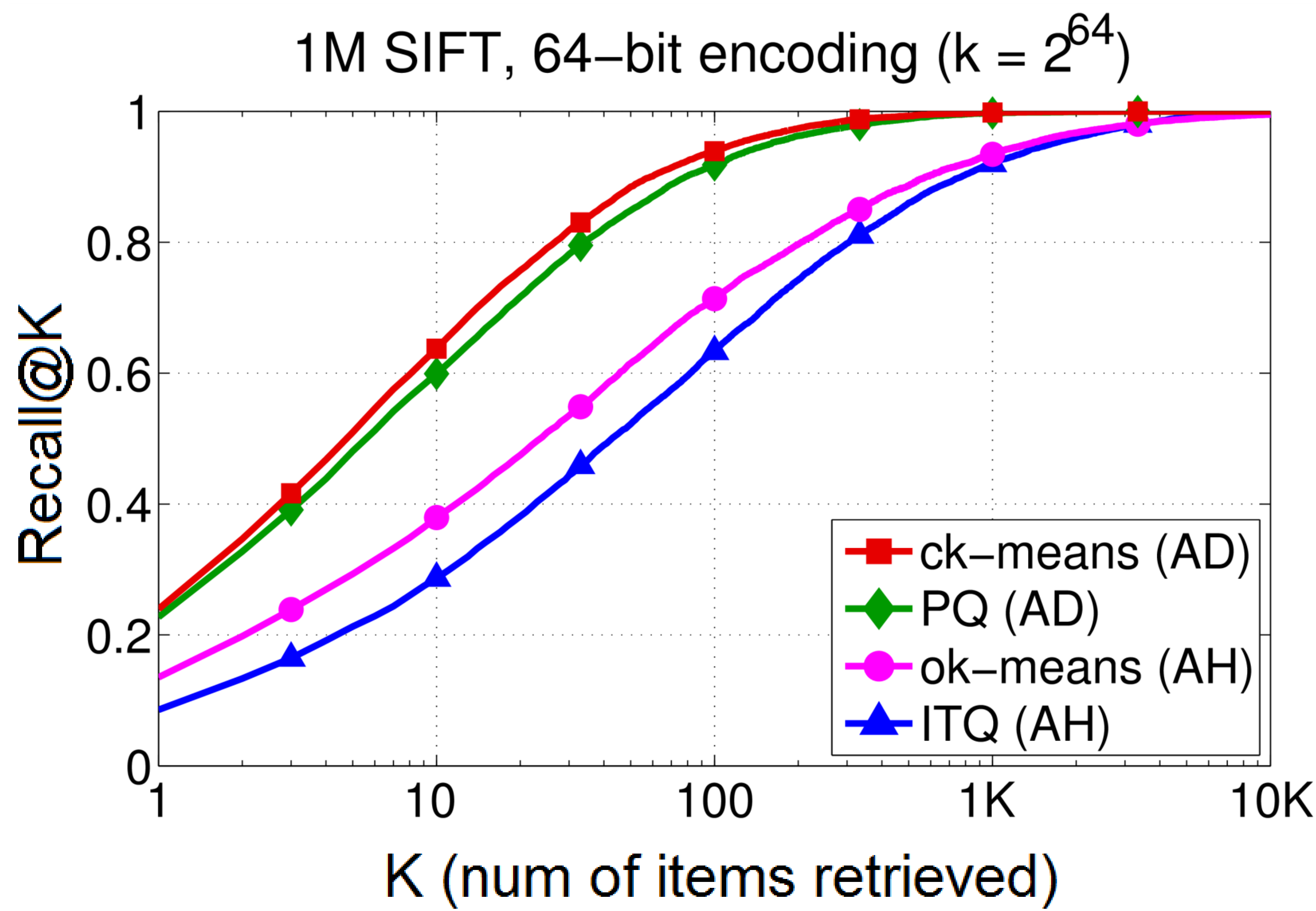
ok-means
$h = 2$
$k = 2^m$

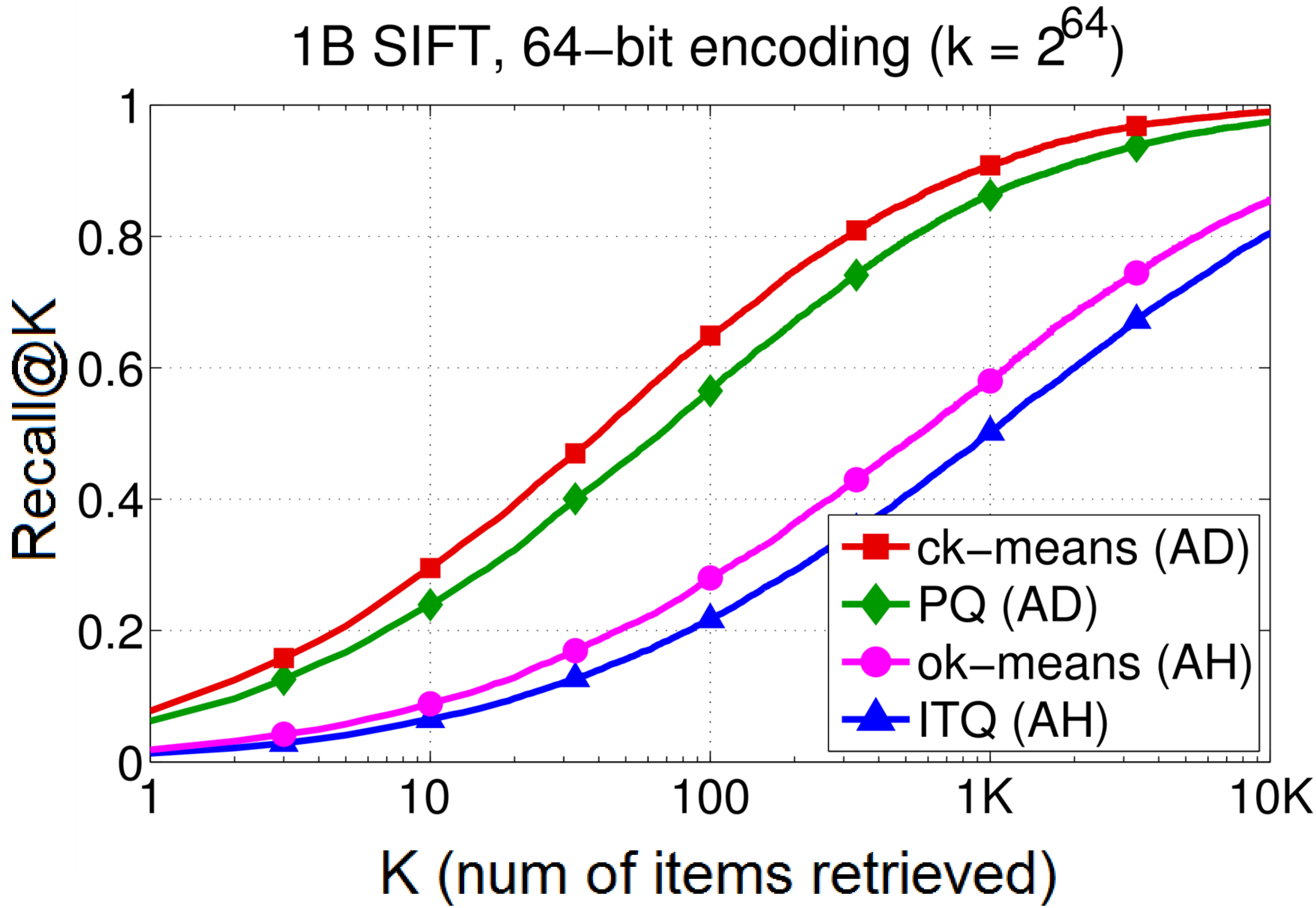
$m = 1$
k-means

compositionality

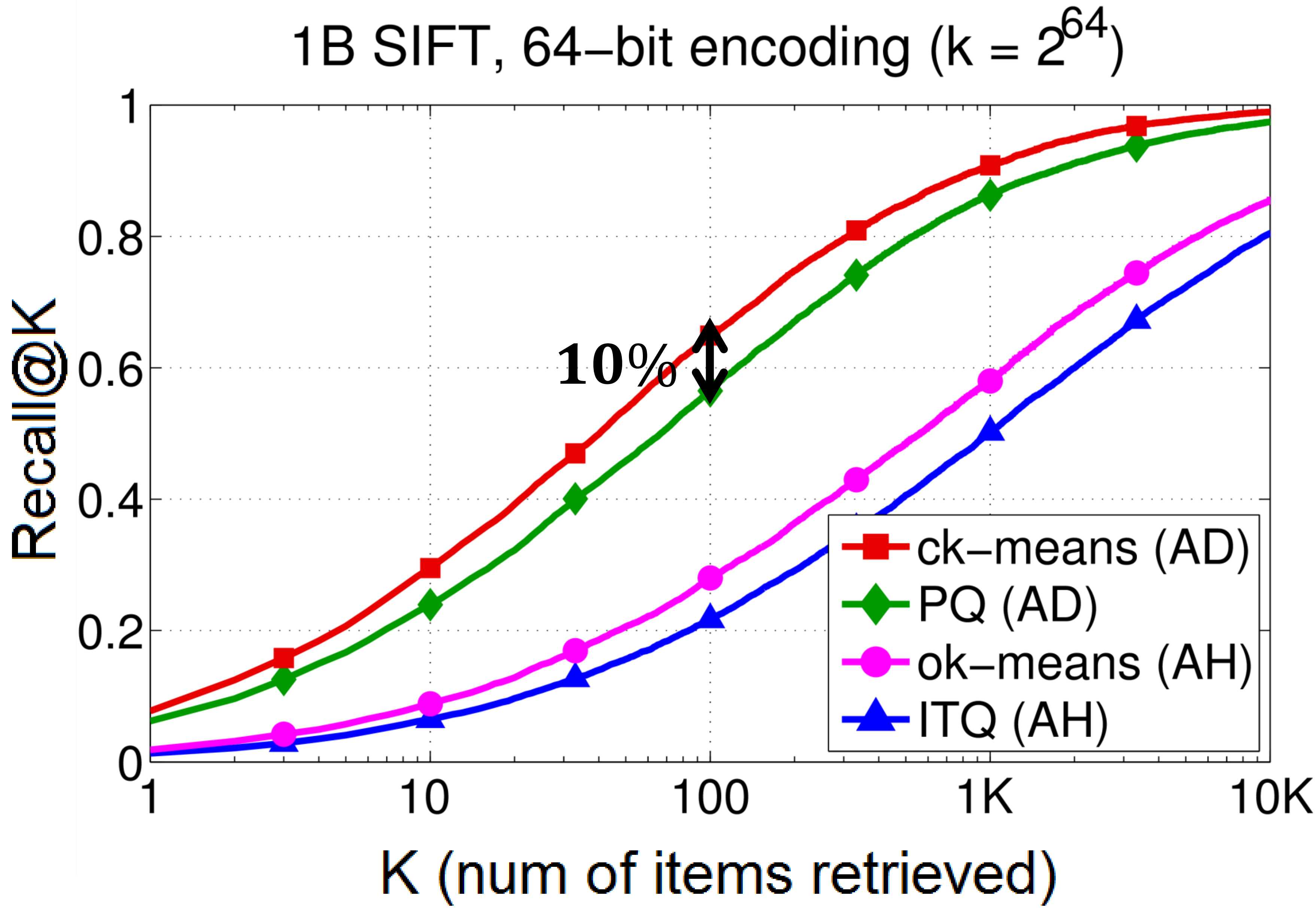


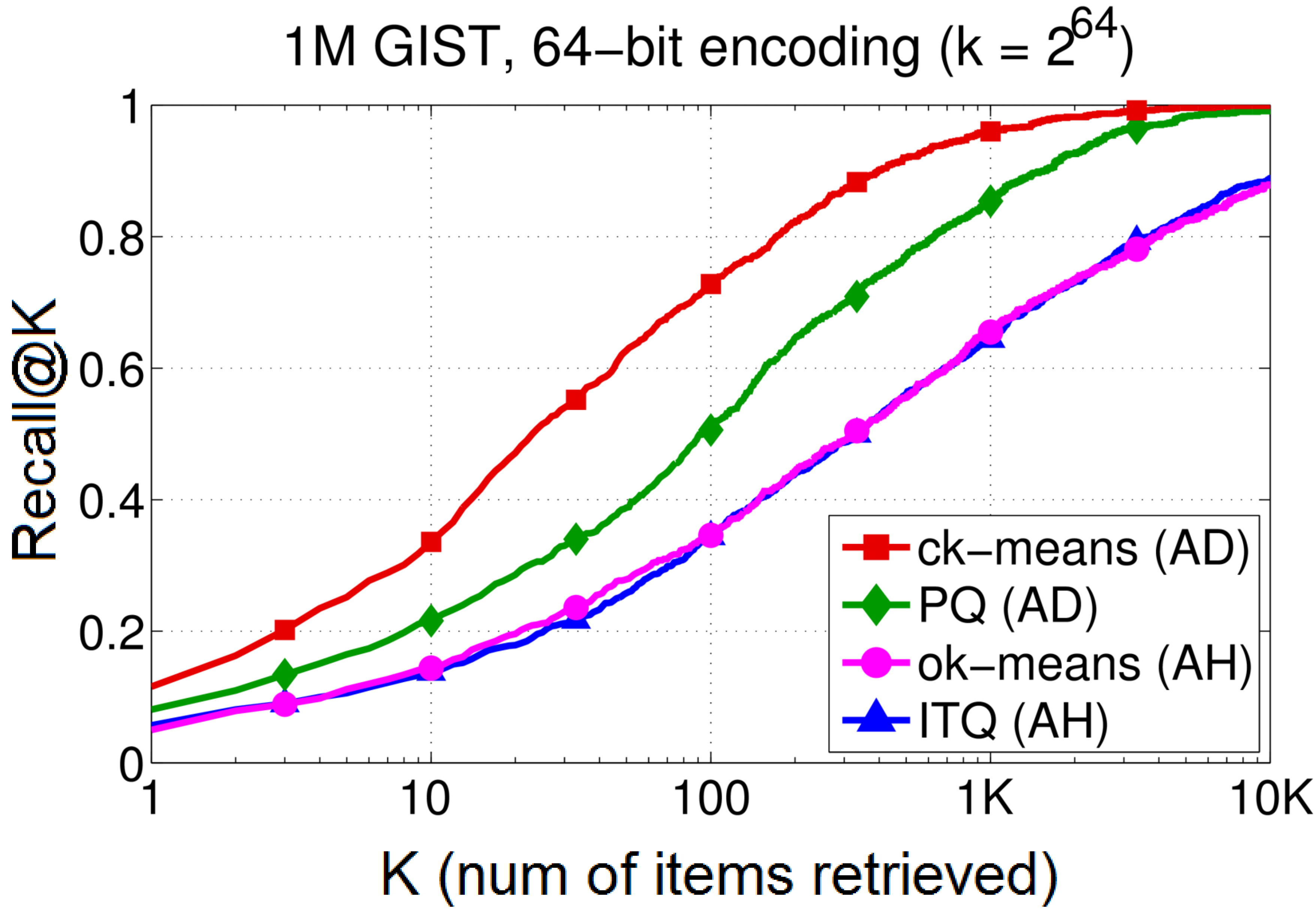
1M SIFT, 64−bit encoding (k = 2^64)
Recall@K
K (num of items retrieved)
1
0.8
0.6
0.4
0.2
0
1
10
100
1K
10K
PQ (AD)
ok−means (AH)
ITQ (AH)

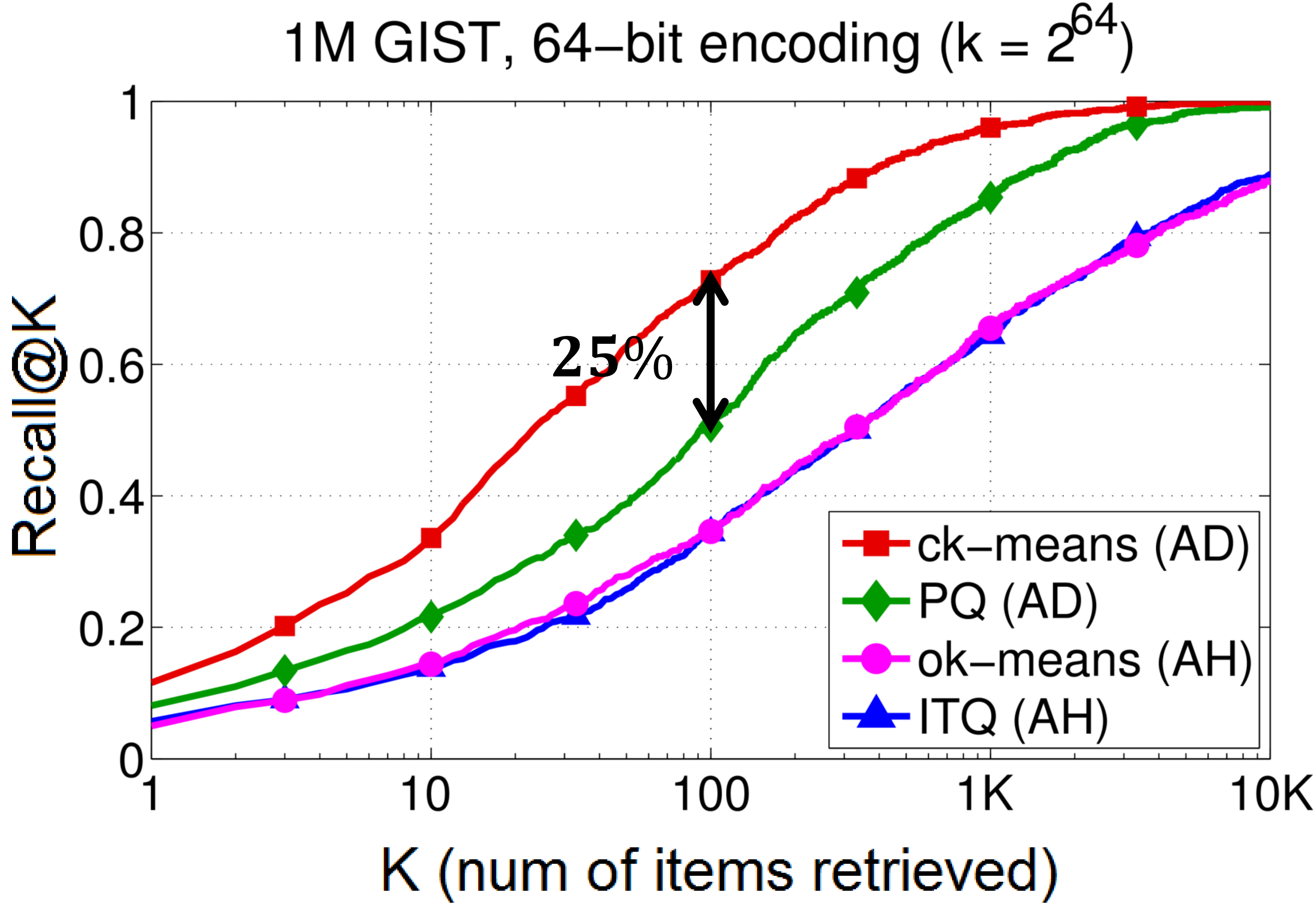

1M SIFT, 64−bit encoding (k = 2^64)
Recall@K
K (num of items retrieved)
1
0.8
0.6
0.4
0.2
0
1
10
100
1K
10K
ck−means (AD)
PQ (AD)
ok−means (AH)
ITQ (AH)



1B SIFT, 64−bit encoding (k = 2^64)
Recall@K
K (num of items retrieved)
0
0.2
0.4
0.6
0.8
1
1
10
100
1K
10K
ck−means (AD)
PQ (AD)
ok−means (AH)
ITQ (AH)


1B SIFT, 64−bit encoding (k = 2^64)
Recall@K
K (num of items retrieved)
10%
ck−means (AD)
PQ (AD)
ok−means (AH)
ITQ (AH)
0
0.2
0.4
0.6
0.8
1
1
10
100
1K
10K


1M GIST, 64−bit encoding ($k = 2^{64}$)
Recall@K
K (num of items retrieved)
0
0.2
0.4
0.6
0.8
1
1
10
100
1K
10K
ck−means (AD)
PQ (AD)
ok−means (AH)
ITQ (AH)



1M GIST, 64−bit encoding (k = 2^64)
Recall@K
K (num of items retrieved)
25%
ck−means (AD)
PQ (AD)
ok−means (AH)
ITQ (AH)
0
0.2
0.4
0.6
0.8
1
1
10
100
1K
10K

Codebook learning (CIFAR-10)

Codebook	Accuracy
k-means ($k = 1600$)	77.9%
k-means ($k = 4000$)	79.6%

Codebook learning (CIFAR-10)

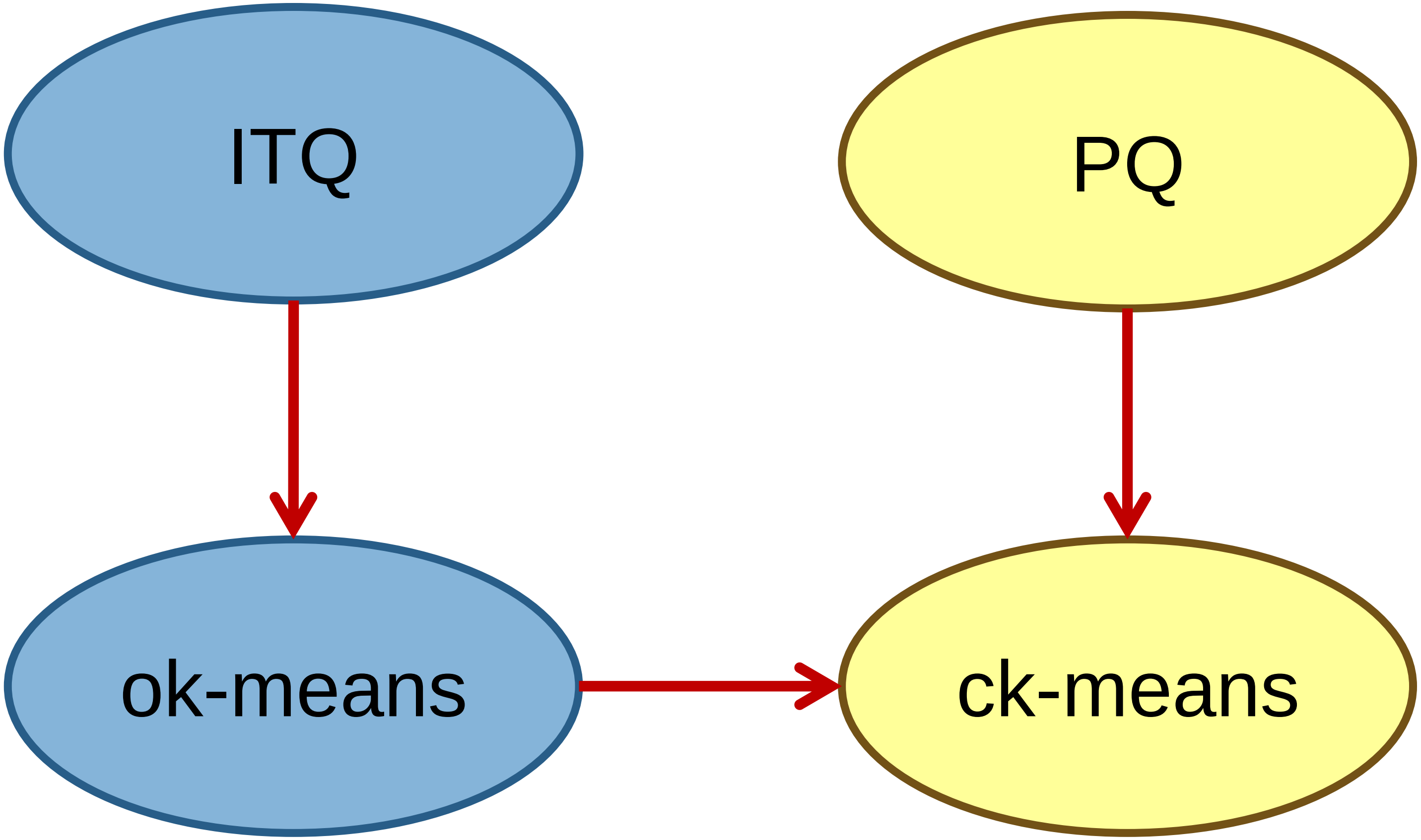
Codebook	Accuracy
k-means ($k = 1600$) ck-means ($k = 40^2$)	77.9% 78.2%
k-means ($k = 4000$) ck-means ($k = 64^2$)	79.6% 79.7%

Codebook learning (CIFAR-10)

Codebook	Accuracy
k-means ($k = 1600$) ck-means ($k = 40^2$) PQ ($k = 40^2$)	77.9% 78.2% 75.9%
k-means ($k = 4000$) ck-means ($k = 64^2$) PQ ($k = 64^2$)	79.6% 79.7% 78.2%

Summary

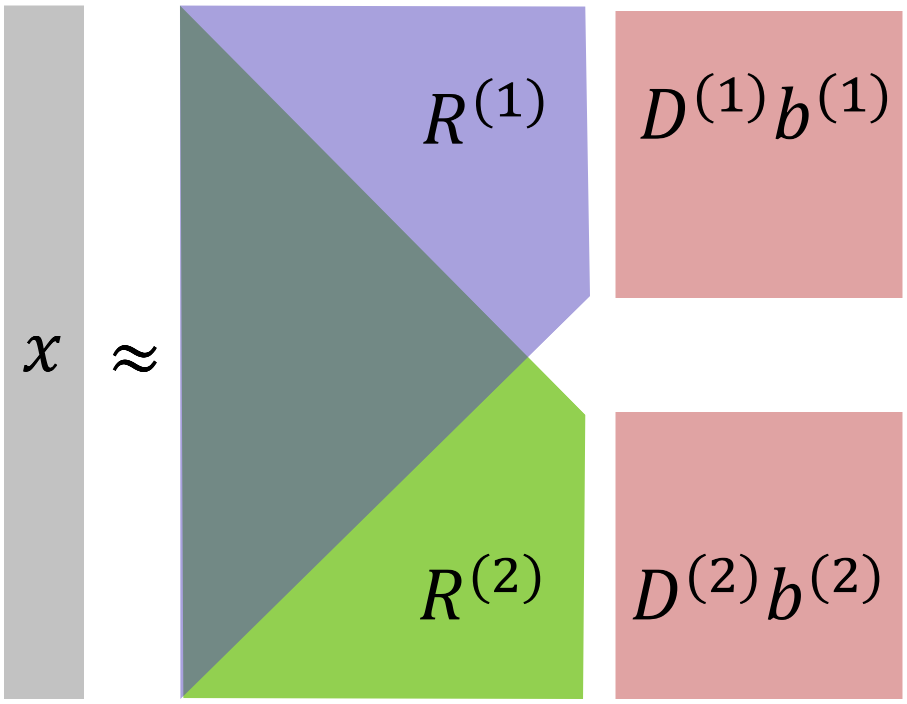
Summary

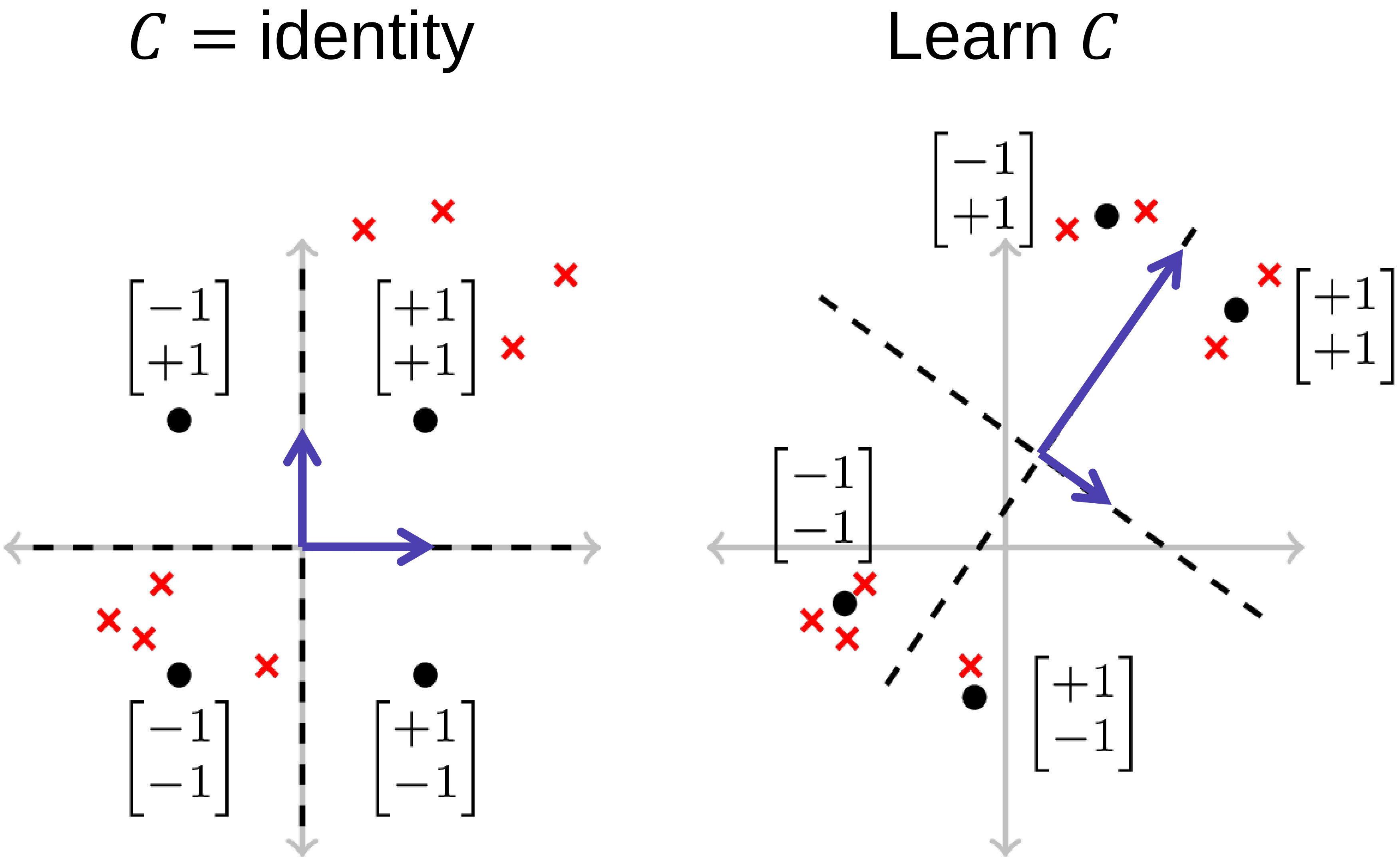


Thank you for your attention!

Please ask questions :-)

Cartesian k-means





C = identity
Learn C
$\begin{bmatrix}-1\\+1\end{bmatrix}$
$\begin{bmatrix}+1\\+1\end{bmatrix}$
$\begin{bmatrix}-1\\-1\end{bmatrix}$
$\begin{bmatrix}+1\\-1\end{bmatrix}$
$\begin{bmatrix}-1\\+1\end{bmatrix}$
$\begin{bmatrix}+1\\+1\end{bmatrix}$
$\begin{bmatrix}-1\\-1\end{bmatrix}$
$\begin{bmatrix}+1\\-1\end{bmatrix}$