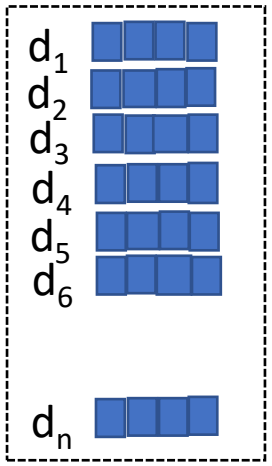
Lesson 4

k-Nearest Neighbors
&
Centroid Based Classifier

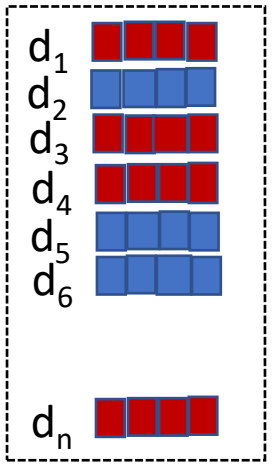
k-Nearest Neighbors

k-Nearest Neighbors



Dataset

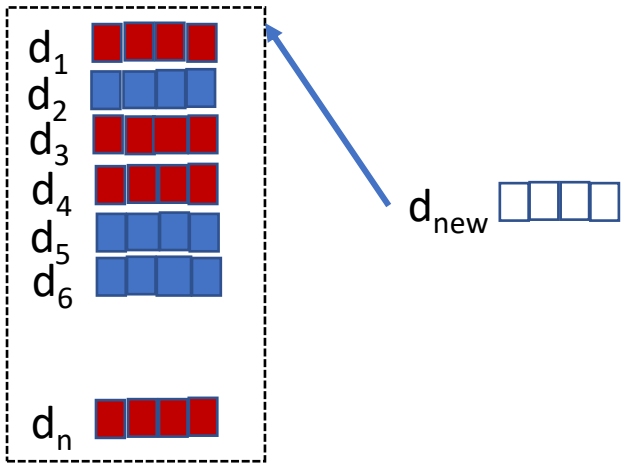
k-Nearest Neighbors



Dataset

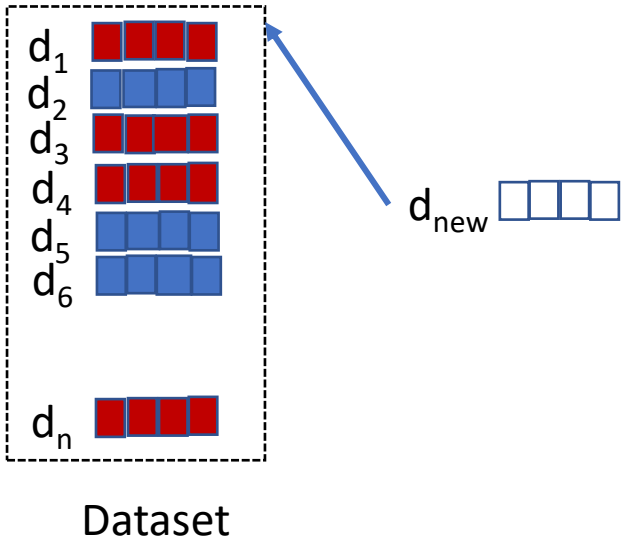
d_{new} 

k-Nearest Neighbors



$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

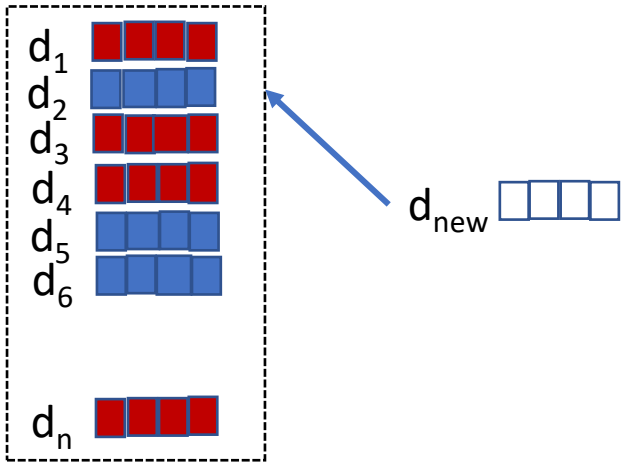
k-Nearest Neighbors



$$\omega(d_1, d_{new})$$

$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

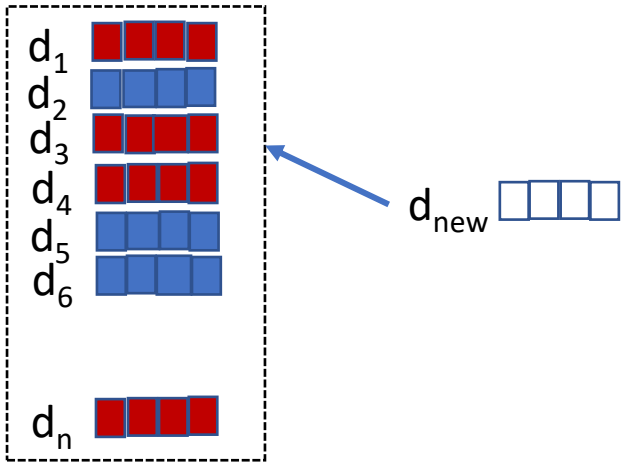
k-Nearest Neighbors



$\omega(d_1, d_{new})$
$\omega(d_2, d_{new})$

$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

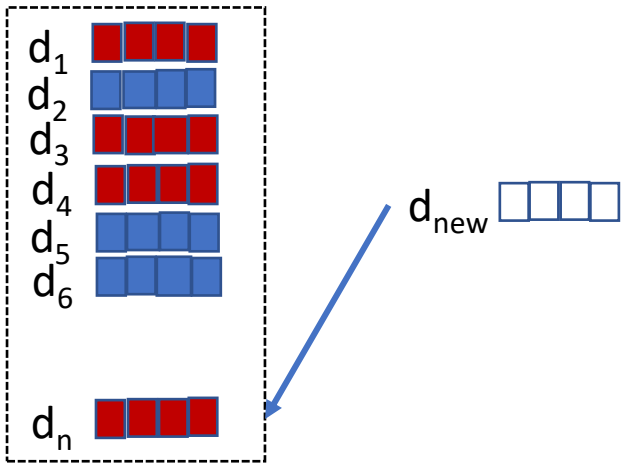
k-Nearest Neighbors



$\omega(d_1, d_{new})$
$\omega(d_2, d_{new})$
$\omega(d_3, d_{new})$

$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

k-Nearest Neighbors

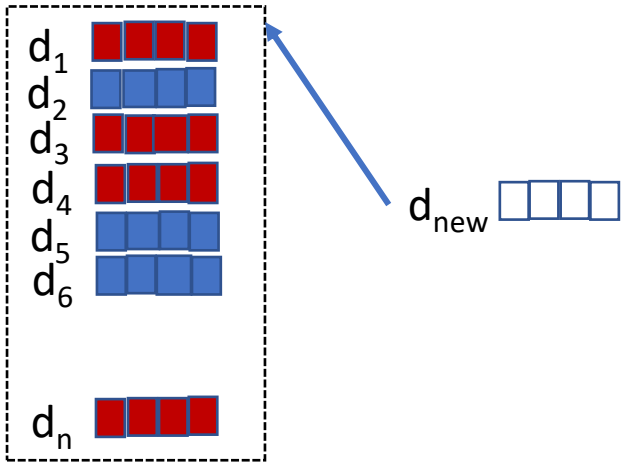


Similarity

$\omega(d_1, d_{new})$
$\omega(d_2, d_{new})$
$\omega(d_3, d_{new})$
$\omega(d_4, d_{new})$
$\omega(d_5, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_n, d_{new})$

$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

k-Nearest Neighbors



Similarity

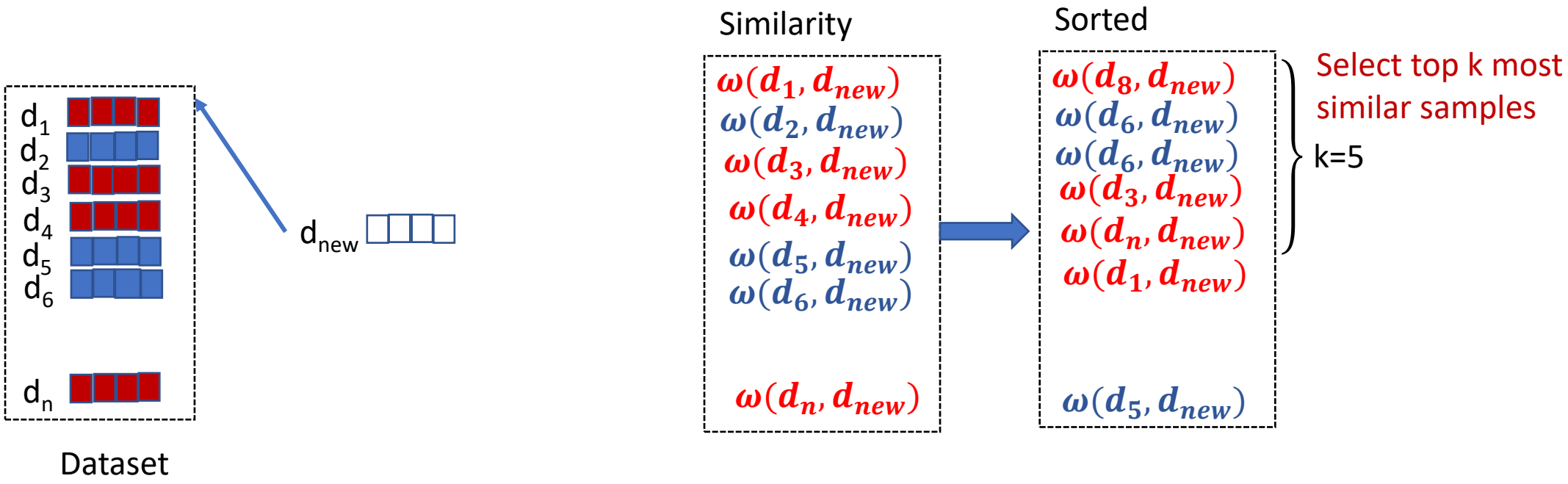
$\omega(d_1, d_{new})$
$\omega(d_2, d_{new})$
$\omega(d_3, d_{new})$
$\omega(d_4, d_{new})$
$\omega(d_5, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_n, d_{new})$

Sorted in descending order

$\omega(d_8, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_3, d_{new})$
$\omega(d_n, d_{new})$
$\omega(d_1, d_{new})$
$\omega(d_5, d_{new})$

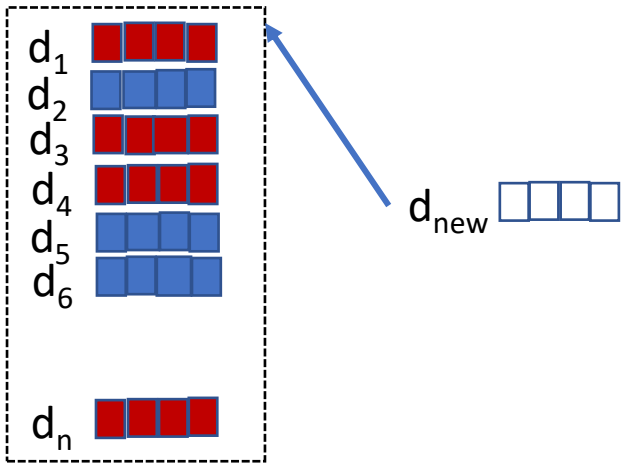
$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

k-Nearest Neighbors



$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

k-Nearest Neighbors



Similarity

$\omega(d_1, d_{new})$
$\omega(d_2, d_{new})$
$\omega(d_3, d_{new})$
$\omega(d_4, d_{new})$
$\omega(d_5, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_n, d_{new})$

Sorted

$\omega(d_8, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_6, d_{new})$
$\omega(d_3, d_{new})$
$\omega(d_n, d_{new})$
$\omega(d_1, d_{new})$
$\omega(d_5, d_{new})$

k=5

Majority class label

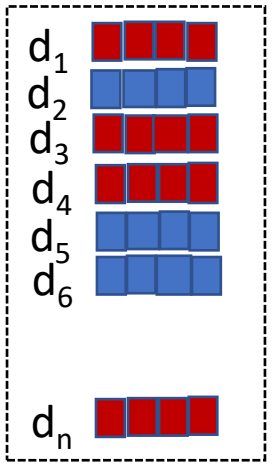
Red = 3
Blue = 2

d_{new} is RED

$$Cosine\ Similarity(d_j, d_{new}) = \frac{\sum_{i=1}^{k}(d_{ji}.d_{newi})}{\sqrt{\sum_{i=1}^{k} d_{ji}^2}\sqrt{\sum_{i=1}^{n} d_{newi}^2}}$$

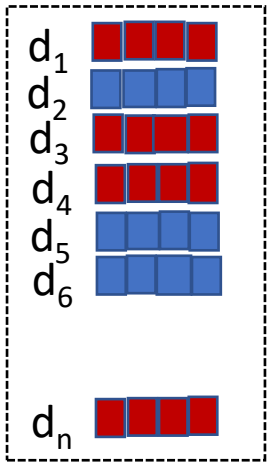
Centroid Based Classifier (Rocchio Classifier)

Centroid Based Classifier



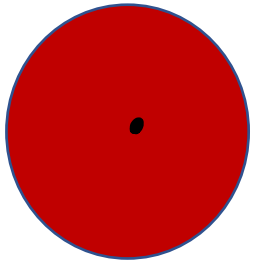
Dataset

Centroid Based Classifier

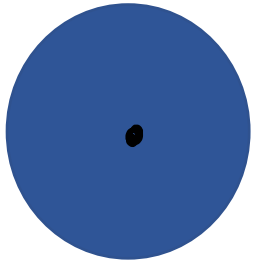


Dataset

d_{new}

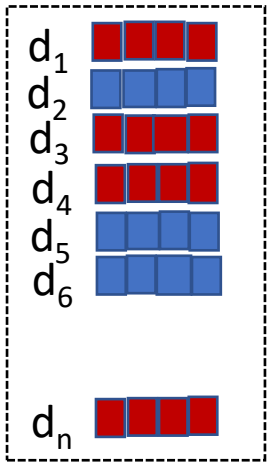


$$c_{red} = \frac{1}{|c_{red}|} \sum_{d_i \in c_{red}} d_i$$

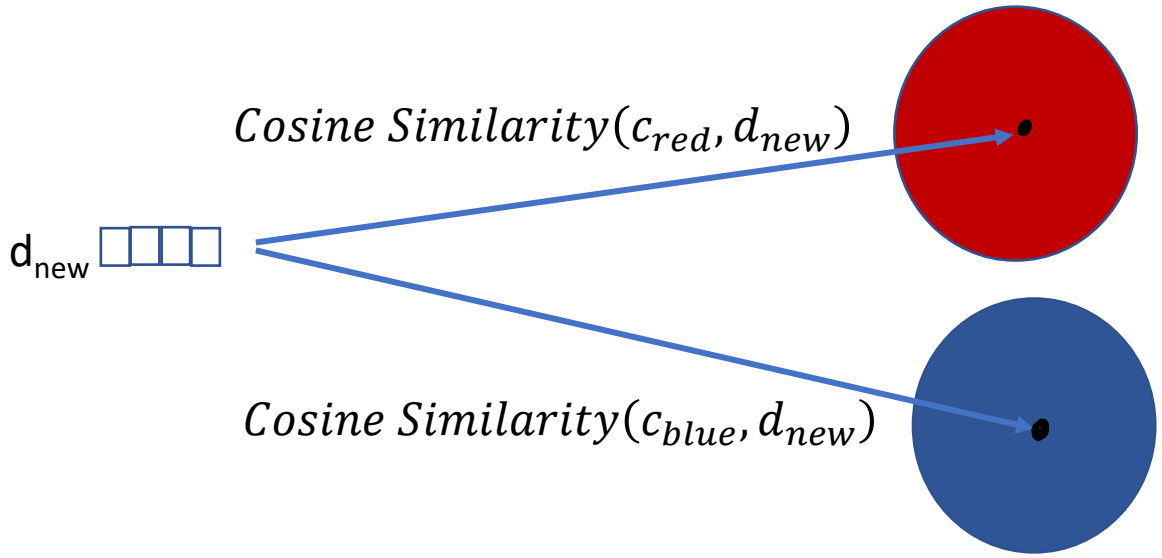


$$c_{blue} = \frac{1}{|c_{blue}|} \sum_{d_i \in c_{blue}} d_i$$

Centroid Based Classifier



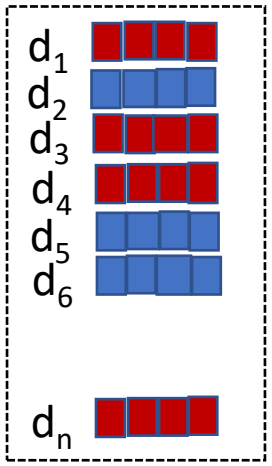
Dataset



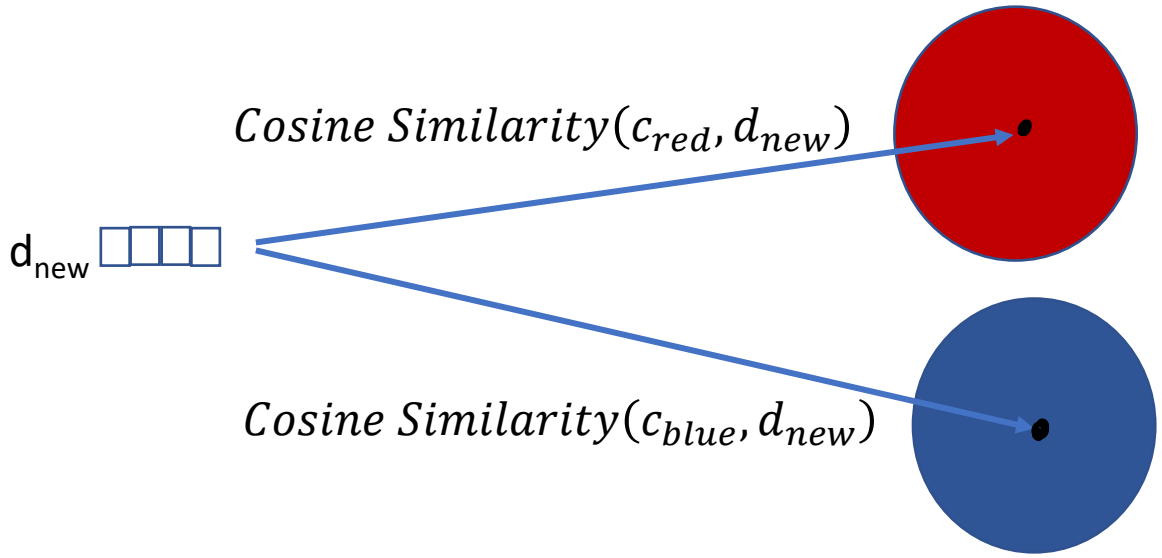
$$c_{red} = \frac{1}{|c_{red}|} \sum_{d_i \in c_{red}} d_i$$

$$c_{blue} = \frac{1}{|c_{blue}|} \sum_{d_i \in c_{blue}} d_i$$

Centroid Based Classifier



Dataset



$$c_{red} = \frac{1}{|c_{red}|} \sum_{d_i \in c_{red}} d_i$$

$$c_{blue} = \frac{1}{|c_{blue}|} \sum_{d_i \in c_{blue}} d_i$$

Assign Class Label with the nearest Centroid.