UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

Sesión 12.0: Stranger things

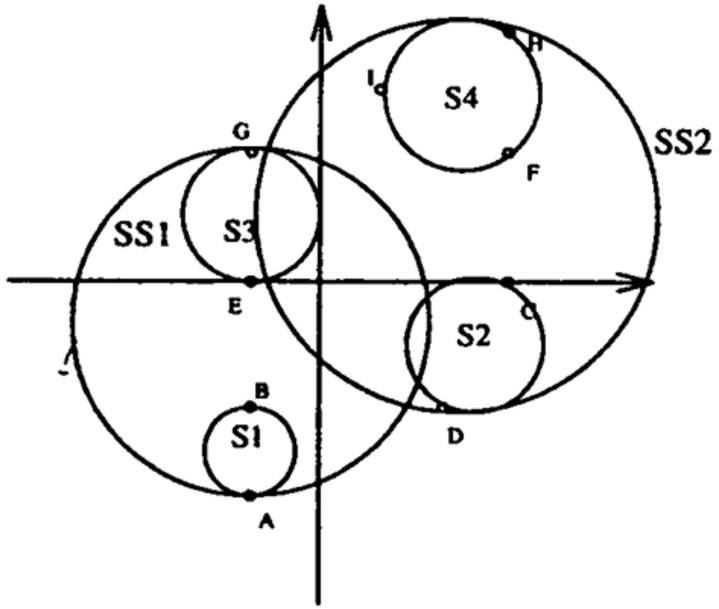
CS3102 EDA

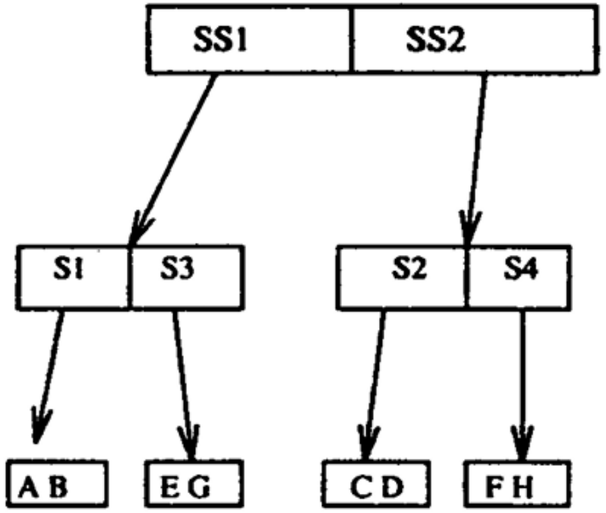
Índice

1. TV-Tree

TV-*Tree*

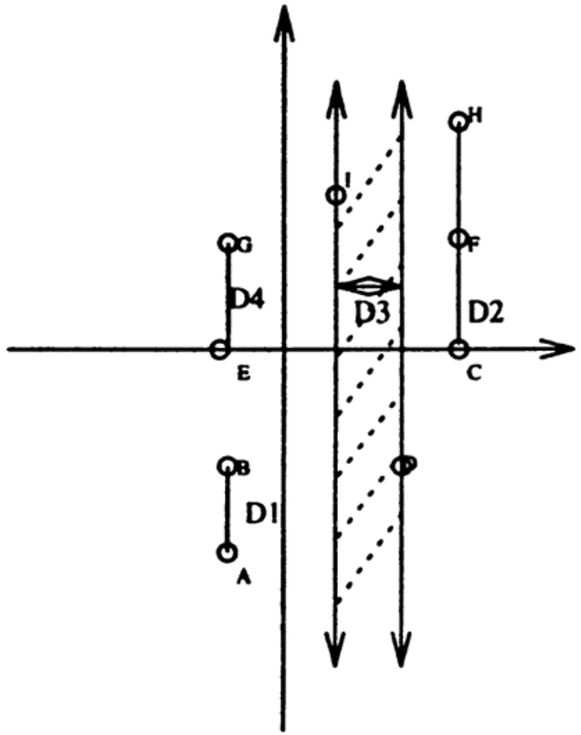
(Telescoping vector tree)

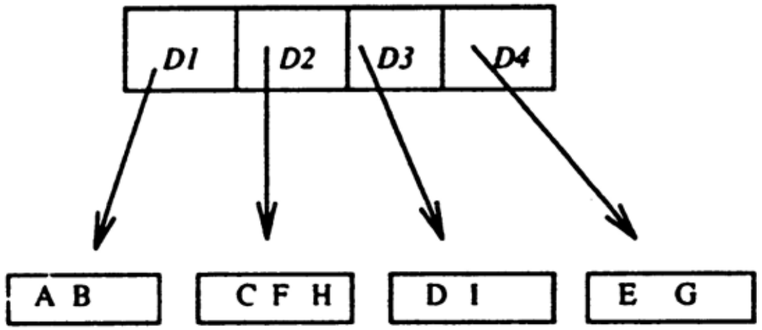




King-Ip Lin et al. (1994) "The TV-tree: An index structure for high-dimensional data".
The VLDB Journal, 1994, vol. 3, no 4, p. 517-542.

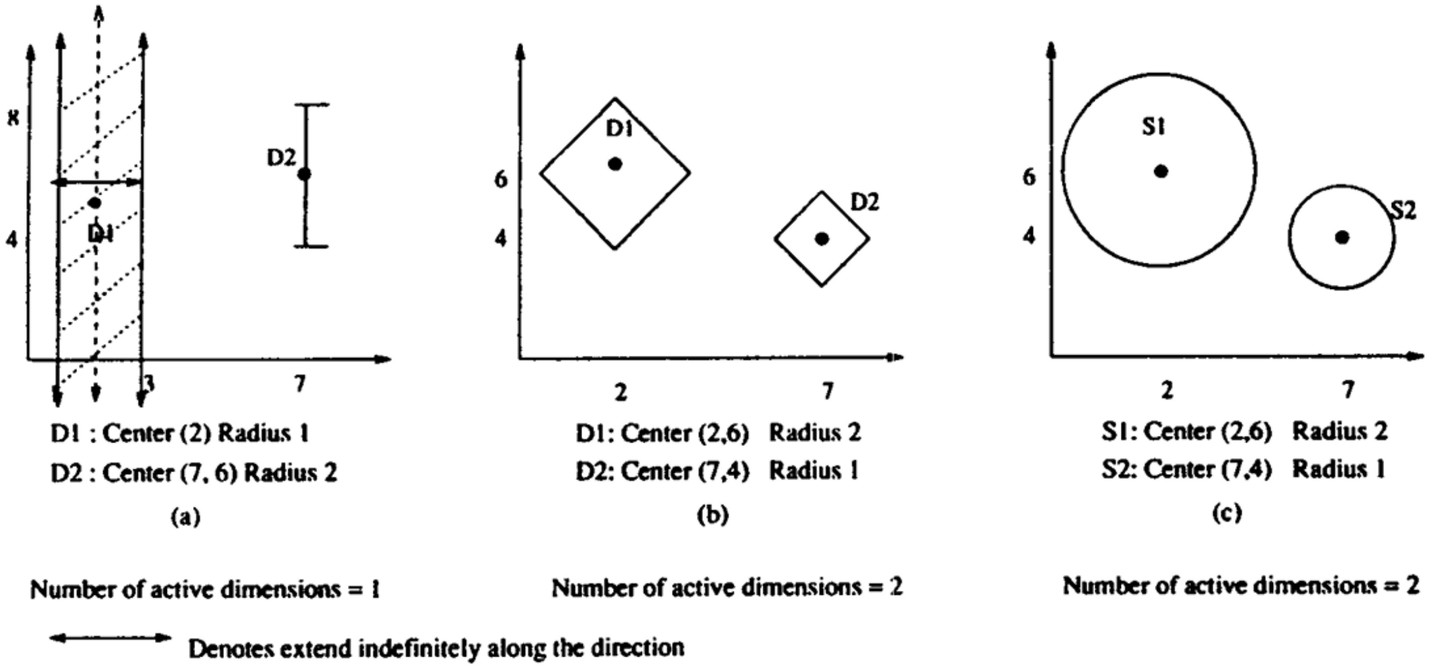
TV-*Tree*





King-Ip Lin et al. (1994) "The TV-tree: An index structure for high-dimensional data".
The VLDB Journal, 1994, vol. 3, no 4, p. 517-542.

TV-*Tree*

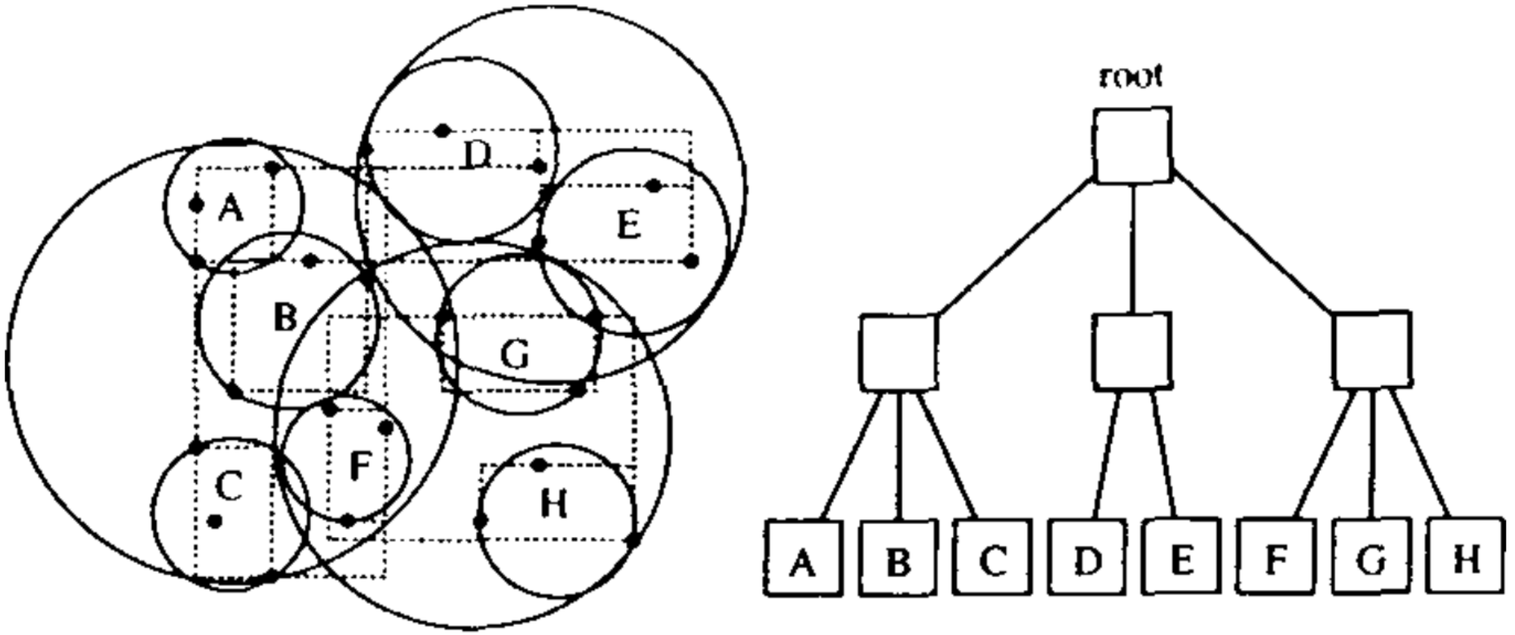


King-Ip Lin et al. (1994) "The TV-tree: An index structure for high-dimensional data".
The VLDB Journal, 1994, vol. 3, no 4, p. 517-542.

2. SR-Tree

SR-*Tree*

(Sphere/Rectangle-tree)



Norio Katayama and Shin'ichi Satoh (1997) "The SR-tree: An index structure for high-dimensional nearest neighbor queries".
ACM Sigmod Record, 1997, vol. 26, no 2, p. 369-380.

SR-*Tree*

Centro del círculo

$$x_i = \frac{\sum_{k=1}^{n} n_k c_i^k}{\sum_{k=1}^{n} n_k} \qquad 1 \le k \le n \quad \text{Índices de los hijos.} \qquad 1 \le i \le D \quad \text{Índices de las dimensiones}$$

Radio del círculo

$$r = \min(d_s, d_r)$$

$$d_s = \max_{1 \le k \le n} (x - c_k + r_k)$$

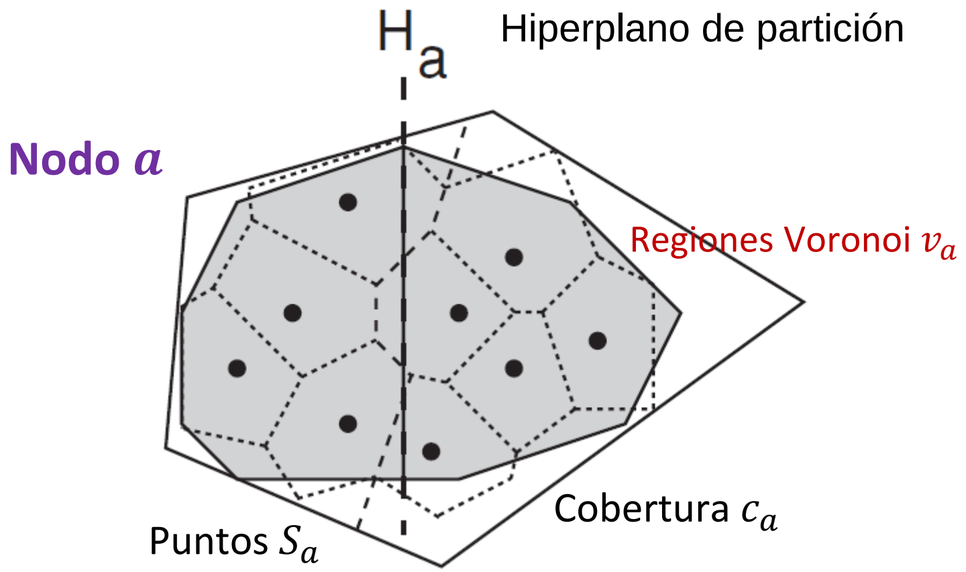
$$d_r = \max_{1 \le k \le n} (\text{MaxDist}(x, \boldsymbol{R}_k))$$

Norio Katayama and Shin'ichi Satoh (1997) "The SR-tree: An index structure for high-dimensional nearest neighbor queries".
ACM Sigmod Record, 1997, vol. 26, no 2, p. 369-380.

3. os-Tree

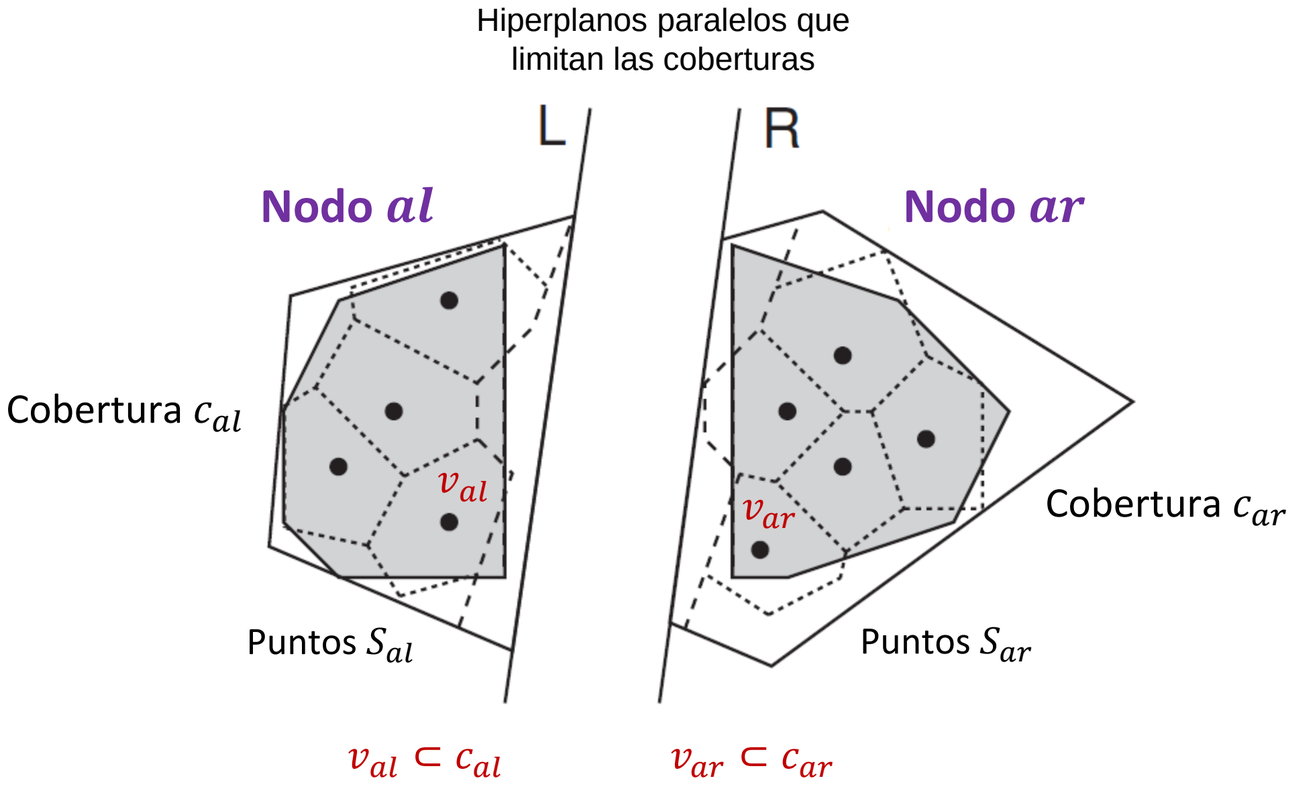
OS-*Tree*

(Overlapped-Split tree)



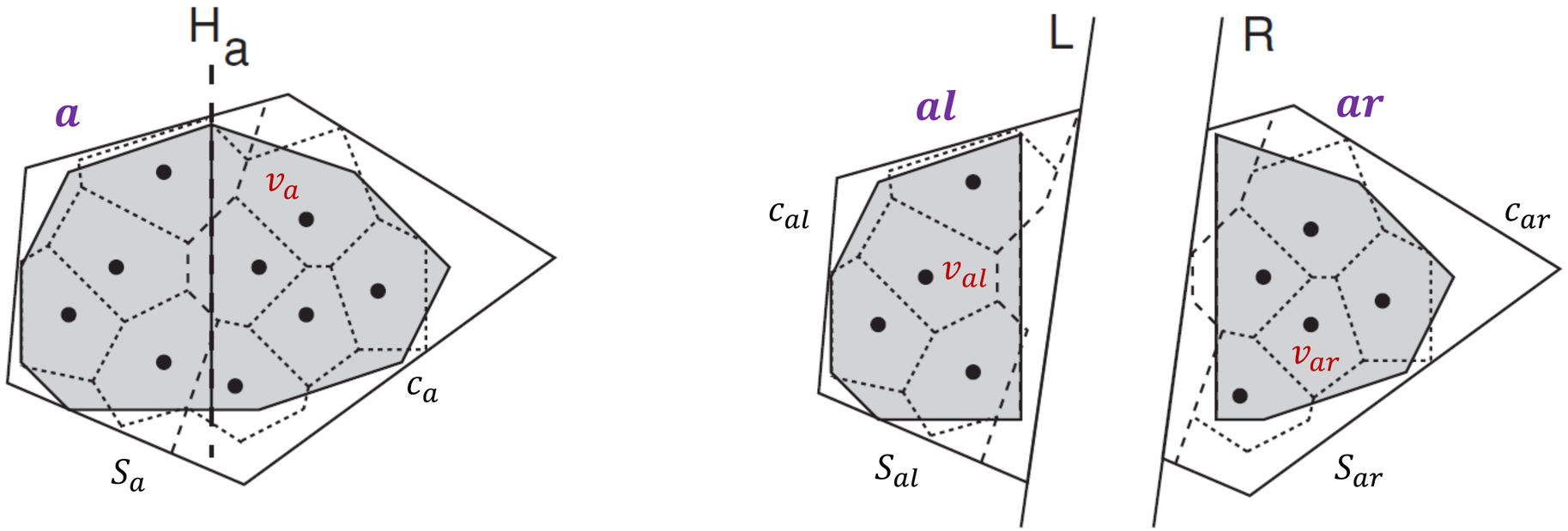
Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

OS-*Tree*



Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

OS-*Tree*



OS-Tree requiere que tengamos el diagrama Voronoi construida

Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

OS-*Tree*

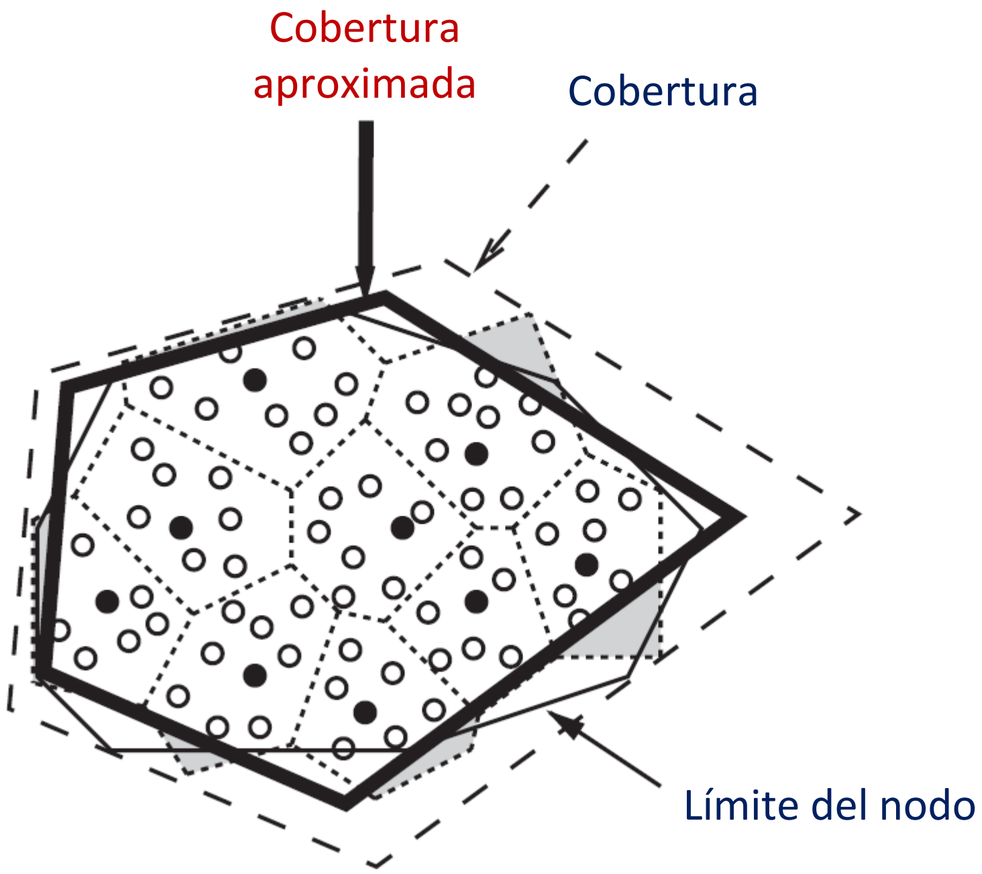
Sin embargo…..

Es poco eficiente para d≥3

:c

Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

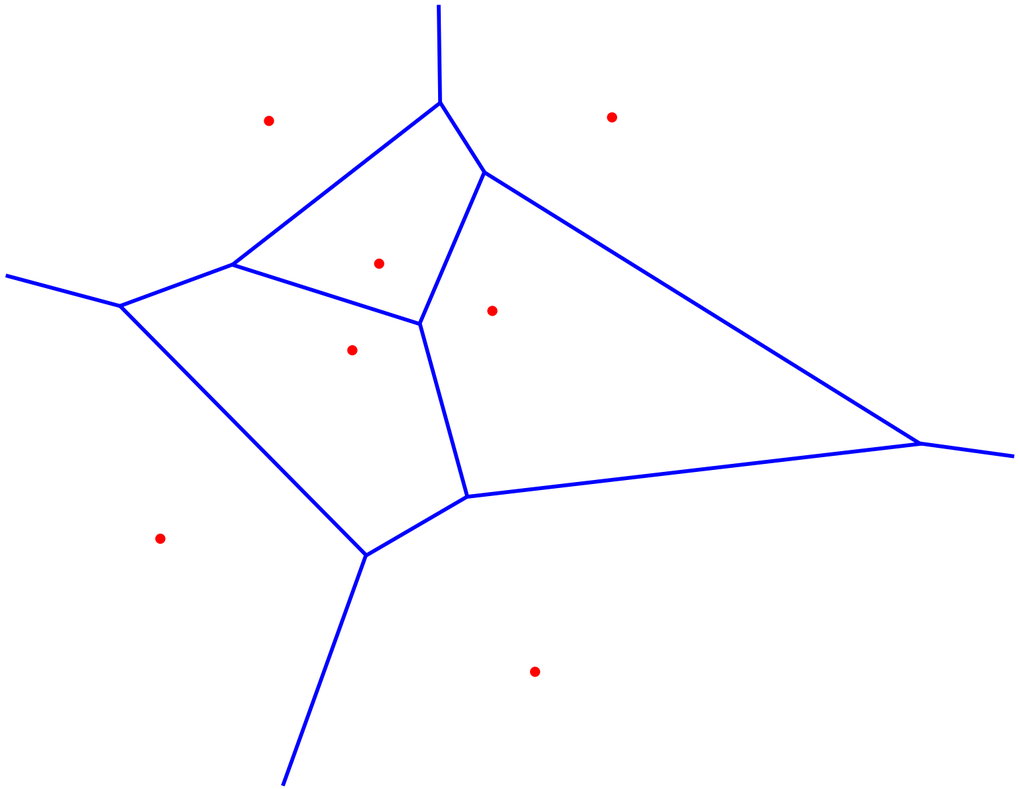
Probably correct *OS-Tree*



- Puntos de datos (S)
- Puntos de entrenamiento (T)

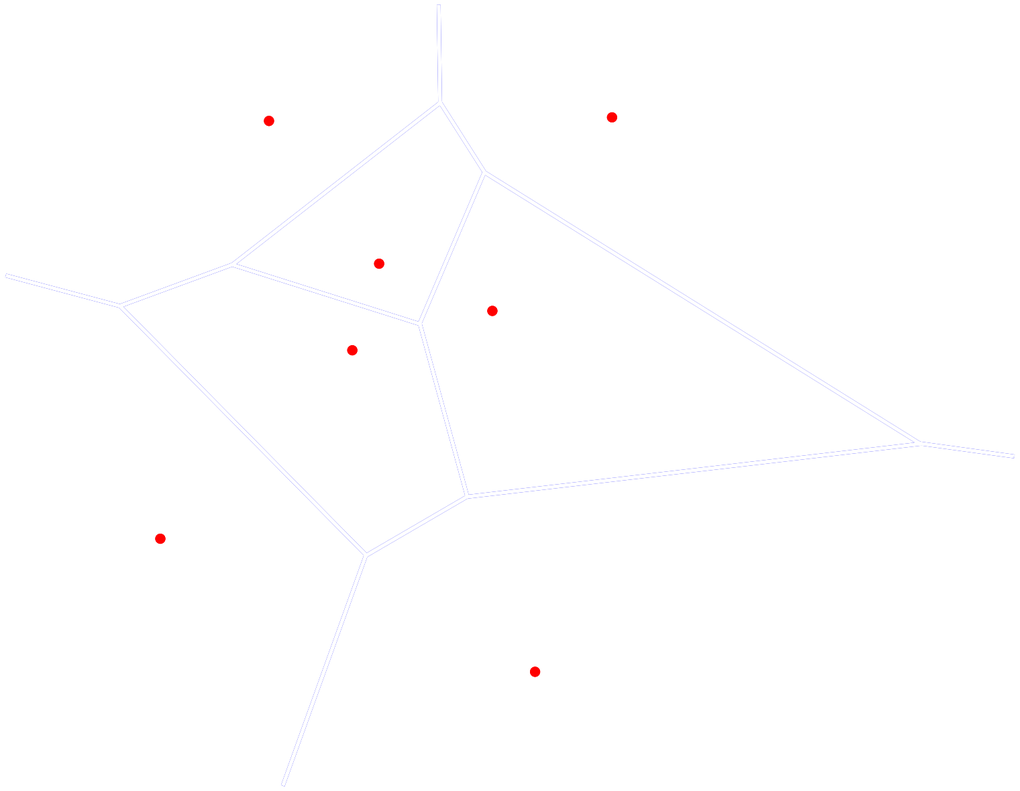
Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

Probably correct *OS-Tree*



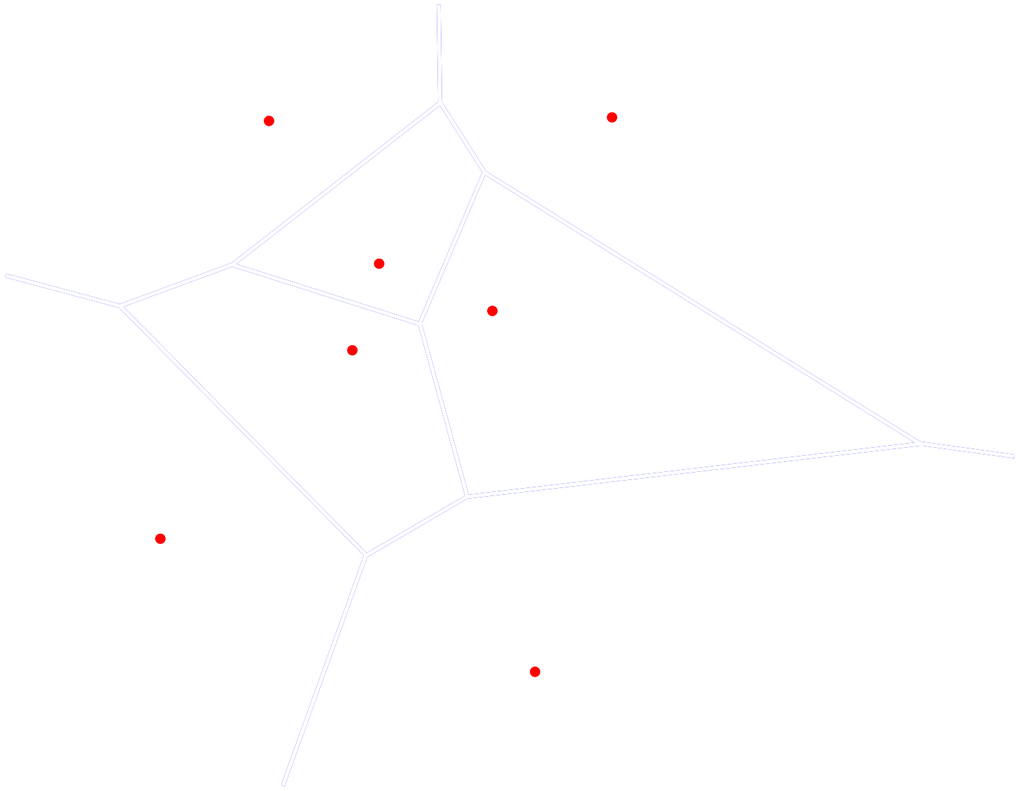
Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

Probably correct *OS-Tree*



Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

Probably correct *OS-Tree*

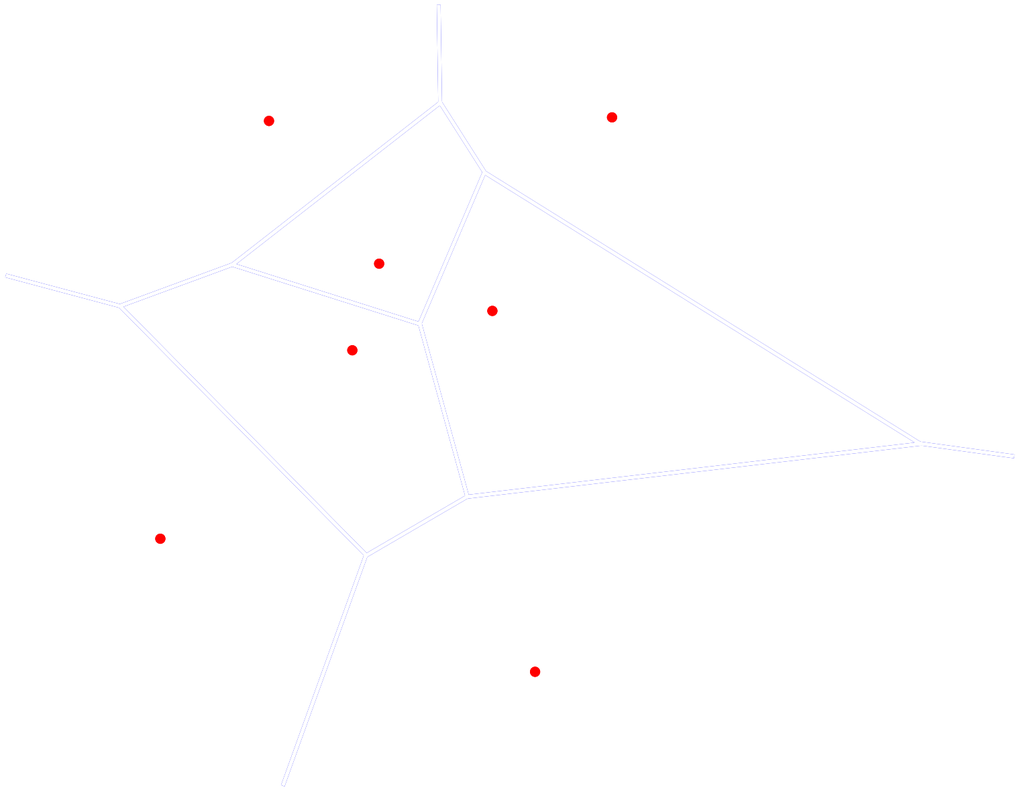


1) Elegir los puntos que componen los subconjuntos S_{al} y S_{ar}

2) Elegir un plano de partición I_a para la celda correspondiente a a.

Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

Probably correct *OS-Tree*



1) Elegir los puntos que componen los subconjuntos S_{al} y S_{ar}

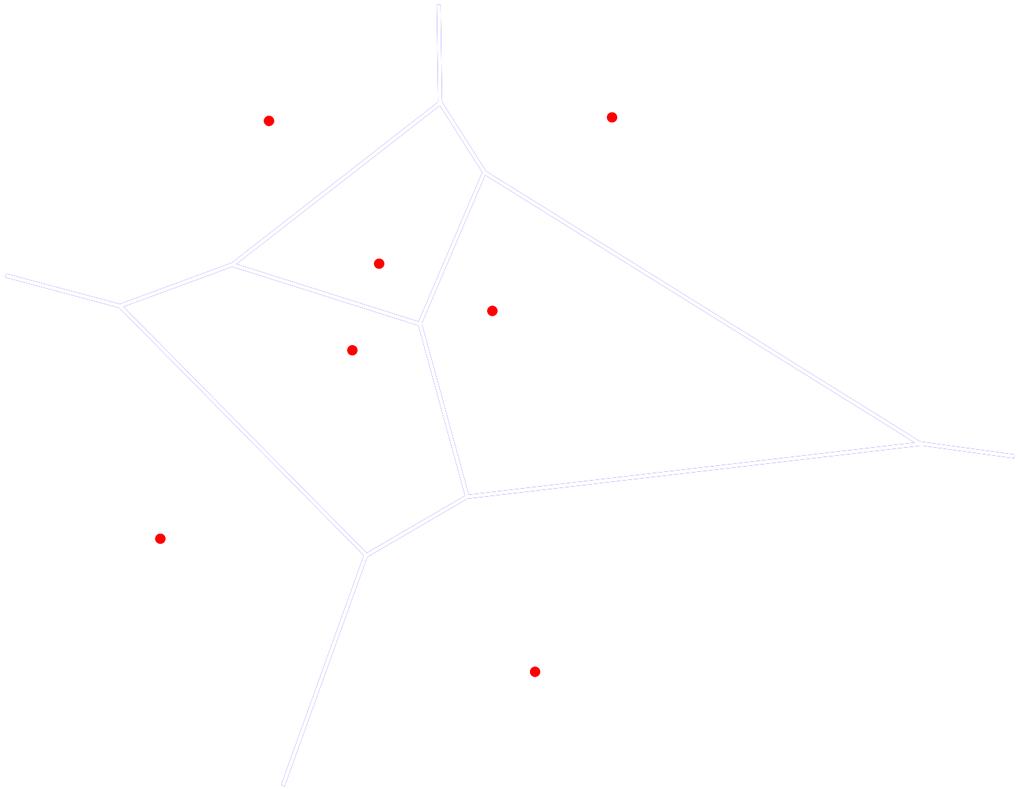
Buscamos el plano ortogonal a la dirección de la mayor variación de los datos.

2) Elegir un plano de partición I_a para la celda correspondiente a a.

Minimizar el grado en que las regiones de Voronoi se superponen al lado contrario

Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

Probably correct *OS-Tree*



1) Elegir los puntos que componen los subconjuntos S_{al} y S_{ar}

Buscamos el plano ortogonal a la dirección de la mayor variación de los datos.

PCA!

2) Elegir un plano de partición I_a para la celda correspondiente a a.

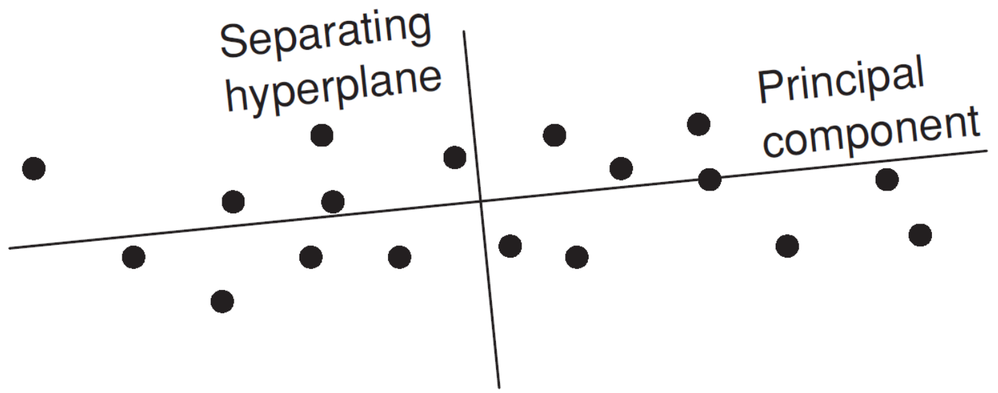
Minimizar el grado en que las regiones de Voronoi se superponen al lado contrario

SVM!

Songrit Maneewongvatana and David M. Mount (2001) "The analysis of a probabilistic approach to nearest neighbor searching". Algorithms and Data Structures: 7th International Workshop, WADS 2001 Providence, RI, USA, August 8–10, 2001 Proceedings (pp. 276-286). Berlin, Heidelberg: Springer Berlin Heidelberg.

Probably correct *OS-Tree*

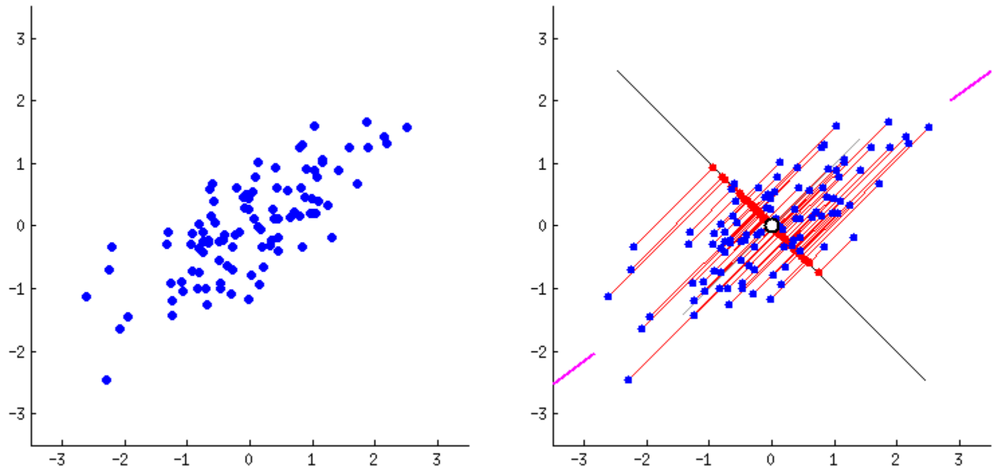
1) Elegir los puntos que componen los subconjuntos S_{al} y S_{ar}



Karl Pearson (1901) "On lines and planes of closest fit to systems of points in space".
The London, Edinburgh, and Dublin philosophical magazine and journal of science, vol. 2, no 11, p. 559-572.

Probably correct *OS-Tree*

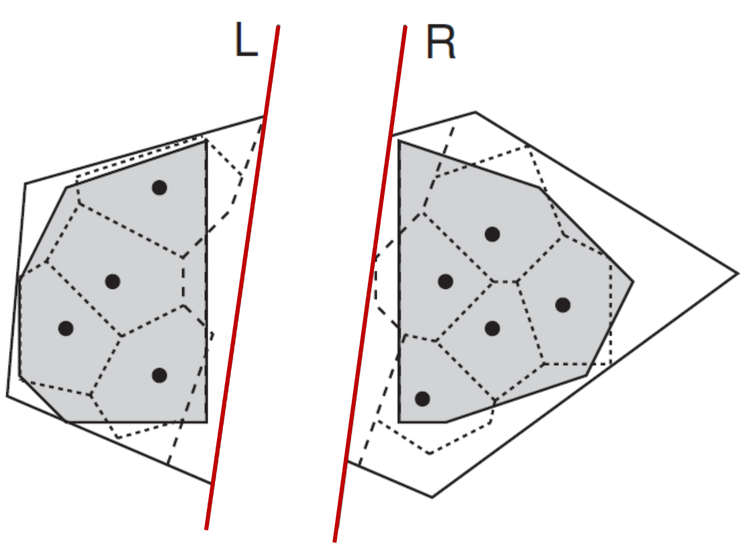
1) Elegir los puntos que componen los subconjuntos S_{al} y S_{ar}

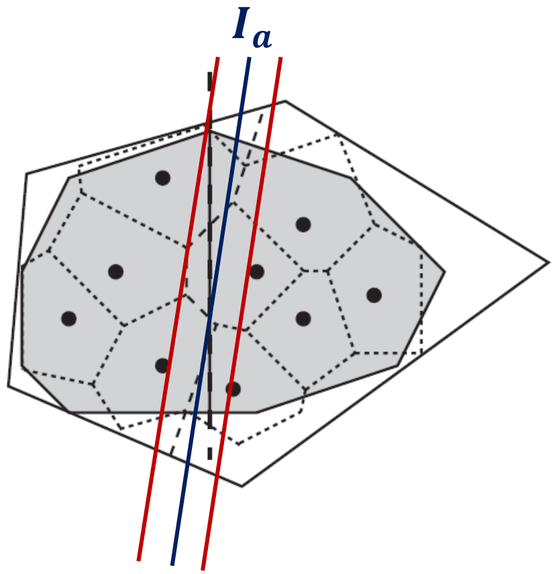


Karl Pearson (1901) "On lines and planes of closest fit to systems of points in space".
The London, Edinburgh, and Dublin philosophical magazine and journal of science, vol. 2, no 11, p. 559-572.

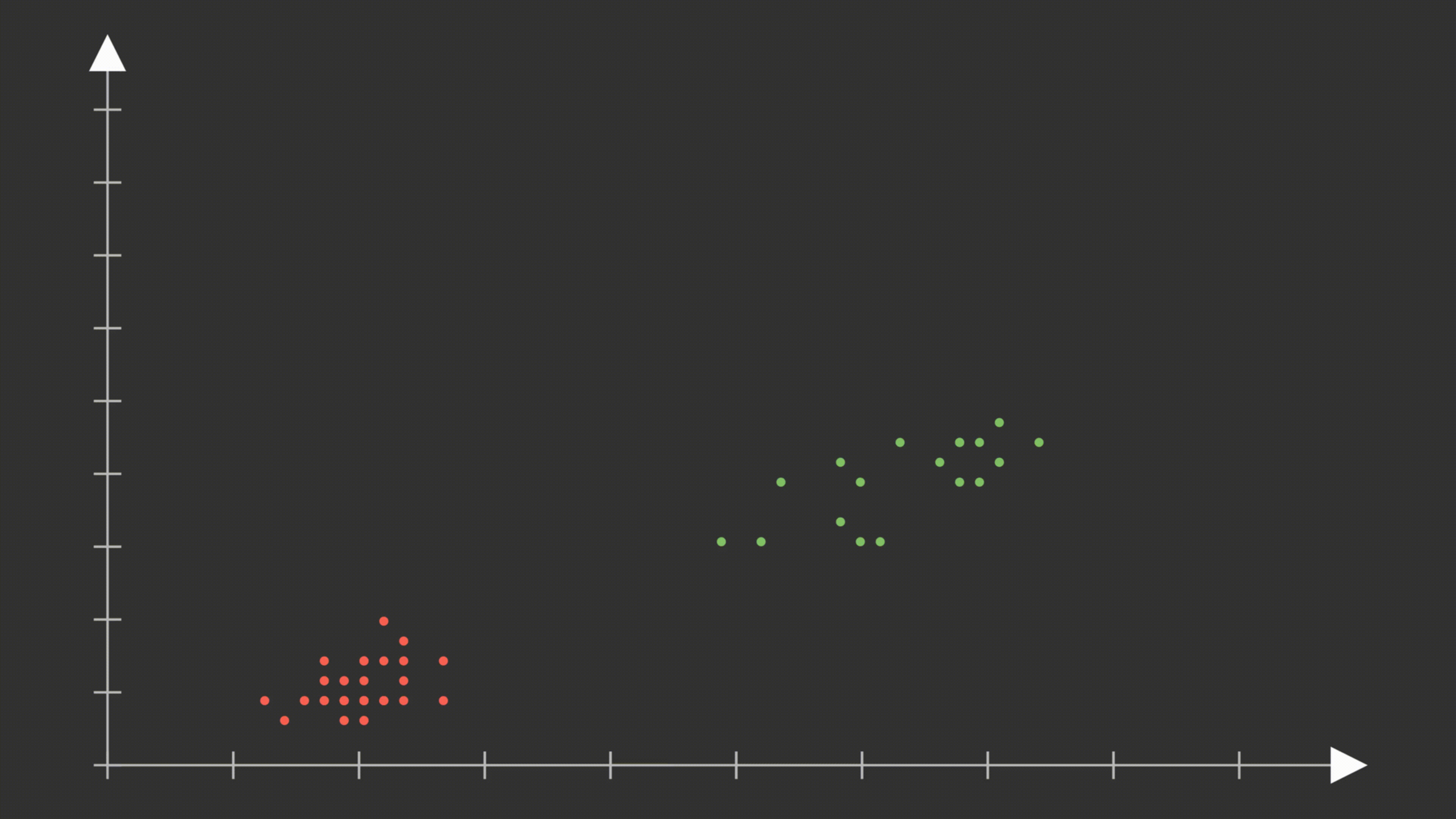
Probably correct *OS-Tree*

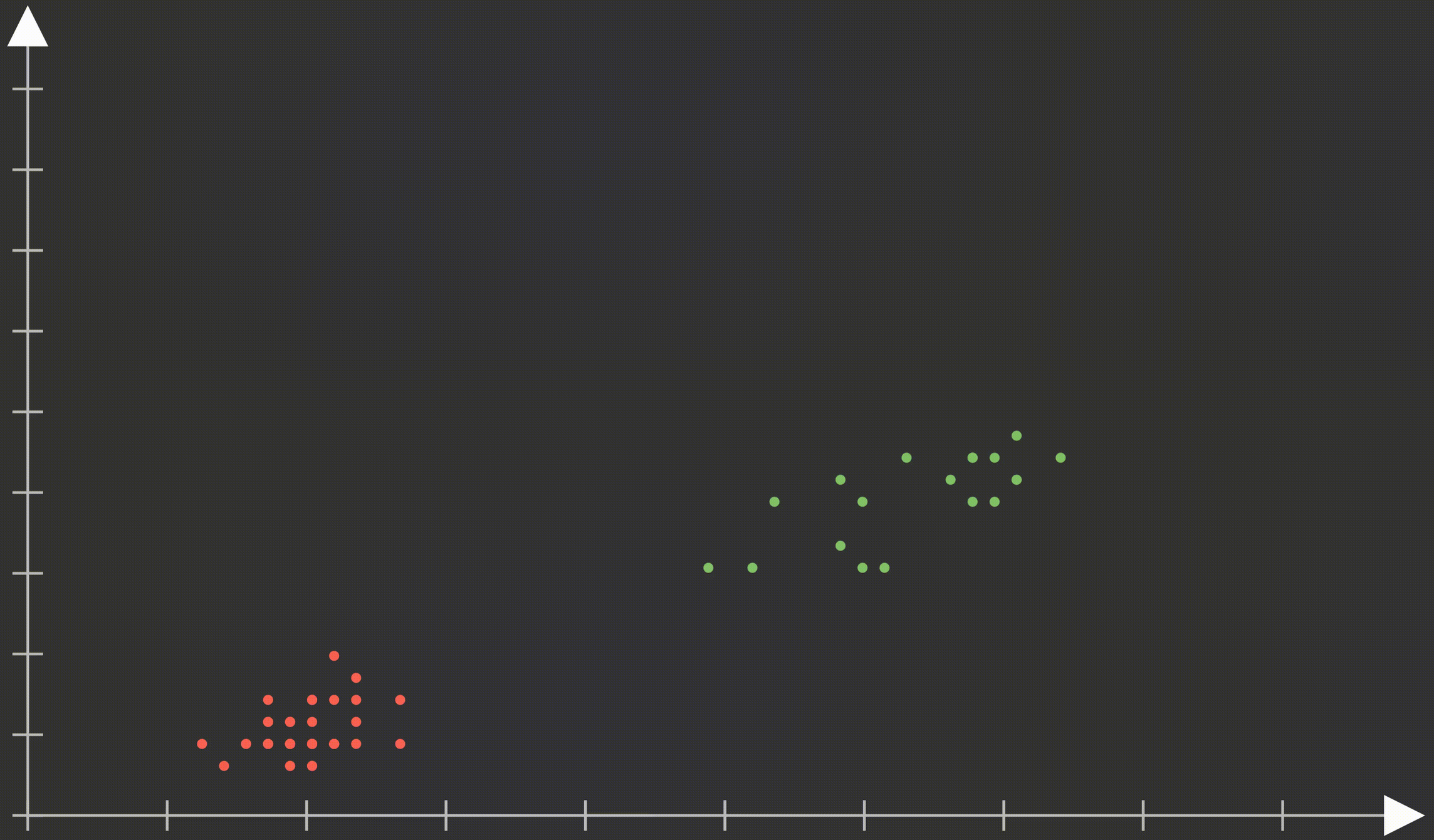
2) Elegir un plano de partición I_a para la celda correspondiente a a

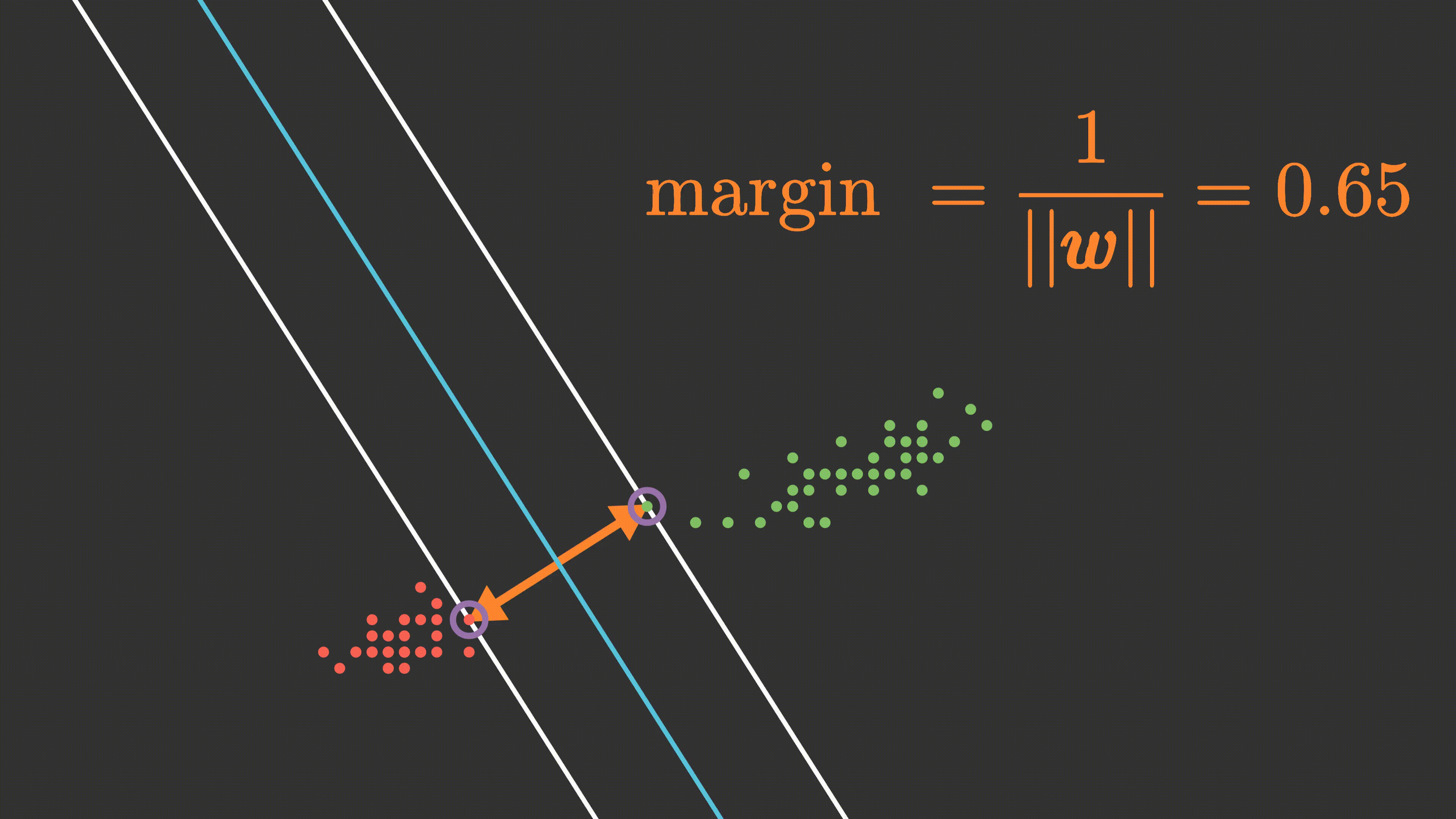




Corinna Cortes and Vladimir Vapnik (1995) "Support-vector networks". Machine learning, vol. 20, p. 273-297.






$\text{margin} = \frac{1}{||\boldsymbol{w}||} = 0.65$

UTEC

UNIVERSIDAD DE INGENIERÍA
Y TECNOLOGÍA