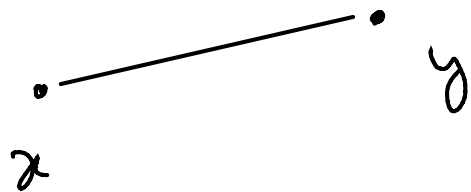


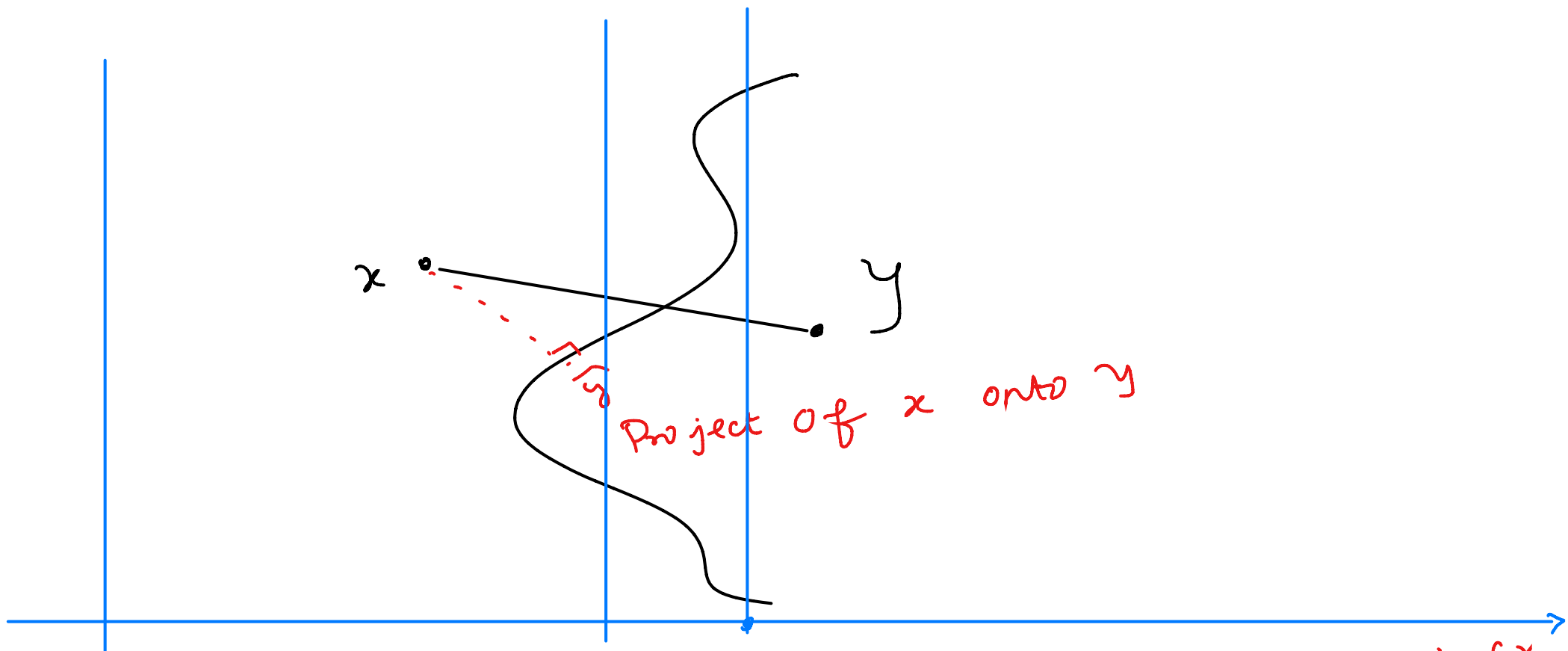
Euclidean Distance between two points $x \in \mathbb{R}^d$, $y \in \mathbb{R}^d$

$$\text{dist}(x, y) = \|x - y\|_2^2 = \underbrace{(x(1) - y(1))^2}_{\text{term 1}} + \underbrace{(x(2) - y(2))^2}_{\text{term 2}} + \dots + (x(d) - y(d))^2$$



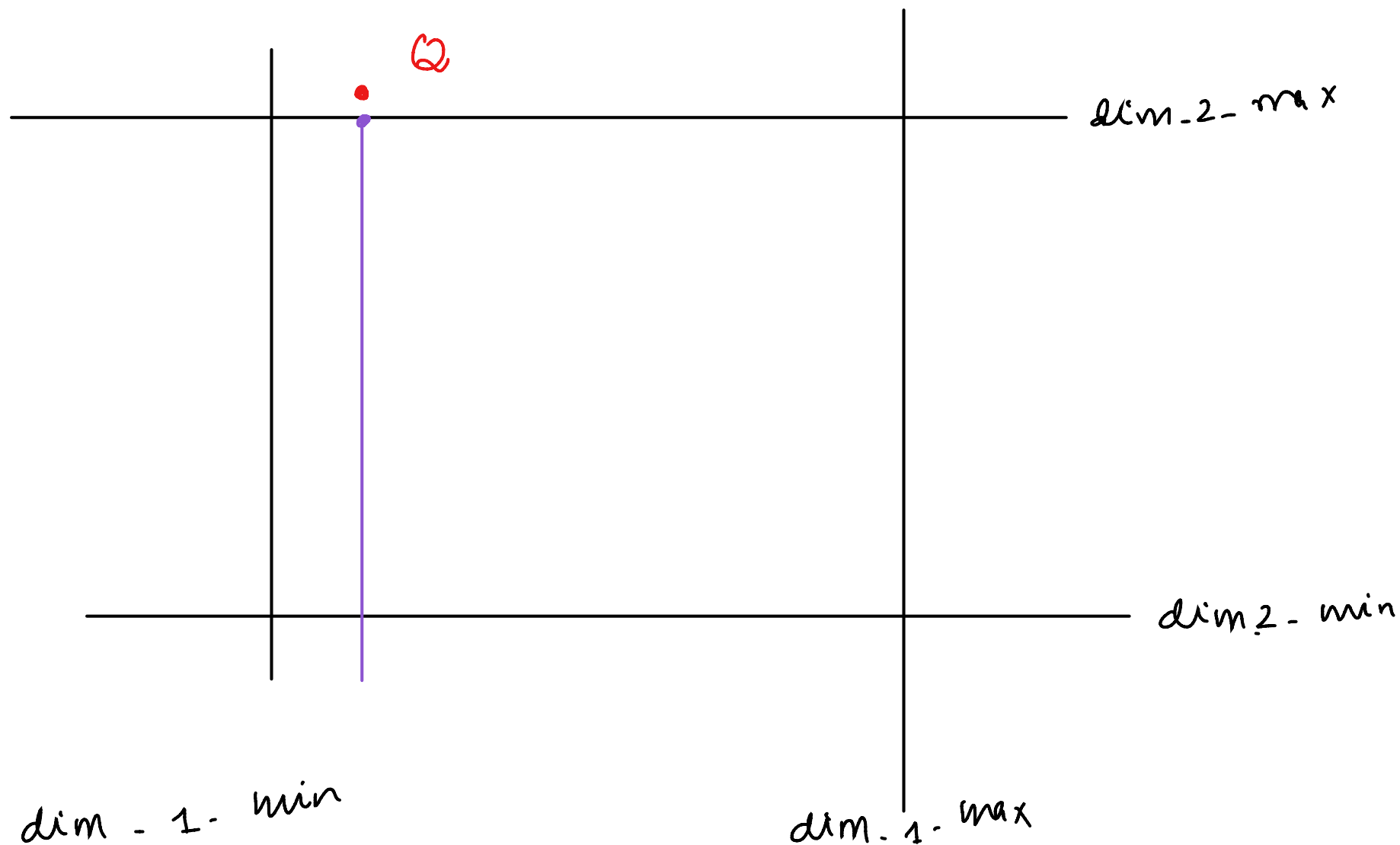
Loss is separable
but terms interact
via dependency
introduced by the set

Distance between a point $x \in \mathbb{R}^d$, and a set \mathcal{Y}



$$\text{dist}(x, \mathcal{Y}) = \min_{y \in \mathcal{Y}} \text{dist}(x, y)$$

$$\hat{y} = \arg \min_{y \in \mathcal{Y}} \text{dist}(x, y)$$



$$\text{dist}(x, y) = \min_{y \in \text{BB}} \underbrace{(x(1) - y(1))^2}_{\text{term 1}} + \underbrace{(x(2) - y(2))^2}_{\text{term 2}} + \dots + (x(d) - y(d))^2$$

\uparrow
 set is simple \Rightarrow term 1, term 2 \dots can be
 taken care of separately

Bloom Filter

(Quick & Dirty)

Setting: Prohibitively large number of items in your World/Universe

Goal: Insert and Search fast

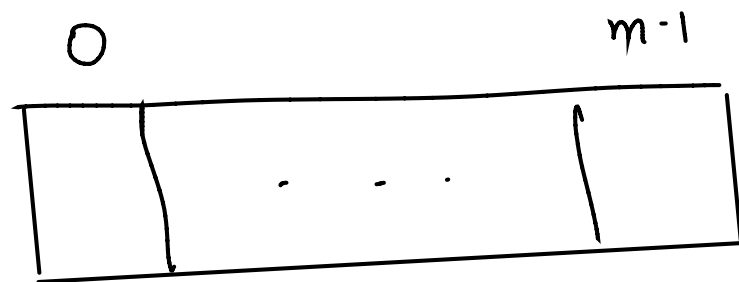
Example: URL needs to be checked for threat

* False positive is okay

* False negative is not allowed

* S = Set of keys

* B = bit array



* k hash functions h_1, \dots, h_k s.t. $h_i: S \rightarrow \{0, \dots, m-1\}$

* Insert: x

$$B[h_1(x)] = B[h_2(x)] = \dots = B[h_k(x)] = 1$$

* Search: x

yes, if $B[h_1(x)] = \dots = B[h_k(x)] = 1$

Toy Example

$$n = 20, \quad S = \{0, \dots, 9\},$$

$$h_1(x) = x \bmod 20, \quad h_2(x) = 3x \bmod 20, \quad h_3(x) = 7x \bmod 20$$

00000 00000 00000 00000

Insert: 4, $h_1(4) = 4, \quad h_2(4) = 12, \quad h_3(4) = 8$

00010 00100 01000 00000

Counting Bloom Filter

* Insert : x

$$B[h_1(x)] \leftarrow B[h_1(x)] + 1, \dots, B[h_k(x)] \leftarrow B[h_k(x)] + 1$$

* Search : x has occurred more than a threshold θ

Yes : if $B[h_1(x)] > \theta$ and $B[h_2(x)] > \theta$ and ... $B[h_k(x)] > \theta$

No if $B[h_1(x)] < \theta$ or $B[h_2(x)] < \theta$... or $B[h_k(x)] < \theta$