

Imp: Time and Space Complexity of KNN

Task $x_q \rightarrow y_q$ what would be time and space complexity.

Input: $D_{\text{train}}, k, x_q \in \mathbb{R}^d$
 $\hookrightarrow d$ -dimensional

output: y_q

Pseudo code

for x_i in D_{train} \rightarrow npts
 - compute $d(x_i, y_q) \rightarrow d_i$
 - keep the smallest k -distances
 $\rightarrow (x_i, y_i, d_i)$
 $\downarrow \quad \uparrow$
 KNN NPT = []
 \leftarrow $O(1)$
 $\therefore k$ is small

\therefore total ^{time} ~~space~~ complexity = $O(nd)$

~~time complexity~~ and

$\boxed{\text{if } d \ll n \Rightarrow O(n)}$

$n \rightarrow$ is significantly large.

space complexity = at evaluation time = space
that is needed to evaluate $x_q \rightarrow y_q$ $O(nd)$

i.e. we have to store the whole of my Dtrain
in memory for calculation.

Let's go little more deeper. So, far we
get time complexity = $O(nd)$

Space complexity = $O(nd)$

$n = 364K$ (Amazon food review example)

$d = 100K \rightarrow$ (Bow)

$nd = 36400m \rightarrow \underline{36 GB}$ of data (Approx)

So, almost needing 36 GB of RAM to run
my test.

Ok, let's talk about time complexity

$O(nd) = 36$ Billions computation

\rightarrow It is huge

* So, we are having space complexity.
large time complexity is higher.

This is the biggest ~~space~~ problem with KNN
That is why people don't use it (most of the
cases)