k-partitions

Page No.
Date : / /

for fixed k

	time	data size
k-partition	$\frac{n^2}{k} d^2$	$k(d+1)$
No partition	$k^2(d+1)^2 d^2$	$(d+1)$

$$f_{train}(k) = \frac{n^2}{k} d^2 + k^2 (d+1)^2 d^2$$

$$f'(k) = -\frac{n^2 d^2}{k^2} + 2k(d+1)^2 d^2$$

$$f'(k) = 0$$

$$\Rightarrow 2k\, d^2 (d+1)^2 = \frac{n^2 d^2}{k^2}$$

$$\Rightarrow k = \left(\frac{n^2}{2(d+1)^2}\right)^{1/3}$$

$$f_{min}(k) = \frac{n^2 d^2}{n^{2/3}} (2^{1/3})(d+1)^{2/3}$$

$$+ \frac{n^{4/3}}{2^{2/3}} \frac{d^2 (d+1)^2}{(d+1)^{4/3}}$$

$$= 2\left(\frac{n^{4/3} d^2 (d+1)^{2/3}}{2^{1/3}\, 2^{2/3}}\right) + \left(\frac{n^{4/3} d^2 (d+1)^{4/3}}{2^{2/3}}\right)$$

Page No.

Date: / /

$$\Rightarrow f_{mdn}(k) = \frac{3}{2^{2/3}} n^{4/3} d^2 (d+1)^{4/3}$$

$$= O\left(n^{4/3} d^{8/3} \right)$$

#

for Variable k

	time	space
k_1-partition	$\frac{n^2}{k_1} d^2$	$k_1(d+1)$
K_2-partition	$\frac{k_1^2 (d+1)^2 d^2}{K_2}$	$K_2(d+1)$
K_3-partition	$\frac{k_2^2 (d+1)^2 d^2}{K_3}$	$k_3(d+1)$
	$\vdots$	$\vdots$
All remaing data	$k_m^2 (d+1)^2 d^2$	$(d+1)$

$$f(k) = \frac{n^2 d^2}{k_1} + k_1 \cdot \frac{k_1}{k_2} (d+1)^2 d^2$$
$$+ \frac{k_2^2}{k_3} \cdot d^2 (d+1)^2 + \dots + k_m^2 (d+1)^2 d^2$$

for simplicity let's say $k_i = r\, k_{i+1}$

$k_1 = r k_2$, $k_2 = r k_3$

$$\Rightarrow f(k) = \frac{n^2 d^2}{k_1} + k_1 \cdot r\, d^2 (d+1)^2$$
$$+ \frac{k_2}{1} \cdot \frac{k_2}{k_3} d^2 (d+1)^2 + \dots$$

$$k_2 = \frac{k_1}{r}, \quad \frac{k_2}{k_3} = r, \quad \frac{k_2^2}{k_3} = k_1$$

$$\frac{k_3^2}{k_4} = k_3 \cdot \frac{k_3}{k_4} = \frac{k_2}{r} \times r = \frac{k_1}{r^2} \times r = \frac{k_1}{r}$$

Page No.

Date : / /

$$\Rightarrow f_{*}(k) = \frac{n^2 d^2}{k_1} + k_1 d^2 (d+1)^2 \left(r + 1 + \frac{1}{r} + \cdots \right)$$

let $r > 1$

$$O(f_{*}(k)) = \frac{n^2 d^2}{k_1} + k_1 d^2 (d+1)^2 \left(\frac{r}{1 - 1/r} \right)$$

$$= \frac{n^2 d^2}{k_1} + k_1 d^2 (d+1)^2 \left(\frac{r^2}{r-1} \right)$$

$$= g(k) \text{ (let)}$$

$$g'(k) = -\frac{n^2 d^2}{k_1^2} + d^2 (d+1)^2 \left[\frac{r^2}{r-1} \right]$$

$$\Rightarrow k_1 = \sqrt{\frac{n^2 (r-1)}{(d+1)^2 (r^2)}}$$

$$= \frac{n (\sqrt{r-1})}{r (d+1)}$$

$$O(f(k)) = \frac{\cancel{n^2} d^2 r (d+1)}{\cancel{n} (\sqrt{r-1})} + \frac{n d^2 (d+1) r}{(\sqrt{r-1})}$$

$$= O\left(n d^3 \right)$$

k-partition + Streaming points for fixed k

	time complexity	Output size
k-partitions	$\left(\frac{n}{k}\right)(d)^3$	$(d+1)\,k$
All data	$(d+1)\,k\,d^3$	$(d+1)$

total time

$$f(k) = \frac{n}{k}(d)^3 + k\,d^3(d+1)$$

$$f'(k) = d^3(d+1) - \frac{n\,d^3}{k^2}$$

$$f'(k) = 0, \Rightarrow k^2 = \frac{n}{d+1}$$

$$\Rightarrow k = \sqrt{\frac{n}{d+1}}$$

$$f(k) = \frac{n}{\sqrt{n}}(d^3)\sqrt{(d+1)} + \cancel{k}\,\frac{d^3(d+1)\sqrt{n}}{\sqrt{(d+1)}}$$

$$O(f(k)) = \sqrt{nd}\;d^3$$

$$= n^{1/2}\,d^{7/2}$$