

Երևանի Պետական Համալսարան

Տրանսֆերային ուսուցման որոշակի
մեթոդի ընհանրացման սխալանքի
գնահատման մասին

Գևորգ Մինասյան

31 Մայիսի 2019

Տրանսֆերային ուսուցում

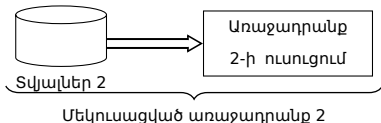
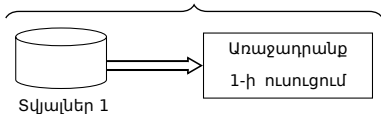
Տրանսֆերային ուսուցում

Դասական մեքենայական ուսուցում

Տրանսֆերային ուսուցում

Դասական մեքենայական ուսուցում

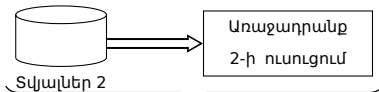
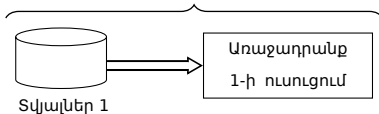
Մեկուսացված առաջադրանք 1



Տրանսֆերային ուսուցում

Դասական մեքենայական ուսուցում

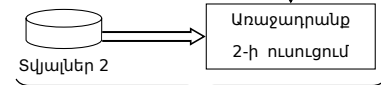
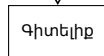
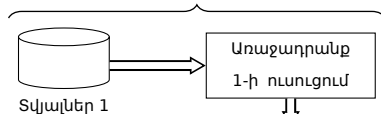
Մեկուսացված առաջադրանք 1



Մեկուսացված առաջադրանք 2

Տրանսֆերային ուսուցում

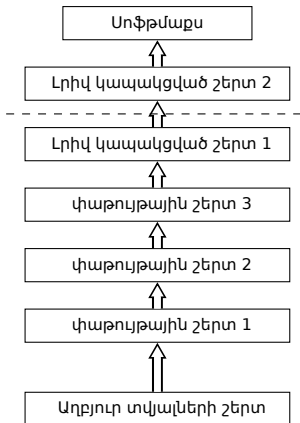
Աղբյուրի տիրույթ



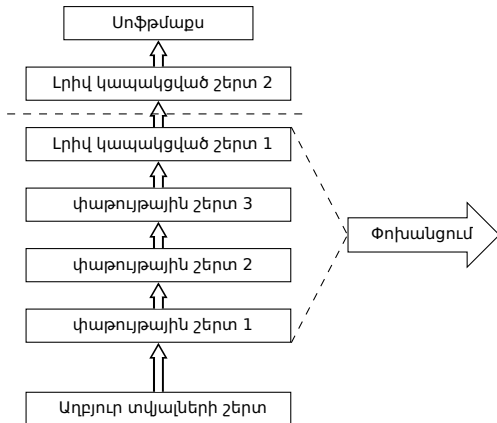
Թիրախի տիրույթ

Տրանսֆերային ուսուցումը վարժեցված
նեյրոնային ցանցի միջոցով

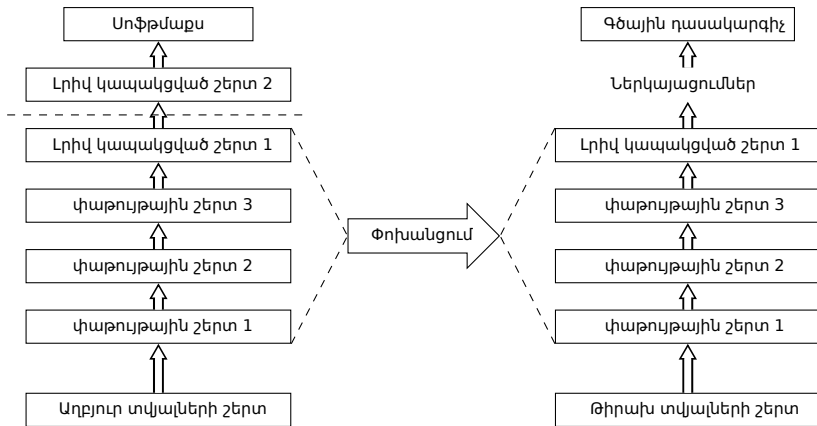
Տրանսֆերային ուսուցումը վարժեցված նեյրոնային ցանցի միջոցով



Տրանսֆերային ուսուցումը վարժեցված նեյրոնային ցանցի միջոցով



Տրանսֆերային ուսուցումը վարժեցված ներոնային ցանցի միջոցով



Նշանակումներ

- Ք բոլոր հնարավոր տվյալների օրինակների բազմություն:

Նշանակումներ

- \mathcal{X} բոլոր հնարավոր տվյալների օրինակների բազմություն:
- \mathcal{C} բոլոր հնարավոր դասերի բազմություն:

Նշանակումներ

- \mathcal{X} բոլոր հնարավոր տվյալների օրինակների բազմություն:
- \mathcal{C} բոլոր հնարավոր դասերի բազմություն:
- \mathcal{F} տվյալների ներկայացումների ֆունկցիաների դաս

$$f: \mathcal{X} \rightarrow \mathbb{R}^d, \forall f \in \mathcal{F}:$$

Նշանակումներ

- \mathcal{X} բոլոր հնարավոր տվյալների օրինակների բազմություն:
- \mathcal{C} բոլոր հնարավոր դասերի բազմություն:
- \mathcal{F} տվյալների ներկայացումների ֆունկցիաների դաս

$$f: \mathcal{X} \rightarrow \mathbb{R}^d, \forall f \in \mathcal{F}:$$

- $\exists R > 0 \forall f \in \mathcal{F} \|f(x)\| \leq R \forall x \in \mathcal{X}:$

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○●○○○○○○○
○○○○○○○○○○○○○○○

Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
○○○○○○○○○○○
○○○○○○○○○○○

Վերահսկվող առաջադրանքներ

Նշանակումներ

Նշանակումներ

- \mathcal{T} վերահսկվող առաջադրանքը բաղկացած է k հատ

$$\{c_1, c_2, \dots, c_k\} \subseteq \mathcal{C}$$

միմյանցից տարբեր դասերից, ուտել $k \geq 2$:

Նշանակումներ

- \mathcal{T} վերահսկվող առաջադրանքը բաղկացած է k հատ

$$\{c_1, c_2, \dots, c_k\} \subseteq \mathcal{C}$$

միմյանցից տարբեր դասերից, ուտել $k \geq 2$:

- $\mathcal{P}(\mathcal{T})$ վերասիկվող առաջադրանքների դիտարկվելու հավանականային բաշխումը:

Նշանակումներ

- \mathcal{T} վերահսկվող առաջադրանքը բաղկացած է k հատ

$$\{c_1, c_2, \dots, c_k\} \subseteq \mathcal{C}$$

միմյանցից տարբեր դասերից, ուտել $k \geq 2$:

- $\mathcal{P}(\mathcal{T})$ վերասիկվող առաջադրանքների դիտարկվելու հավանականային բաշխումը:
- $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k)$ k դասերից բաղկացած վերասիկվող առաջադրանքների դիտարկվելու հավանականային բաշխումը:

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○●○○○
○○○○○○○○○○○○○○○

Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
○○○○○○○○○○○
○○○○○○○○○○○

Վերահսկվող առաջադրանքներ

Նշանակումներ

Նշանակումներ

- $\mathcal{D}_c(x)$ c դասին համապատասխան հավանականային բաշխումն է, ցույց է տալիս, թե x օրինակը ինչքանով է համապատասխան c դասին:

Նշանակումներ

- $\mathcal{D}_c(x)$ c դասին համապատասխան հավանականային բաշխումն է, ցույց է տալիս, թե x օրինակը ինչքանով է համապատասխան c դասին:
- $\mathcal{D}_{\mathcal{T}}(x, c) = \mathcal{D}_{\mathcal{T}}(c)\mathcal{D}_c(x)$ \mathcal{T} վերահսկվող առաջադրանքի պիտակավորված տվյալների հավանականային բաշխումն է:

Նշանակումներ

- $\mathcal{D}_c(x)$ c դասին համապատասխան հավանականային բաշխումն է, ցույց է տալիս, թե x օրինակը ինչքանով է համապատասխան c դասին:
- $\mathcal{D}_{\mathcal{T}}(x, c) = \mathcal{D}_{\mathcal{T}}(c)\mathcal{D}_c(x)$ \mathcal{T} վերահսկվող առաջադրանքի պիտակավորված տվյալների հավանականային բաշխումն է:
- $S = \{(x_1, y_1), \dots, (x_M, y_M) \mid x_i \in \mathcal{X}, y_i \in \mathcal{T}\}$ \mathcal{T} առաջադրանքի պիտակավորված օրինակների ուսուցման բազմությունն է՝ ընտրված միմյանցից անկախ և $\mathcal{D}_{\mathcal{T}}(x, c)$ հավանականային բաշխումից:

Սահմանումներ

Հիմնական արդյունքներ

Նշանակումներ

Նշանակումներ

- \mathcal{T} առաջադրանքի բազմադաս դասակարգիչը ֆուկցիա է՝ $g: \mathcal{X} \rightarrow \mathbb{R}^k$, որի արժեքի կորդինատները ինդեքսավորված են այդ առաջադրանքի դասերով:

Նշանակումներ

- \mathcal{T} առաջադրանքի բազմադաս դասակարգիչը ֆուկցիա է՝ $g: \mathcal{X} \rightarrow \mathbb{R}^k$, որի արժեքի կորդինատները ինդեքսավորված են այդ առաջադրանքի դասերով:
- $l(\{g(x)_y - g(x)_{y'}\}_{y \neq y'})$ -ը $(x, y) \in \mathcal{X} \times \mathcal{T}$ կետում g դասակարգիչով պայմանավորված կորուստն է:

Նշանակումներ

- \mathcal{T} առաջադրանքի բազմադաս դասակարգիչը ֆուկցիա է՝ $g: \mathcal{X} \rightarrow \mathbb{R}^k$, որի արժեքի կորդինատները ինդեքսավորված են այդ առաջադրանքի դասերով:
- $l(\{g(x)_y - g(x)_{y'}\}_{y \neq y'})$ -ը $(x, y) \in \mathcal{X} \times \mathcal{T}$ կետում g դասակարգիչով պայմանավորված կորուստն է:
- $l(v) = \max\{0, 1 + \max_i\{-v_i\}\}$:
- $l(v) = \log_2(1 + \sum_i e^{-v_i})$:

Սահմանումներ

Հիմնական արդյունքներ

Նշանակումներ

Նշանակումներ

- \mathcal{T} առաջադրանքի համար g դասակարգիչի կորուստը հետևյալն է՝

$$L(\mathcal{T}, g) = \mathbb{E}_{(x, c) \sim \mathcal{D}_{\mathcal{T}}} [l(\{g(x)_c - g(x)_{c'}\}_{c \neq c'})]:$$

Նշանակումներ

- \mathcal{T} առաջադրանքի համար g դասակարգիչի կորուստը հետևյալն է՝

$$L(\mathcal{T}, g) = \mathbb{E}_{(x, c) \sim \mathcal{D}_{\mathcal{T}}} [l(\{g(x)_c - g(x)_{c'}\}_{c \neq c'})] :$$

- $\mathcal{V} = \{W : \|W\|_{\infty} \leq Q \text{ և } Q > 0\}$:

Նշանակումներ

- \mathcal{T} առաջադրանքի համար g դասակարգիչի կորուստը հետևյալն է՝

$$L(\mathcal{T}, g) = \mathbb{E}_{(x, c) \sim \mathcal{D}_{\mathcal{T}}} [l(\{g(x)_c - g(x)_{c'}\}_{c \neq c'})] :$$

- $\mathcal{V} = \{W : \|W\|_{\infty} \leq Q \text{ և } Q > 0\}$:
- $g(x) = Wf(x)$, որտեղ $W \in \mathcal{V}$:

Նշանակումներ

- \mathcal{T} առաջադրանքի համար g դասակարգիչի կորուստը հետևյալն է՝

$$L(\mathcal{T}, g) = \mathbb{E}_{(x, c) \sim \mathcal{D}_{\mathcal{T}}} [l(\{g(x)_c - g(x)_{c'}\}_{c \neq c'})] :$$

- $\mathcal{V} = \{W : \|W\|_{\infty} \leq Q \text{ և } Q > 0\}$:
- $g(x) = Wf(x)$, որտեղ $W \in \mathcal{V}$:
- $L(\mathcal{T}, f) = \inf_{W \in \mathcal{V}} L(\mathcal{T}, f, W)$:

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○○○○○○○
○○○○○○○○○○○○●○○○

Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
○○○○○○○○○○○○○○
○○○○○○○○○○○○

Վերահսկվող ներկայացումների գնահատման չափը

Նշանակումներ

Նշանակումներ

Վերահսկիչ միջին կորուստ

k դասերից բաղկացած առաջադրանքների վերահսկիչ միջին կորուստը f ներկայացման համար սահմանվում է որպես՝

$$L_k(f) = \mathbb{E}_{\mathcal{T} \sim \mathcal{P}} [L(\mathcal{T}, f) \mid |\mathcal{T}| = k]:$$

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○○○○○
○○○○○○○○○○○○●○

Ռադեմախերի բարոգությունը
○○○○

Հիմնական արդյունքներ
○○○○○○○○○○○
○○○○○○○○○○○

Վերահսկվող ներկայացումների գնահատման չափը

Նշանակումներ

Նշանակումներ

Էմպիրիկ վերահսկիչ միջին կորուստ

Դիցուք ունենք միմյանցից անկախ $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k)$ բաշխումից ընտրված N հատ առաջադրանքներ՝ $\mathcal{T}_1, \dots, \mathcal{T}_N$: Էմպիրիկ վերահսկիչ միջին կորուստը f ներկայացման համար հետևյալն է՝

$$\hat{L}_k(f) = \frac{1}{N} \sum_{i=1}^N L(\mathcal{T}_i, f):$$

Սահմանումներ

Հիմնական արդյունքներ

Ռադեմախերի բարդությունը

Ռադեմախերի բարդությունը

Ներկայացումների էմպիրիկ Ռադեմախերի բարդությունը

Դիցուք \mathcal{F} տվյալների ներկայացումների ֆունկցիաների ընտանիք է՝

$$\forall f \in \mathcal{F}, f: \mathcal{X} \rightarrow \mathbb{R}^d,$$

իսկ S -ը՝ m հզորությամբ տվյալների ֆիքսված օրինակների բազմություն՝

$$S = \{x_i | x_i \in \mathcal{X}, \forall i \in [m]\}:$$

Այդ դեպքում ներկայացումների \mathcal{F} ընտանիքի էմպիրիկ Ռադեմախերի բարդությունը ֆիքսված օրինակների S բազմության համար սահմանվում է հետևյալ կերպ՝

$$\hat{\mathcal{R}}_S(\mathcal{F}) = \frac{1}{m} \mathbb{E}_{\sigma \sim \{\pm 1\}^{md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^m \sum_{j=1}^d \sigma_{ij} f_j(x_i):$$

Անհավասարություն Ռադեմախերի բարդությունների վերաբերյալ

Թեորեմ

Դիցուք \mathcal{X} -ը և \mathcal{Y} -ը որևէ բազմություններ են, և $(x_1, x_2, \dots, x_n) \in X^n$: Տրված է նաև \mathcal{F} ֆունկցիաների բազմություն, որի կամայական $f \in \mathcal{F}$ ֆունկցիա \mathcal{X} բազմությունը արտապատկերում է \mathbb{R}^d եվկլիդյան տարածություն՝ $f: \mathcal{X} \rightarrow \mathbb{R}^d$: Դիցուք ունենք h_i ֆունկցիաներ՝

$$h_i: \mathbb{R}^d \times \mathcal{Y} \rightarrow \mathbb{R}$$

կամայական $i \in [n]$ համար: Ենթադրենք, որ բոլոր $h_i(v, y)$ ֆունկցիաները ինչ-որ L դրական հաստատունով Լիպշիցի հատկությամբ օժտված ֆունկցիաներ են՝ ըստ v -ի կամայական $y \in \mathcal{Y}$ համար: Այդ դեպքում տեղի ունի հետևյալ անհավասարությունը՝

$$\mathbb{E}_{\sigma \sim \{\pm 1\}^n} \left[\sup_{\substack{f \in \mathcal{F} \\ y \in \mathcal{Y}}} \sum_{i=1}^n \sigma_i h_i(f(x_i), y) \right] \leq \sqrt{2} L \mathbb{E}_{\sigma \sim \{\pm 1\}^{nd}} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^n \sum_{j=1}^d \sigma_{ij} f_j(x_i) \right] :$$

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○○○○○○○
○○○○○○○○○○○○○○○○

Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
●○○○○○○○○○○
○○○○○○○○○○

Վերահսկիչ միջին կորուստի գնահատականը

Ներկայացումների ուսուցման ալգորիթմը

Մենք հասնում ենք ներկայումս ֆունկցիա սովորելու ալգորիթմը հետևյալն է՝

$$(\hat{f}, \hat{W}) = \underset{\substack{f \in \mathcal{F} \\ W \in \mathcal{V}}}{\operatorname{argmin}} \hat{L}(\mathcal{T}, f, W),$$

որտեղ \hat{f} -ը փնտրվող ներկայացումն է:

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○○○○○○○
○○○○○○○○○○○○○○○○

Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
○○●○○○○○○○○
○○○○○○○○○○

Վերահսկիչ միջին կորուստի գնահատականը

Վերահսկիչ միջին կորուստի գնահատականը

Վերահսկիչ միջին կորուստի գնահատականը

■ $\rho_{min} = \min_{c \in \mathcal{T}} D_{\mathcal{T}}(c):$

Վերահսկիչ միջին կորուստի գնահատականը

- $\rho_{min} = \min_{c \in \mathcal{T}} D_{\mathcal{T}}(c)$:
- $m(c)$ այն T_i առաջադրանքների քանակը, որոնցում c դասն է մասնակցում:

Վերահսկիչ միջին կորուստի գնահատականը

- $\rho_{min} = \min_{c \in \mathcal{T}} D_{\mathcal{T}}(c)$:
- $m(c)$ այն T_i առաջադրանքների քանակը, որոնցում c դասն է մասնակցում:
- $m_{max} = \max_{c \in T} m(c)$:

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○○○○○○○
○○○○○○○○○○○○○○○○

Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
○○○○○○●○○○
○○○○○○○○○○

Վերահսկիչ միջին կորուստի գնահատականը

Վերահսկիչ միջին կորուստի գնահատականը

Տրանսֆերային ուսուցում
○○○○

Սահմանումներ
○○○○○○○○○○○○○○○○
○○○○○○○○○○○○○○○○

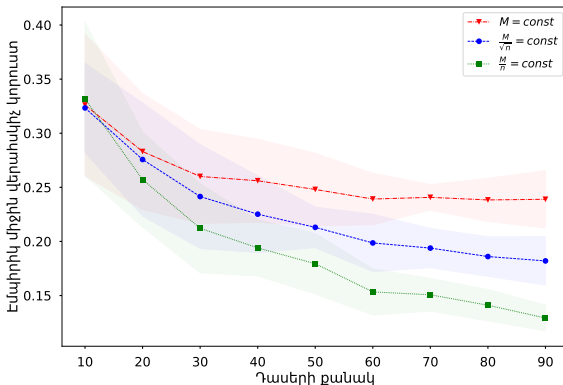
Ռադեմախերի բարդությունը
○○○○

Հիմնական արդյունքներ
○○○○○○○○●○
○○○○○○○○○○

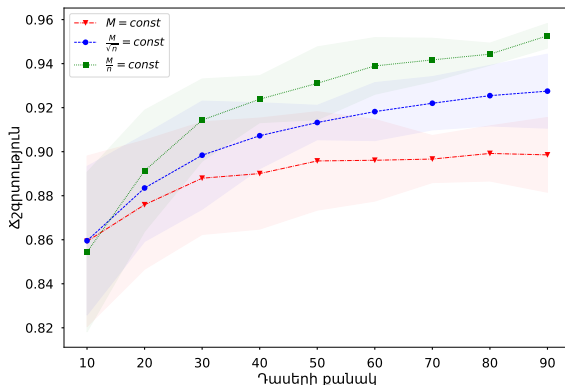
Վերահսկիչ միջին կորուստի գնահատականը

Վերահսկիչ միջին կորուստի գնահատականը

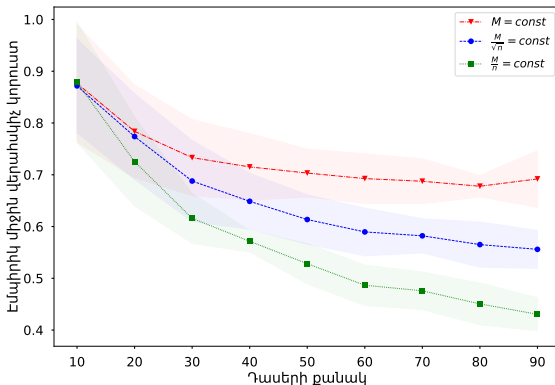
Իսկ եթե $|\mathcal{C}| \rightarrow \infty$, ապա $m_{\text{max}} \rightarrow 1$:



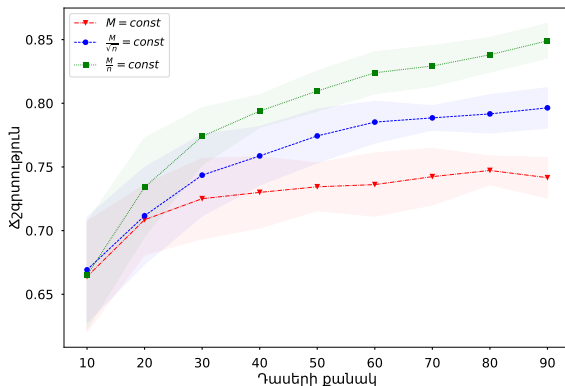
$k = 2$ քանակությամբ դասերից բաղկացած առաջադրանքի էմպիրիկ միջին վերահսկիչ կորուստի կախվածությունը՝ ներկայացումների ցանցի վարժեցման ժամանակ օգտագործված դասերի քանակից:



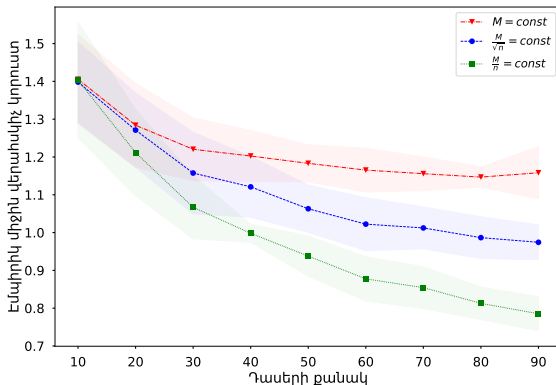
$k = 2$ քանակությամբ դասերից բաղկացած առաջադրանքի ճշգրտության կախվածությունը՝ ներկայացումների ցանցի վարժեցման ժամանակ օգտագործված դասերի քանակից:



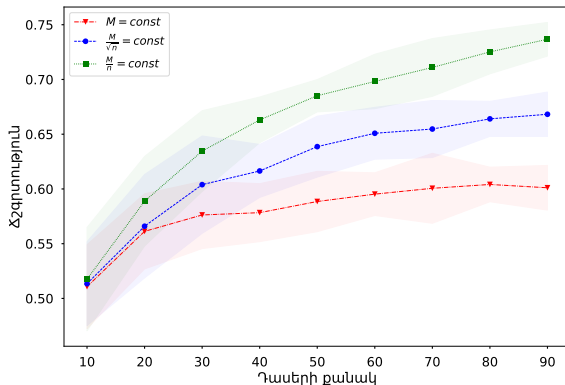
$k = 5$ քանակությամբ դասերից բաղկացած առաջադրանքի էմպիրիկ միջին վերահսկիչ կորուստի կախվածությունը՝ ներկայացումների ցանցի վարժեցման ժամանակ օգտագործված դասերի քանակից:



$k = 5$ քանակությամբ դասերից բաղկացած առաջադրանքի ճշգրտության կախվածությունը՝ ներկայացումների ցանցի վարժեցման ժամանակ օգտագործված դասերի քանակից:



$k = 10$ քանակությամբ դասերից բաղկացած առաջադրանքի էմպիրիկ միջին վերահսկիչ կորուստի կախվածությունը՝ ներկայացումների ցանցի վարժեցման ժամանակ օգտագործված դասերի քանակից:



$k = 10$ քանակությամբ դատերից բաղկացած առաջադրանքի ճշգրտության կախվածությունը՝ ներկայացումների ցանցի վարժեցման ժամանակ օգտագործված դատերի քանակից:

	$M = const$			$\frac{M}{\sqrt{n}} = const$			$\frac{M}{n} = const$		
	$k = 2$	$k = 5$	$k = 10$	$k = 2$	$k = 5$	$k = 10$	$k = 2$	$k = 5$	$k = 10$
$n = 10$	85.930	66.386	51.077	85.955	66.918	51.352	85.440	66.498	51.760
$n = 20$	87.590	70.856	56.126	88.350	71.168	56.594	89.130	73.404	58.875
$n = 30$	88.795	72.506	57.624	89.840	74.356	60.393	91.435	77.400	63.458
$n = 40$	89.005	73.002	57.835	90.725	75.866	61.643	92.385	79.386	66.309
$n = 50$	89.580	73.442	58.853	91.325	77.436	63.865	93.105	80.966	68.515
$n = 60$	89.610	73.614	59.526	91.820	78.522	65.085	93.895	82.396	69.817
$n = 70$	89.665	74.238	60.060	92.200	78.854	65.473	94.170	82.928	71.096
$n = 80$	89.920	74.724	60.410	92.545	79.166	66.401	94.430	83.810	72.518
$n = 90$	89.855	74.156	60.107	92.750	79.642	66.821	95.265	84.906	73.671

Իրակացված փորձարկումների ծրագիրը հասանելի է
հետևյալ հասցեով`

<https://github.com/Gevorg-Minasyan/fine-tune-tj>:

Շնորհակալություն