Persistent Homology Captures the Generalization of Neural Networks Without A Validation Set
Supplementary Material

Anonymous Author(s)
Affiliation
Address
email

1 Appendix I

2 Mathematical definitions are contained in this Appendix.

3 **Definition 1 (simplex)** *A k-simplex is a k-dimensional polytope which is the convex hull of its k + 1*
4 *vertices. i.e. the set of all convex combinations $\lambda_0 v_0 + \lambda_1 v_1 + ... + \lambda_k v_k$ where $\lambda_0 + \lambda_1 + ... + \lambda_k = 1$*
5 *and $0 \leq \lambda_j \leq 1 \ \ \forall j \in \{0, 1, ..., k\}$.*

6 Some examples of simplices are points (0-simplex), line segments (1-simplex), triangles (2-simplex)
7 or tetrahedrons (3-simplex).

8 **Definition 2 (simplicial complex)** *We define a simplicial complex $\mathcal{K}$ is a set of simplices that satisfies*
9 *the following conditions:*

10   *1. Every subset (or face) of a simplex in $\mathcal{K}$ also belongs to $\mathcal{K}$.*

11   *2. For any two simplices σ_1 and σ_2 in $\mathcal{K}$, if $\sigma_1 \cap \sigma_2 \neq \emptyset$, then $\sigma_1 \cap \sigma_2$ is a common subset, or*
12   *face, of both σ_1 and σ_2.*

13 **Definition 3 (directed flag complex)** *The directed flag complex (for example, on a directed graph*
14 *$G = (V, E)$), $FC(G)$, is defined to be the ordered simplicial complex whose k-simplices are all*
15 *ordered $(k + 1)$-cliques, i.e., $(k + 1)$-tuples $\sigma = (v_0, v_1, \ldots, v_k)$, such that $v_i \in V \ \ \forall i$, and*
16 *$(v_i, v_j) \in E$ for $i < j$.*

17 Boundary function, denoted by ∂ symbol, is a function that maps i-simplex to the sum of its (i-
18 1)-dimensional faces. Formally speaking, for an i-simplex $\sigma = [v_0, \ldots, v_i]$, its *boundary* (∂) is:
19

$$\partial_i \sigma = \sum_{j=0}^{i} [v_0, \ldots, \hat{v}_j, \ldots, v_i] \tag{1}$$

20 where the hat indicates the v_j is omitted.

21 This definition can be expanded to i-chains. For an i-chain $c = c_i \sigma_i$, $\partial_i(c) = \sum_i c_i \partial_i \sigma_i$.

22 We can now distinguish two special types of chains using the boundary map that will be useful to
23 define homology:

24   • The first one is an i-*cycle*, which is defined as an i-chain with empty boundary. In other
25   words, an i-chain c is an i-cycle if and only if $\partial_i(c) = 0$, i.e. $c \in Ker(\partial_i)$.

26   • An i-chain c is i-*boundary* if there exists an $(i + 1)$-chain d such that $c = \partial_{i+1}(d)$, i.e.
27     $c \in Im(\partial_{i+1})$.

Definition 4 (graph) *A graph G is a pair (V, E), where V is a finite set referred to as the vertices*
or nodes of G, and E is a subset of the set of unordered pairs $e = \{u, v\}$ of distinct points in V,
which we call the edges of G. Geometrically the pair $\{u, v\}$ indicates that the vertices u and v are
adjacent in G. A directed graph, or a digraph, is similarly a pair (V, E) of vertices V and edges E,
except the edges are ordered pairs of distinct vertices, i.e.,the pair (u, v) indicates that there is an
edge from u to v in G. In a digraph, we allow reciprocal edges, i.e., both (u, v) and (v, u) may be
edges in G, but we exclude loops, i.e., edges of the form (v, v).

Definition 5 (homology group) *Given these two special subspaces, i-cycles $Z_i(K)$ and i-*
boundaries $B_i(K)$ of $C_i(K)$, we now take the quotient space of $B_i(K)$ as a subset of $Z_i(K)$.
In this quotient space, there are only the i-cycles that do not bound an $(i + 1)$-complex, or i-voids of
K. This quotient space is called i-th homology group of the simplicial complex K:

$$H_i(K) = \frac{Z_i(K)}{B_i(K)} = \frac{Ker(\partial_i)}{Im(\partial_{i+1})} \tag{2}$$

where Ker and Im are the function kernel and image respectively.

The dimension of i-th homology is called the i-th *Betti number* of K, $\beta_i(K)$, where:

$$\beta_i(K) = dim(Ker(\partial_i)) - dim(Im(\partial_{i+1})) \tag{3}$$
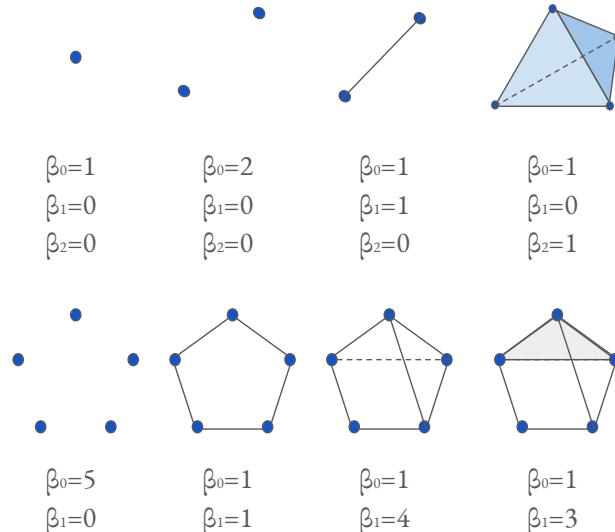
Figure 1: Betti numbers example

Definition 6 (Wasserstein distance) *The p-Wasserstein distance between two PDs D_1 and D_2 is*
the infimum over all bijections: $\gamma : D_1 \rightarrow D_2$ of:

$$d_W(D_1, D_2) = \Big(\sum_{x \in D_1} ||x - \gamma(x)||_\infty^p \Big)^{1/p} \tag{4}$$

where $|| - ||_\infty$ is defined for $(x, y) \in \mathbb{R}^2$ by $\max\{|x|, |y|\}$. The limit $p \rightarrow \infty$ defines the Bottleneck
distance. More explicitly, it is the infimum over the same set of bijections of the value

$$d_B(D_1, D_2) = \sup_{x \in D_1} ||x - \gamma(x)||_\infty. \tag{5}$$

Definition 7 (Persistence landscape) *Given a collection of intervals $\{(b_i, d_i)\}_{i \in I}$ that compose a*
PD, its persistence landscape *is the set of functions $\lambda_k : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ defined by letting $\lambda_k(t)$ be the k-th*
largest value of the set $\{\Lambda_i(t)\}_{i \in I}$ where:

$$\Lambda_i(t) = [\min\{t - b_i, d_i - t\}]_+ \tag{6}$$

48   *and $c_+ := \max(c, 0)$. The function λ_k is referred to as the k-layer of the persistence landscape.*

49   *Now we define a vectorization of the set of real-valued function that compose PDs on $\mathbb{N} \times \mathbb{R}$. For*
50   *any $p = 1, \ldots, \infty$ we can restrict attention to PDs D whose associated persistence landscape λ is*
51   *p-integrable, that is to say,*

$$||\lambda||_p = \left(\sum_{i \in \mathbb{N}} ||\lambda_i||_p^p \right)^{1/p} \tag{7}$$

52   *is finite. In this case, we refer to Equation (7) as the p-landscape norm of D. For $p = 2$, we define*
53   *the value of the* landscape kernel *or similarity of two vectorized PDs D and E as*

$$\langle \lambda, \mu \rangle = \left(\sum_{i \in \mathbb{N}} \int_{\mathbb{R}} |\lambda_i(x) - \mu_i(x)|^2 \, dx \right)^{1/2} \tag{8}$$

54   *where λ and μ are their associated persistence landscapes.*

55   λ_k is geometrically described as follows. For each $i \in I$, we draw an isosceles triangle with base the
56   interval (b_i, d_i) on the horizontal t-axis, and sides with slope 1 and -1. This subdivides the plane
57   into a number of polygonal regions that we label by the number of triangles contained on it. If P_k is
58   the union of the polygonal regions with values at least k, then the graph of λ_k is the upper contour of
59   P_k, with $\lambda_k(a) = 0$ if the vertical line $t = a$ does not intersect P_k.

60   **Definition 8 (Weighted Silhouette)** *Let $D = \{(b_i, d_i)\}_{i \in I}$ be a PD and $w = \{w_i\}_{i \in I}$ a set of*
61   *positive real numbers. The Silhouette of D weighted by w is the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by:*

$$\phi(t) = \frac{\sum_{i \in I} w_i \Lambda_i(t)}{\sum_{i \in I} w_i}, \tag{9}$$

62   *where*

$$\Lambda_i(t) = [\min\{t - b_i, d_i - t\}]_+ \tag{10}$$

63   *and $c_+ := \max(c, 0)$ When $w_i = |d_i - b_i|^p$ for $0 < p \leq \infty$ we refer to ϕ as the p-power-weighted*
64   *Silhouette of D. It defines a vectorization of the set of PDs on the vector space of continuous*
65   *real-valued functions on $\mathbb{R}$.*

66   **Definition 9 (Heat vectorizations)** *Considering PD as the support of Dirac deltas, one can con-*
67   *struct, for any $t > 0$, two vectorizations of the set of PDs to the set of continuous real-valued function*
68   *on the first quadrant $\mathbb{R}^2_{>0}$. The Heat vectorization is constructed for every PD D by solving the heat*
69   *equation:*

$$\begin{aligned}
\Delta_x(u) &= \partial_t u & on \ \Omega \times \mathbb{R}_{>0} \\
u &= 0 & on \ \{x_1 = x_2\} \times \mathbb{R}_{\geq 0} \\
u &= \sum_{p \in D} \delta_p & on \ \Omega \times 0
\end{aligned} \tag{11}$$

70   *where $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \leq x_2\}$, then solving the same equation after precomposing the data*
71   *of Equation (11) with the change of coordinates $(x_1, x_2) \mapsto (x_2, x_1)$, and defining the image of D to*
72   *be the difference between these two solutions at the chosen time t.*

73   *We recall that the solution to the Heat equation with initial condition given by a Dirac delta supported*
74   *at $p \in \mathbb{R}^2$ is:*

$$\frac{1}{4\pi t} \exp\left(-\frac{||p - x||^2}{4t} \right) \tag{12}$$

75   *To highlight the connection with normally distributed random variables, it is customary to use the the*
76   *change of variable $\sigma = \sqrt{2t}$.*

77   For a complete reference on vectorized persistence summaries and PH approximated metrics, see
78   Tauzin et al. [2], Berry et al. [1] and Giotto-TDA package documentation appendix[1].

---

[1]`https://giotto-ai.github.io/gtda-docs/0.3.1/theory/glossary.html#`
`persistence-landscape`

3
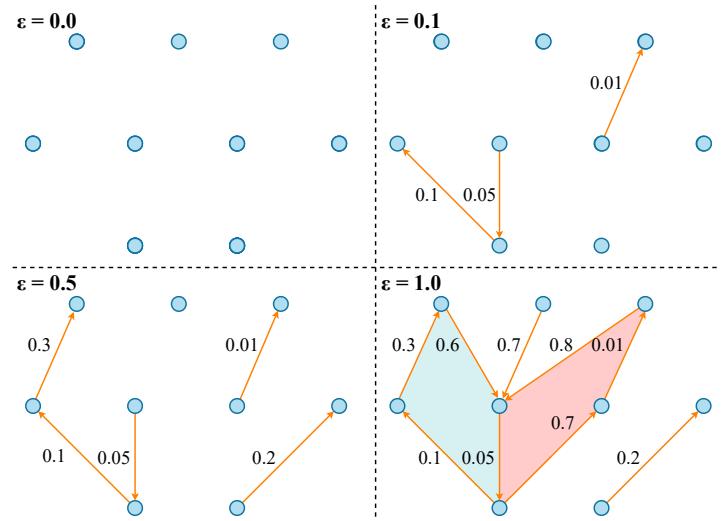
Appendix II

Figure 2: MLP Simplicial complex filtration example.
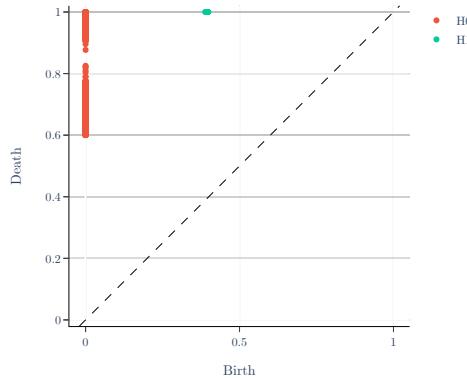

Figure 3: An example of Persistence Homology diagram
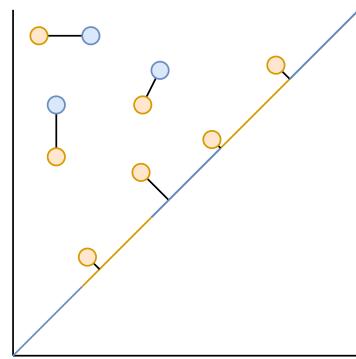
Figure 4: Persistence Homology diagram distance example.
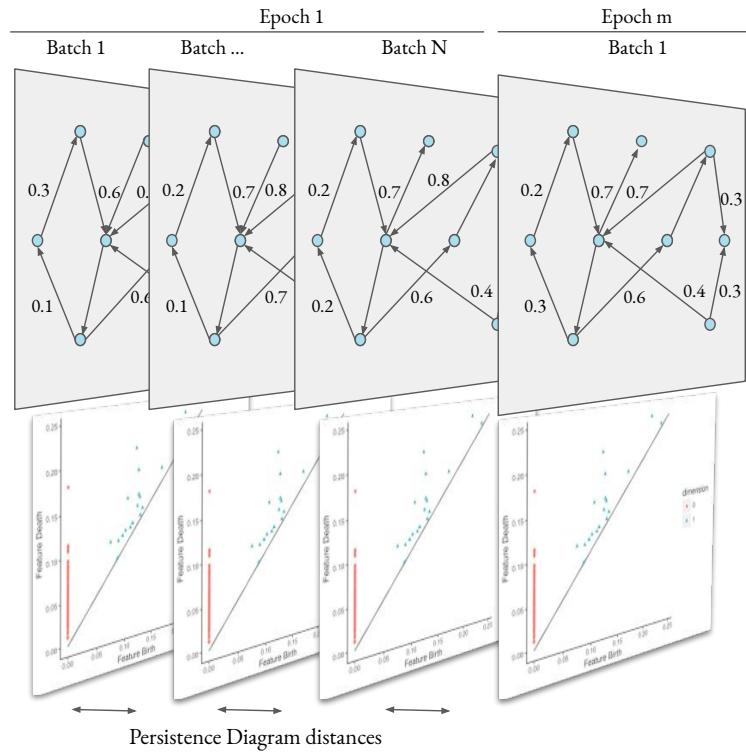
Figure 5: MLP learning evolution PD similarity process.

Appendix III

# Cumulative plots
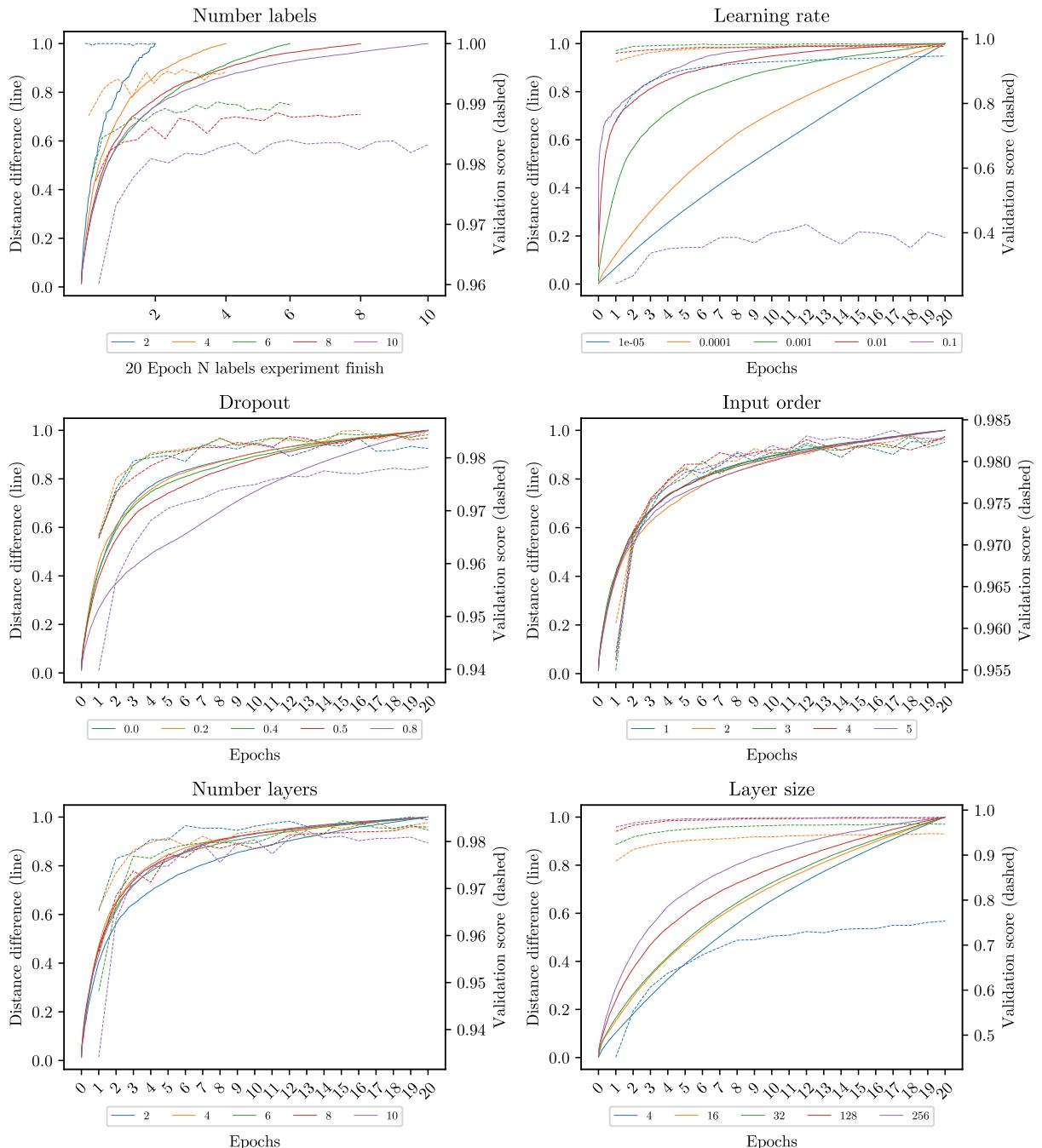
Figure 6: MNIST cumulative using Heat discretization.
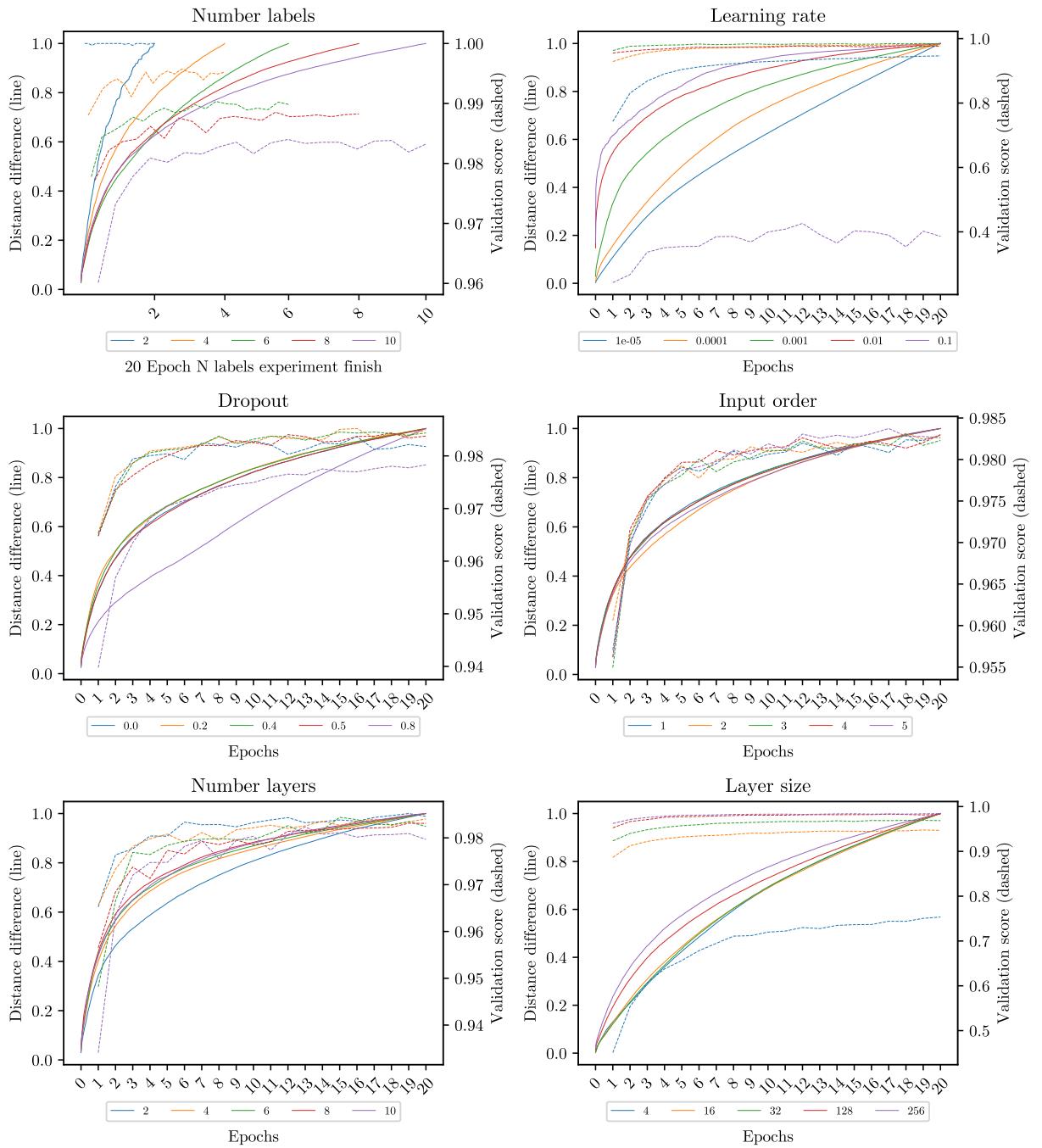
Figure 7: MNIST cumulative using Silhouette discretization.
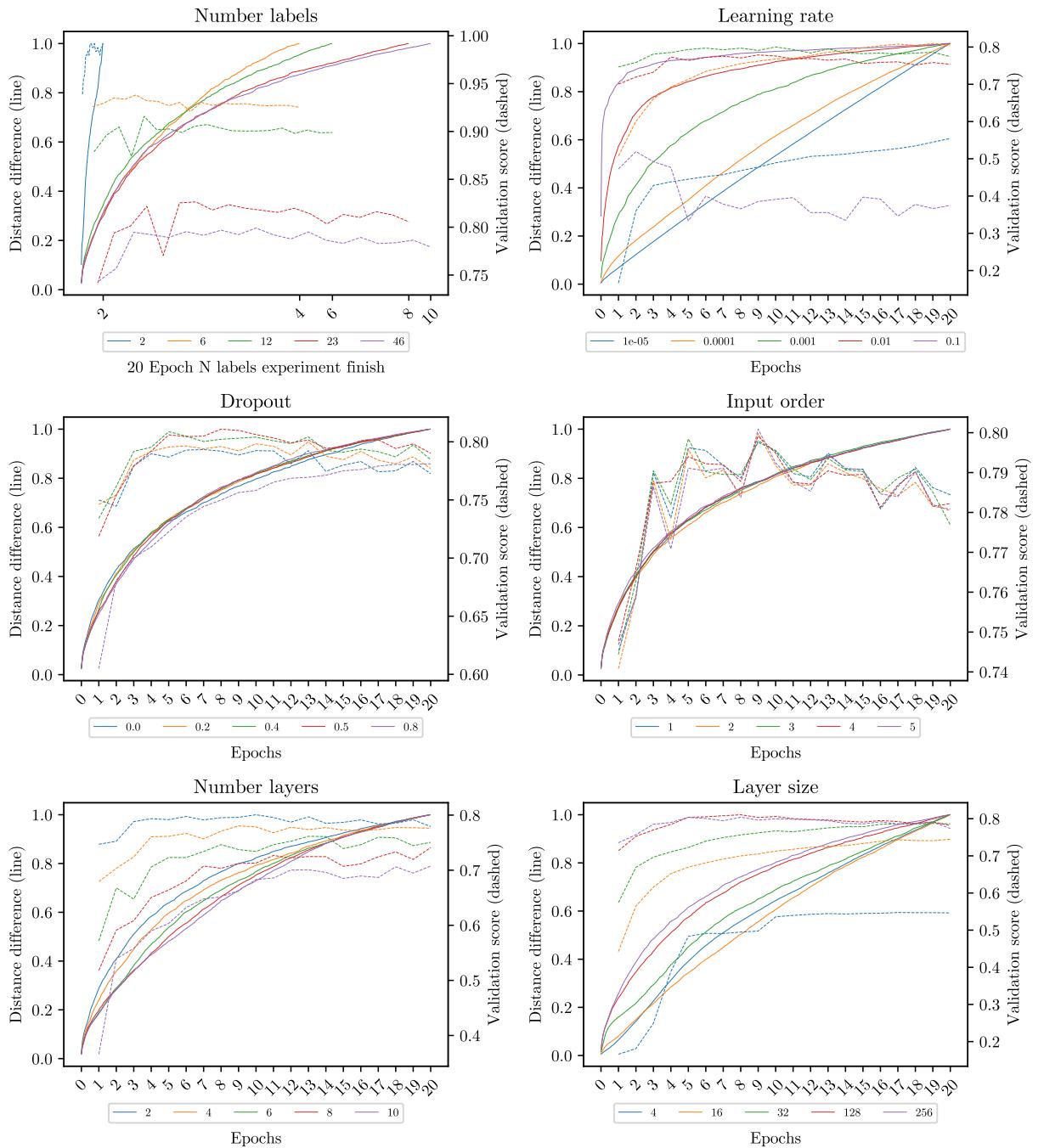
Figure 8: Reuters cumulative using Heat discretization.
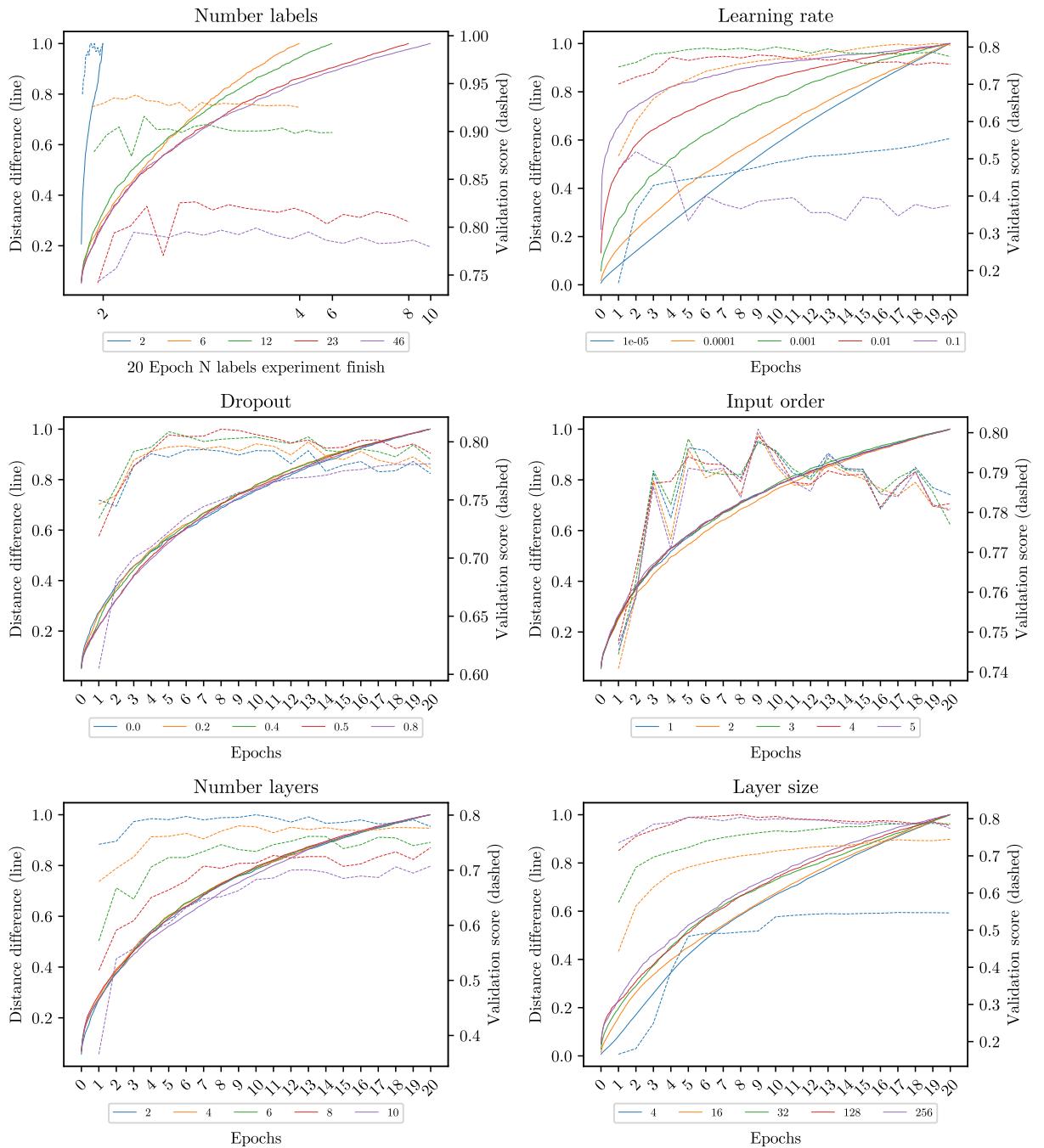
Figure 9: Reuters cumulative using Silhouette discretization.
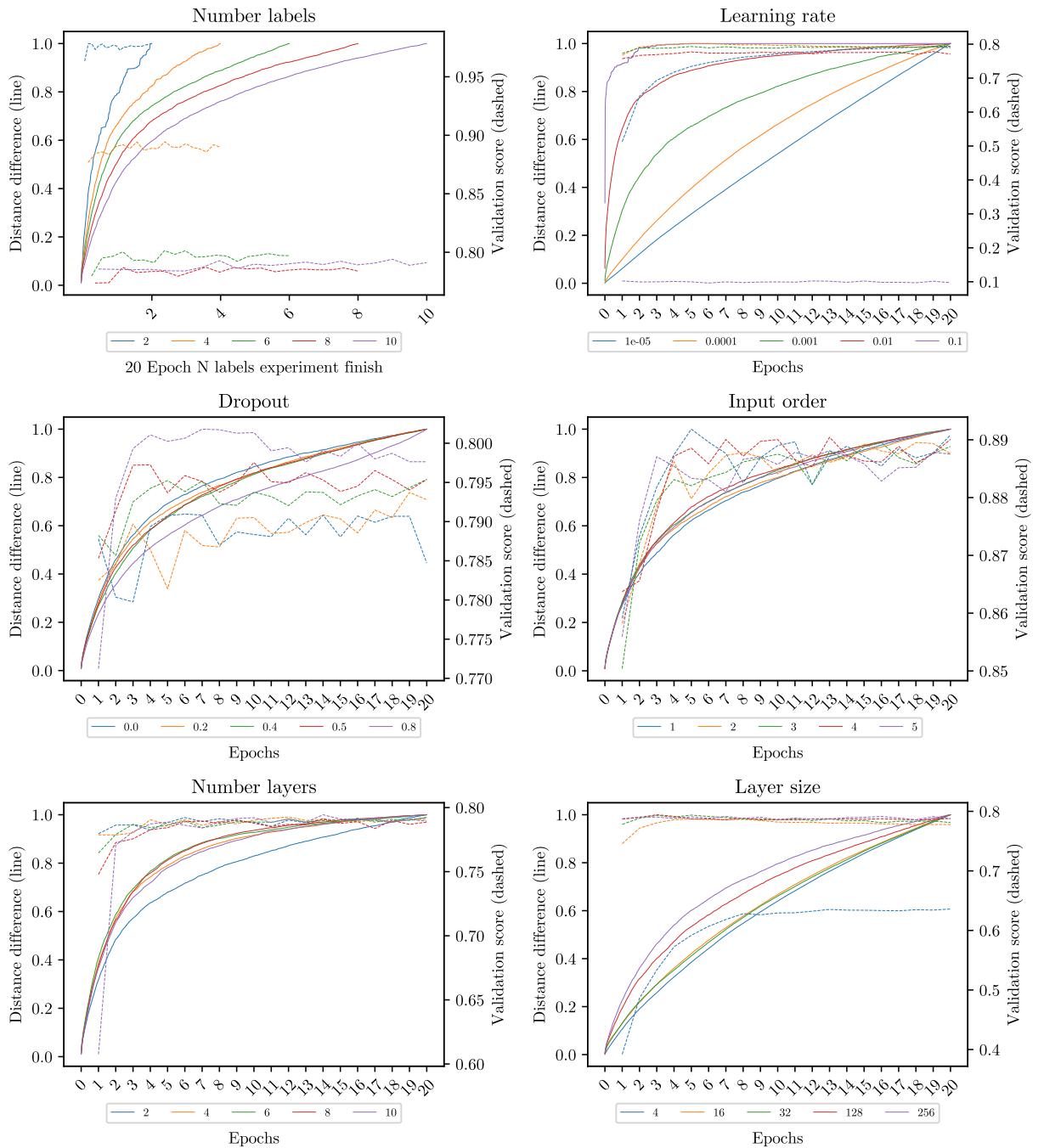
Figure 10: CIFAR-10 CNN cumulative using Heat discretization.
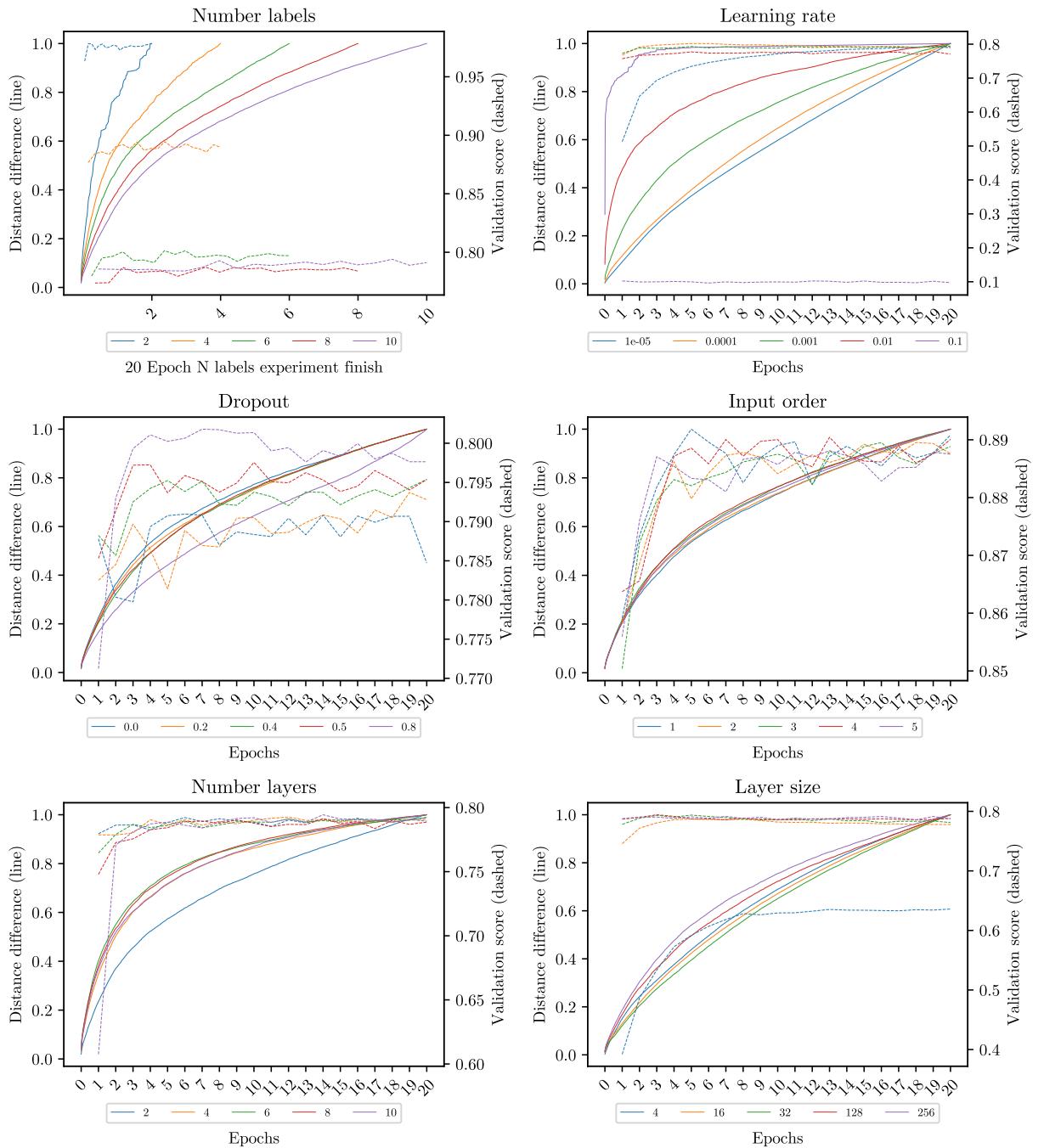
Figure 11: CIFAR-10 CNN cumulative using Silhouette discretization.
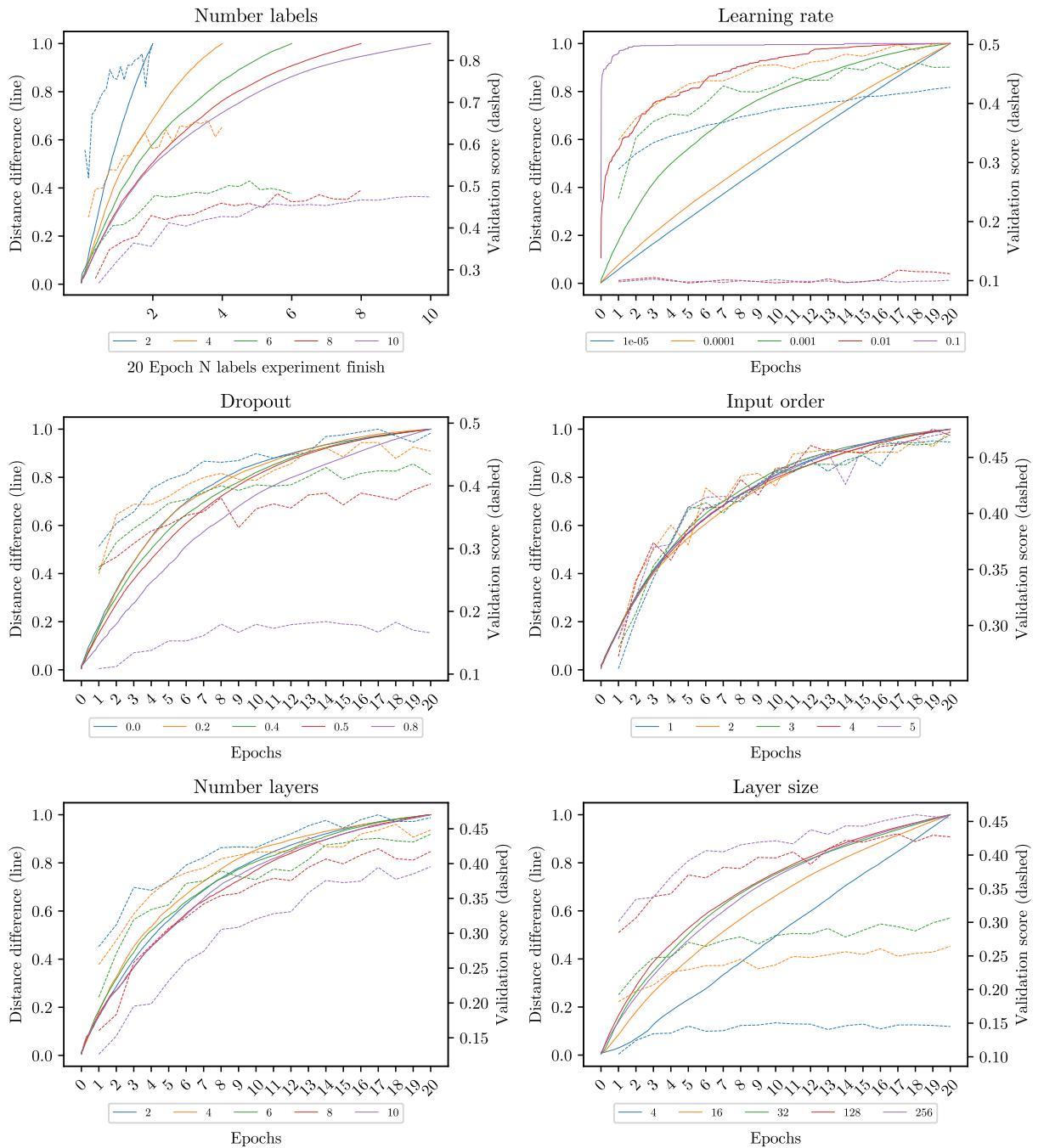
Figure 12: CIFAR-10 MLP cumulative using Heat discretization.
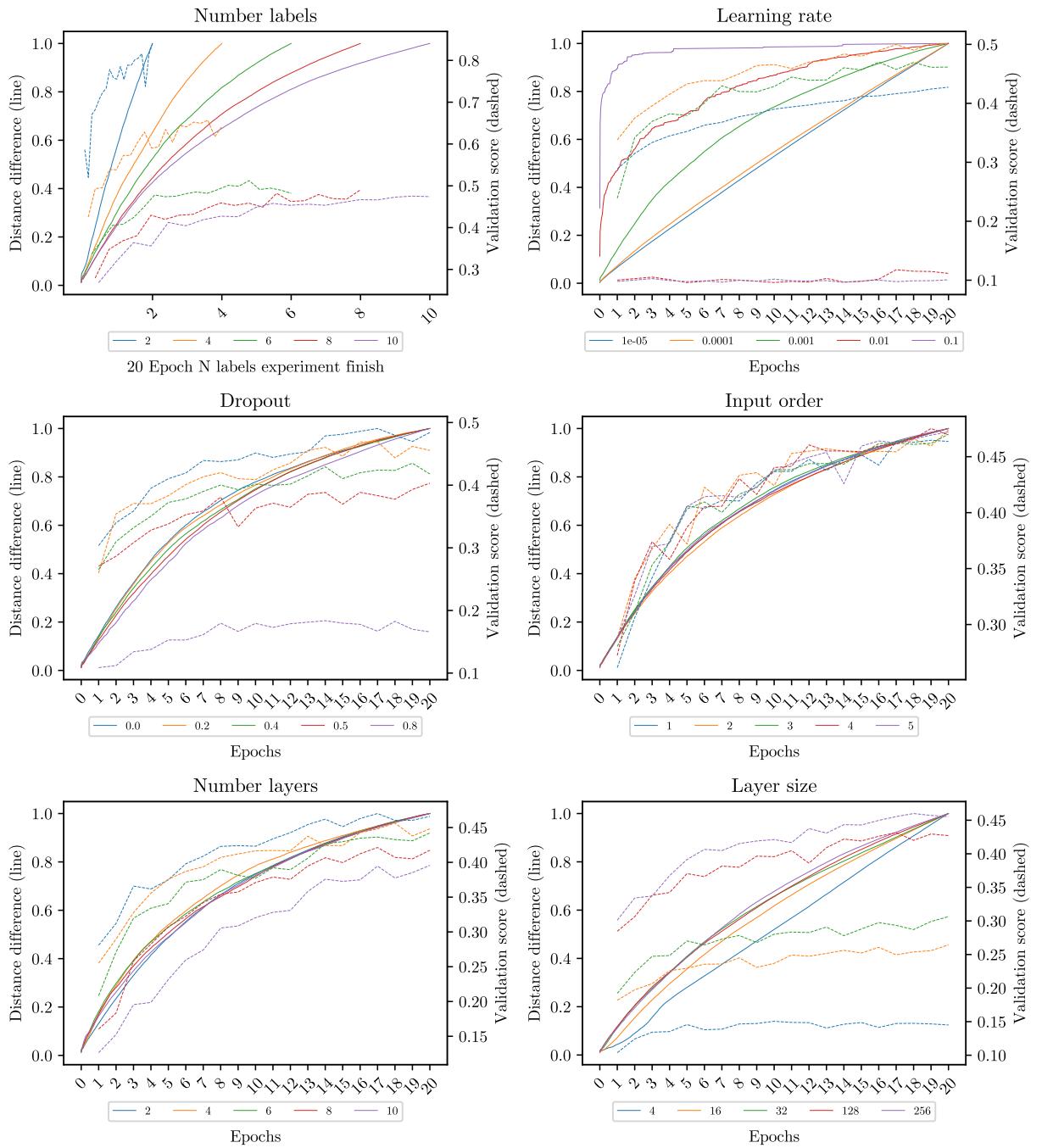
Figure 13: CIFAR-10 MLP cumulative using Silhouette discretization.
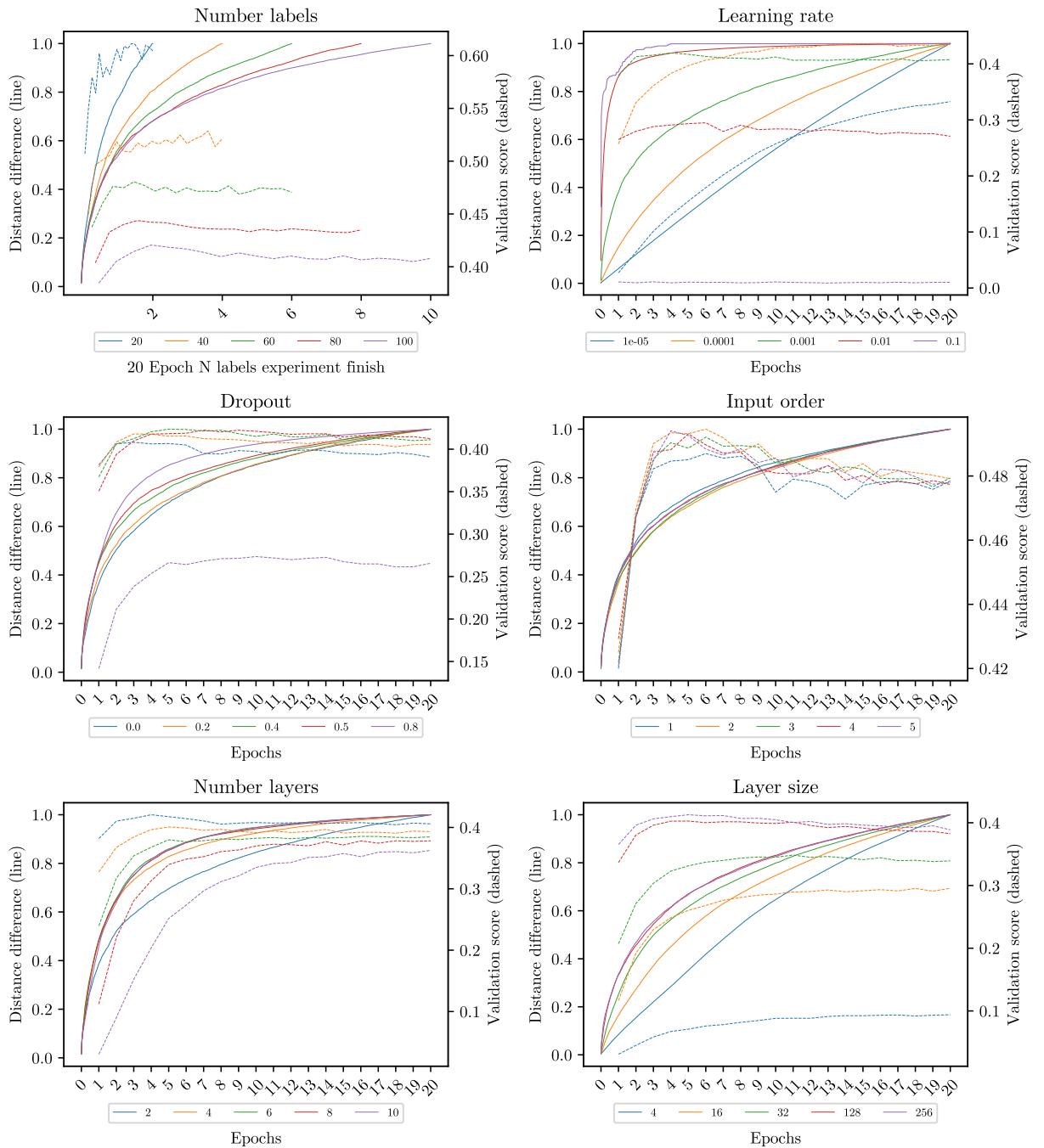
Figure 14: CIFAR-100 CNN cumulative using Heat discretization.
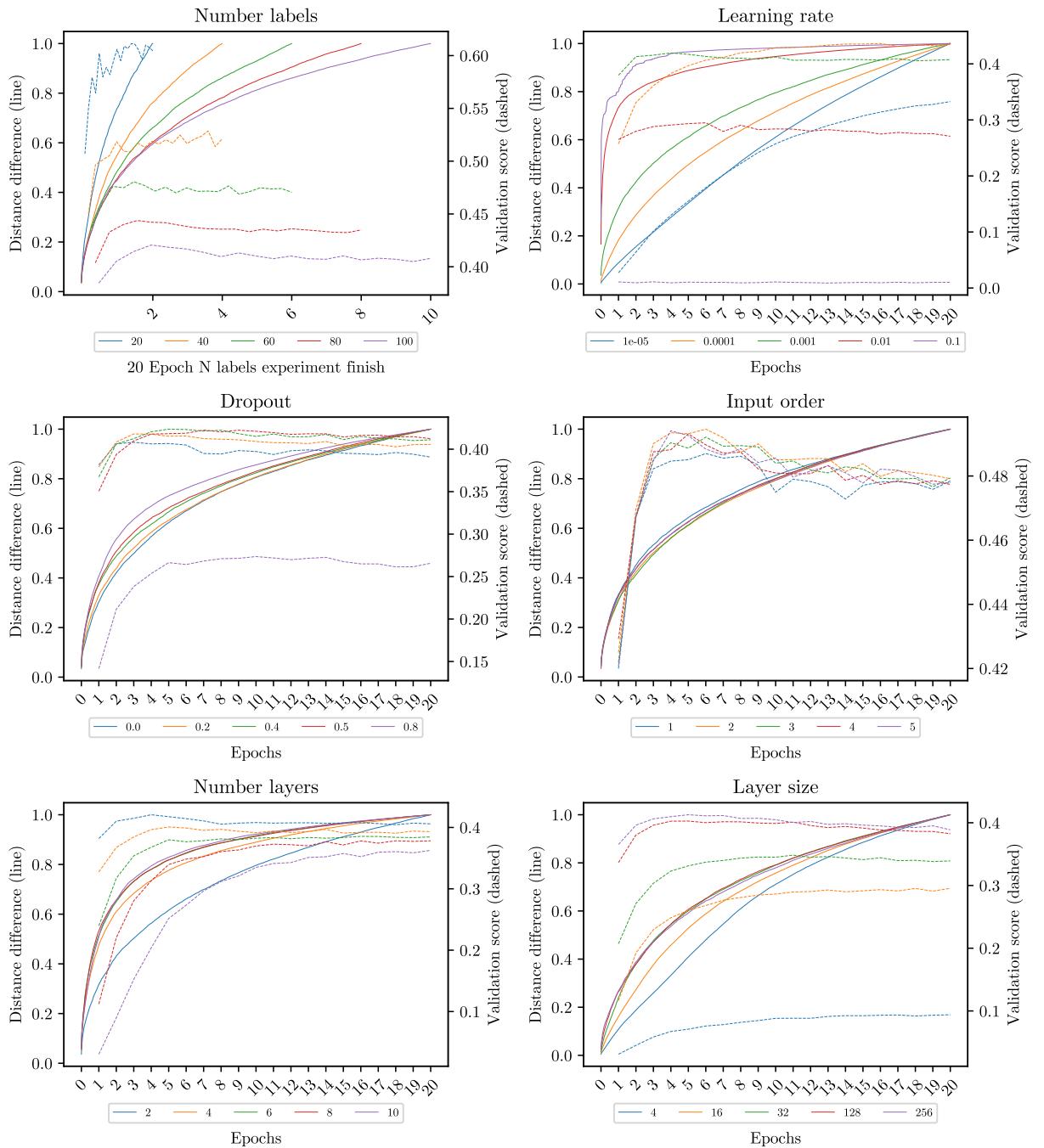
Figure 15: CIFAR-100 CNN cumulative using Silhouette discretization.
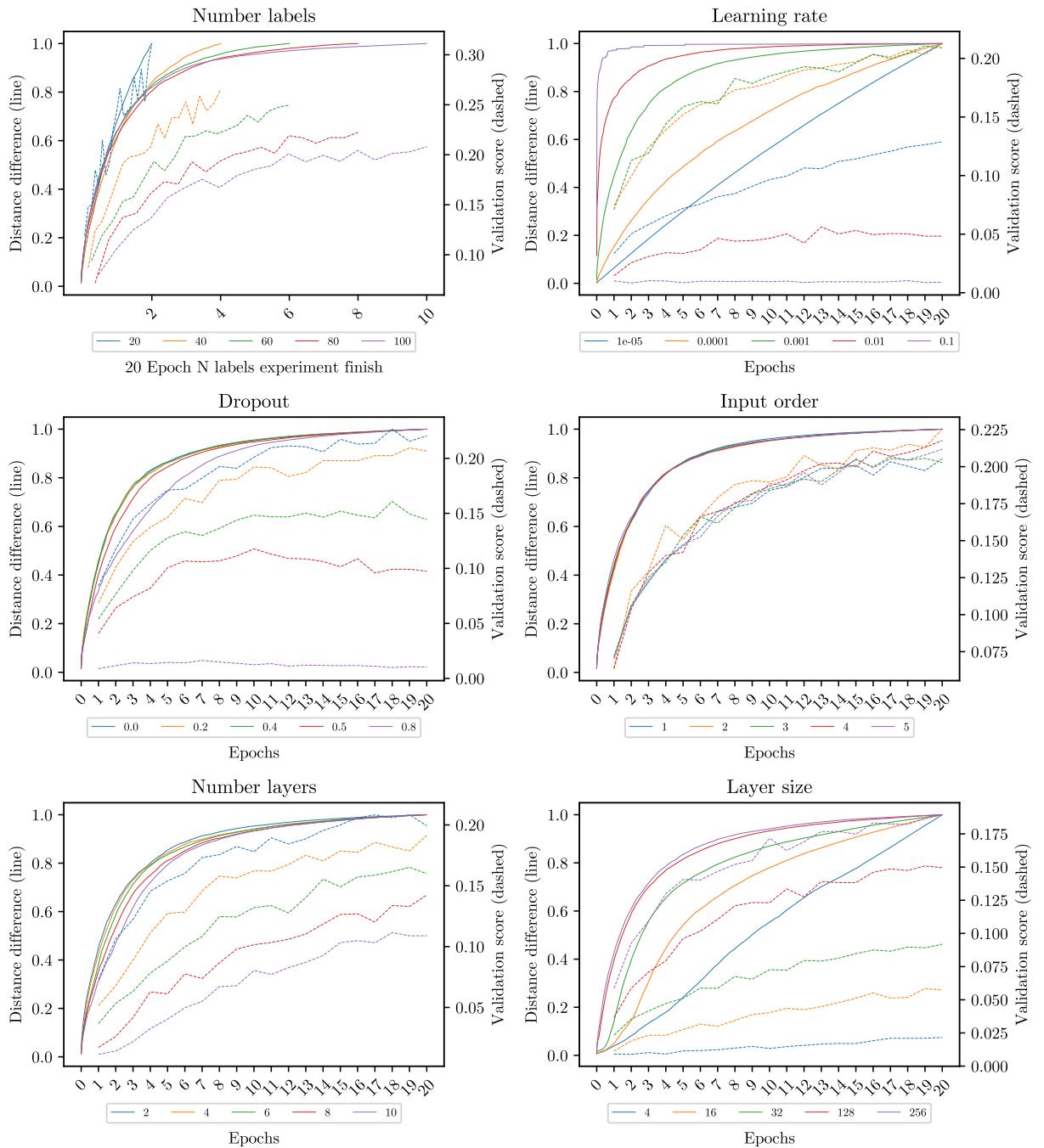
Figure 16: CIFAR-100 MLP cumulative using Heat discretization.
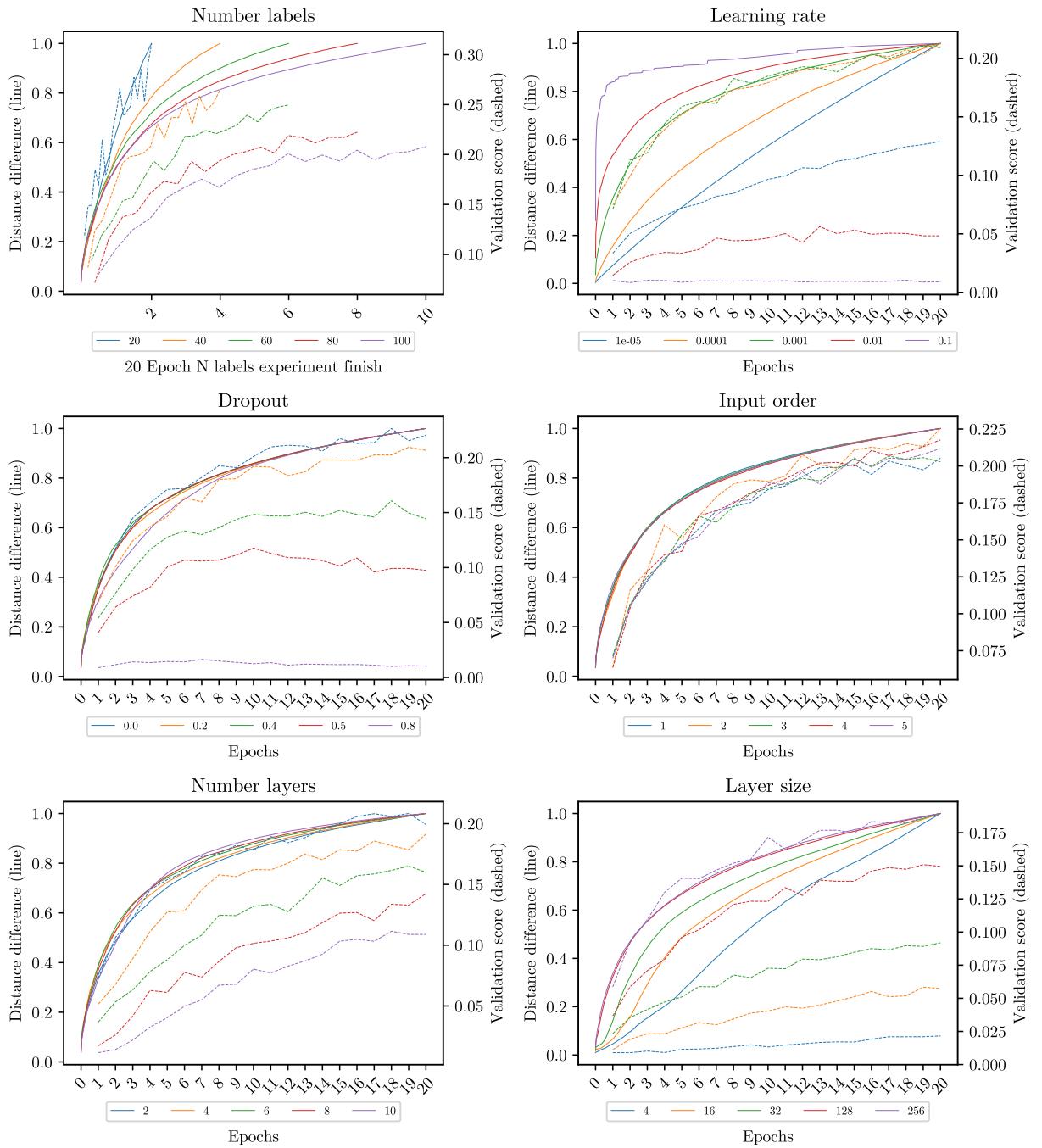
Figure 17: CIFAR-100 MLP cumulative using Silhouette discretization.

Cumulative plots without normalization
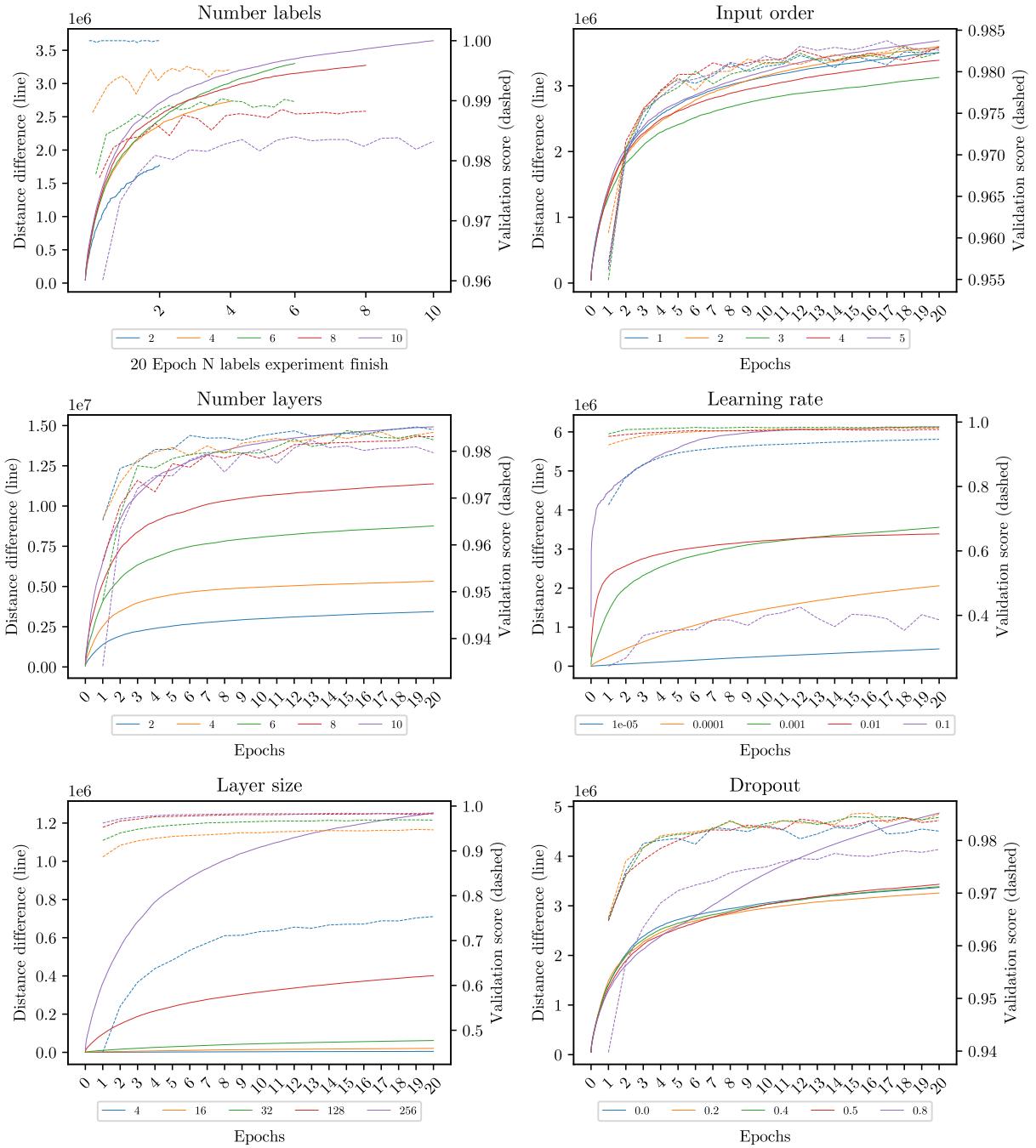
Figure 18: MNIST cumulative using Heat discretization.
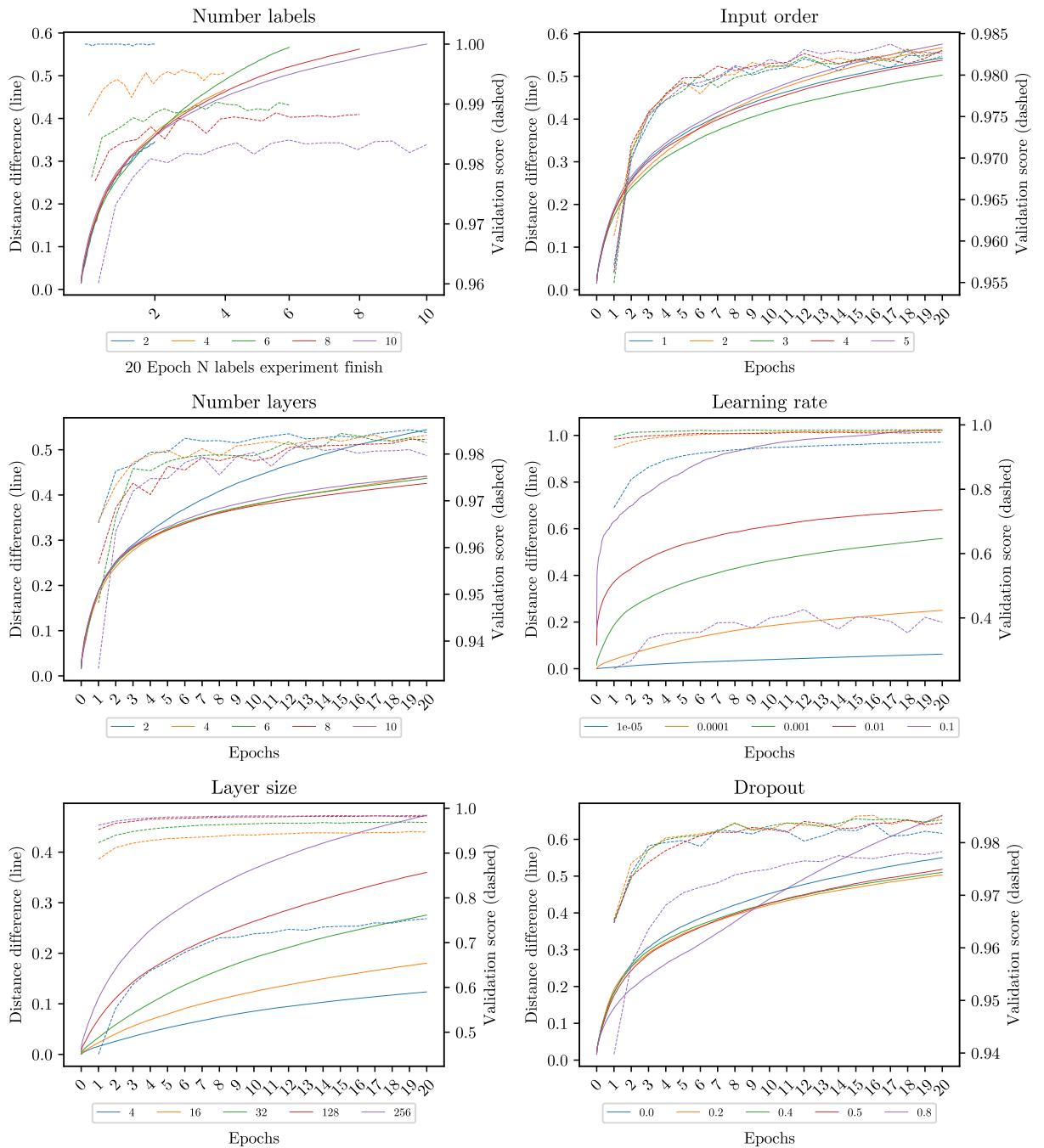
Figure 19: MNIST cumulative using Silhouette discretization.
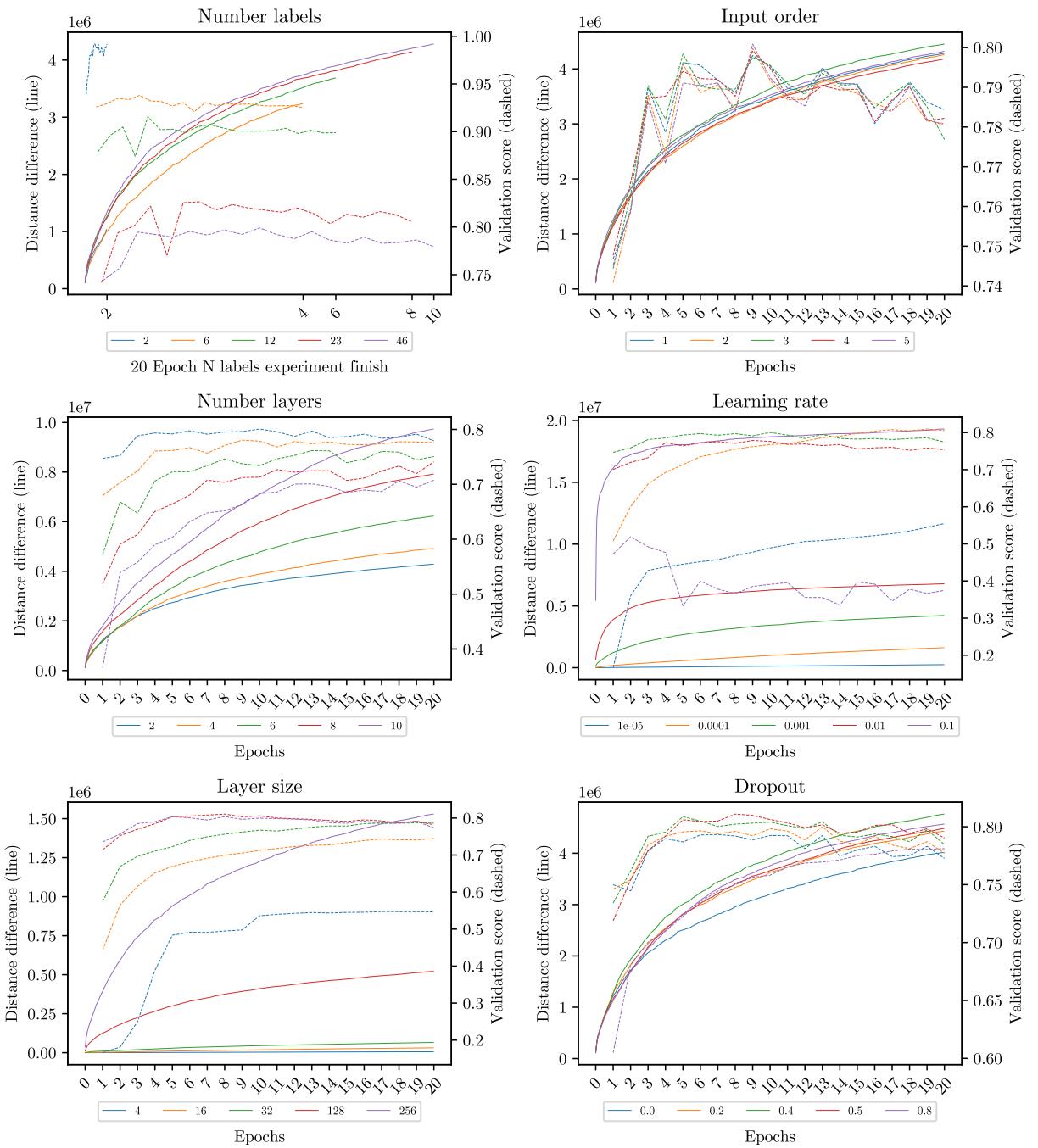
Figure 20: Reuters cumulative using Heat discretization.
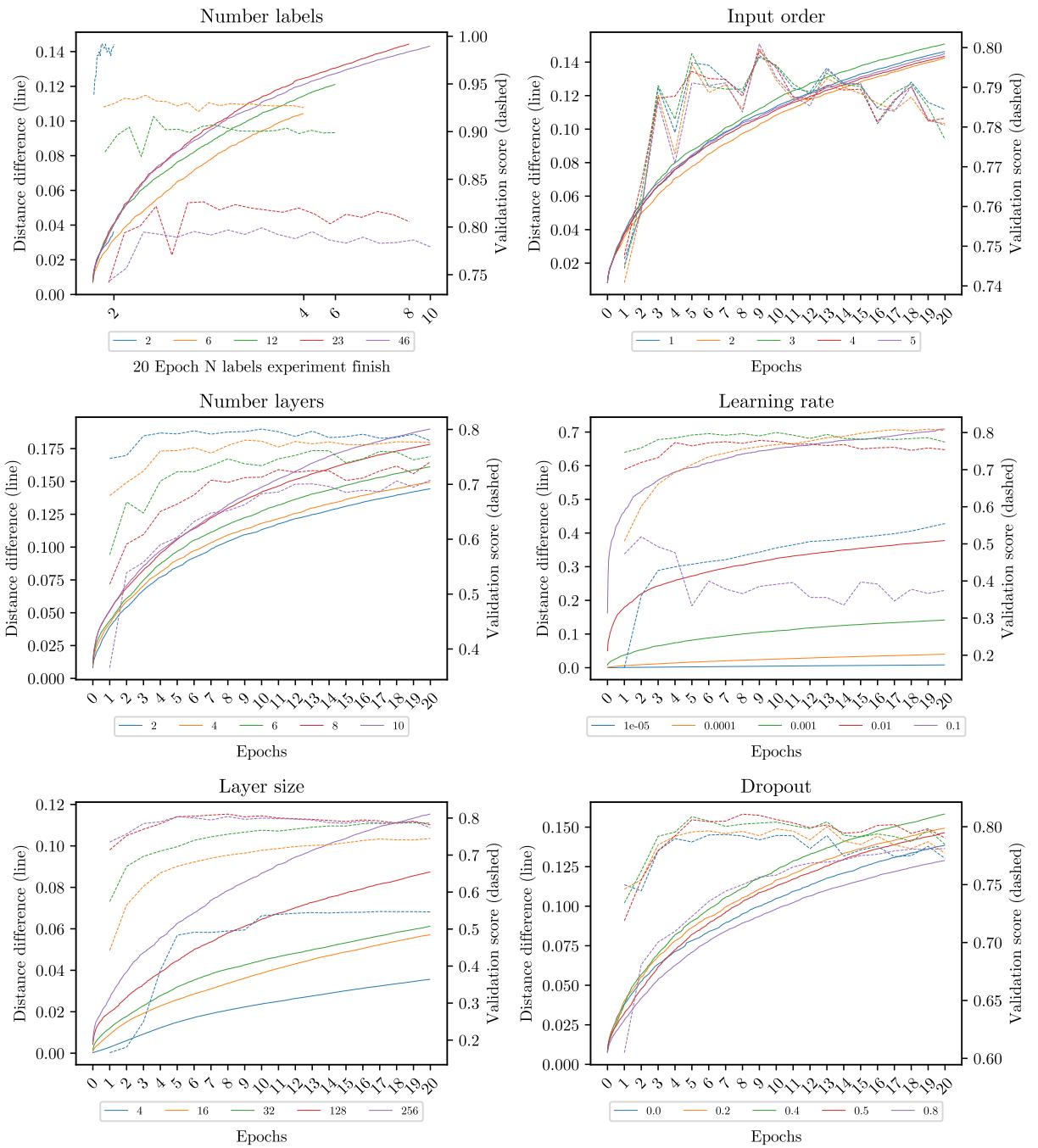
Figure 21: Reuters cumulative using Silhouette discretization.
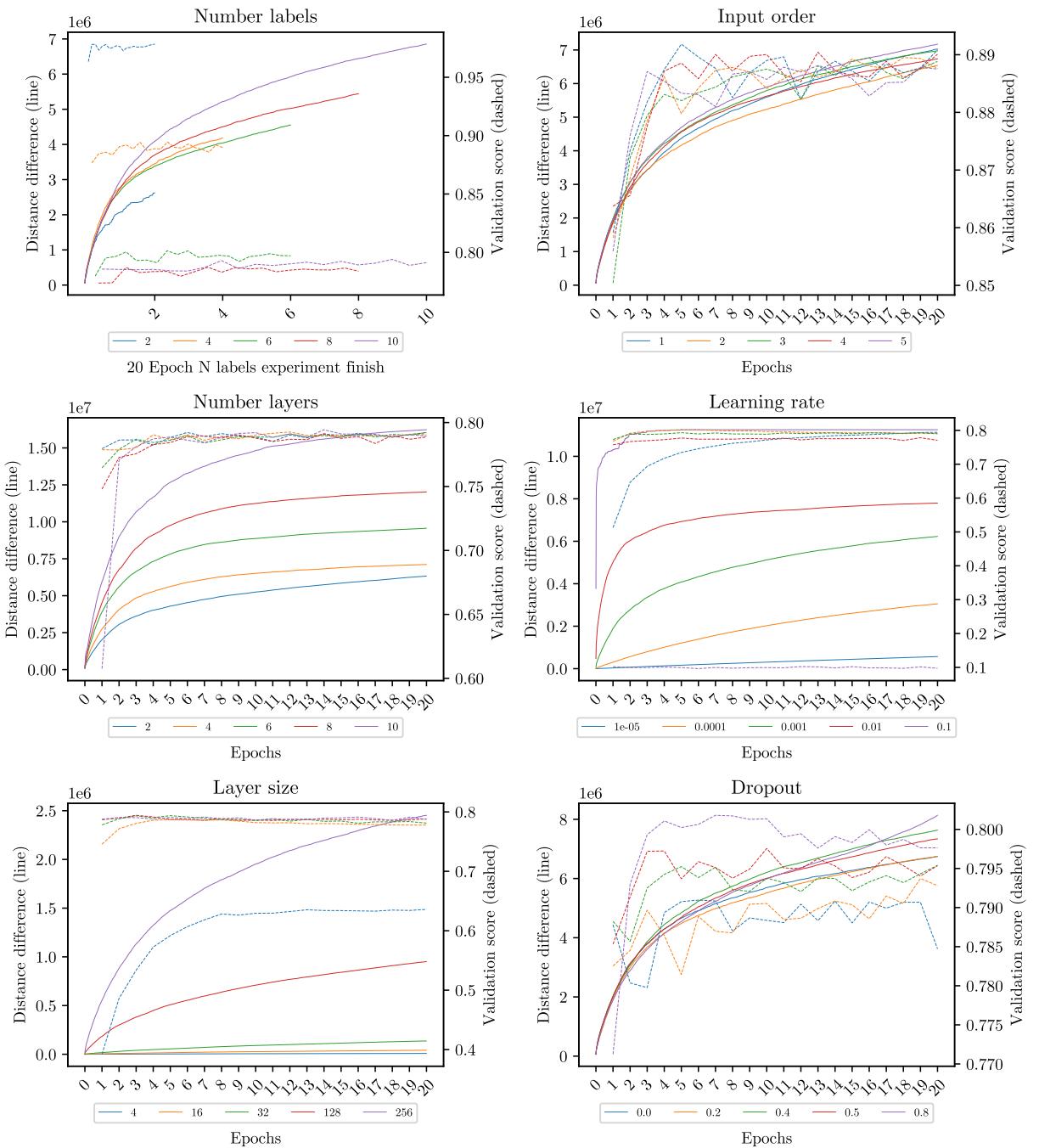
Figure 22: CIFAR-10 CNN cumulative no normalized using Heat discretization.
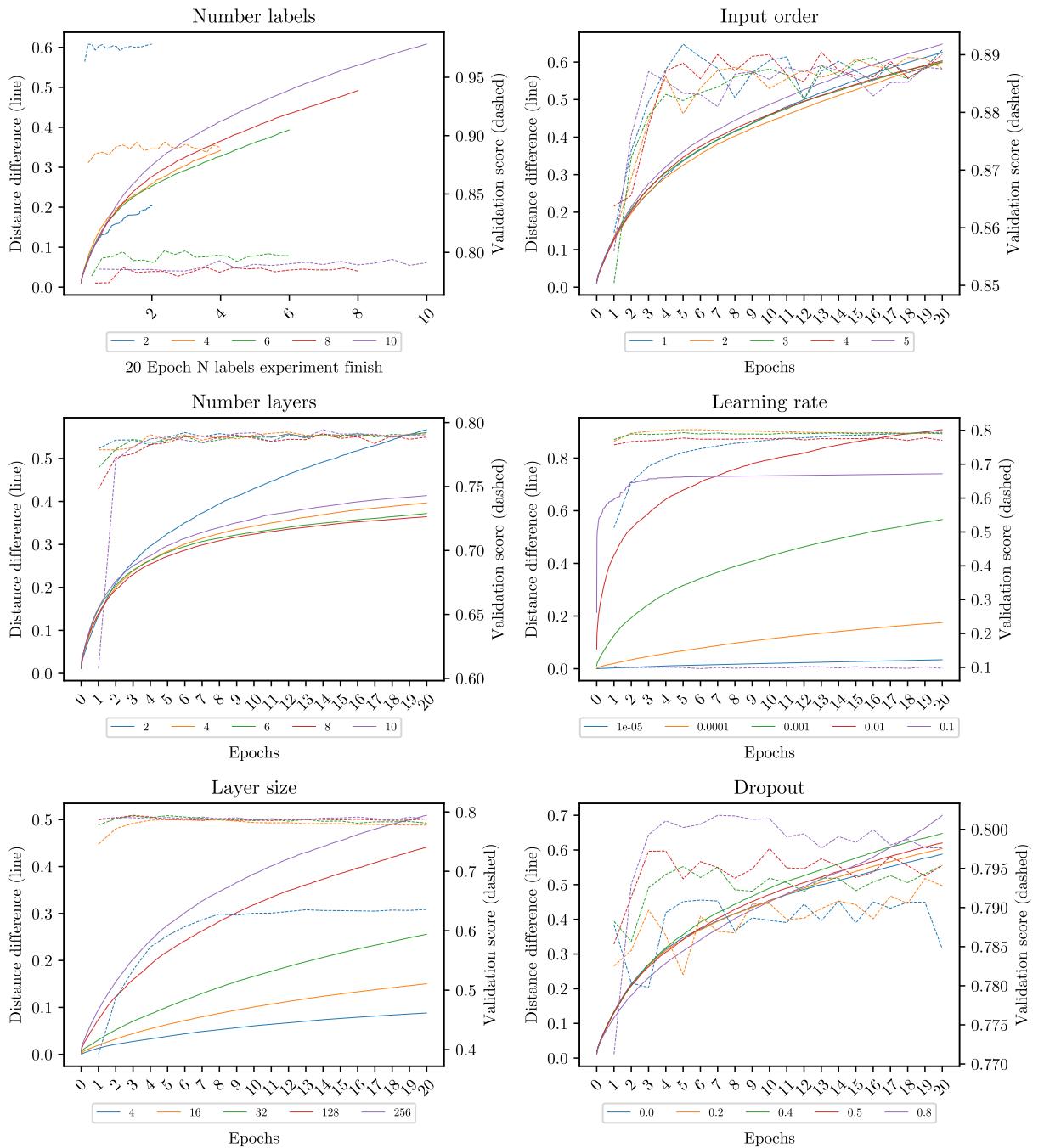
Figure 23: CIFAR-10 CNN cumulative no normalized using Silhouette discretization.
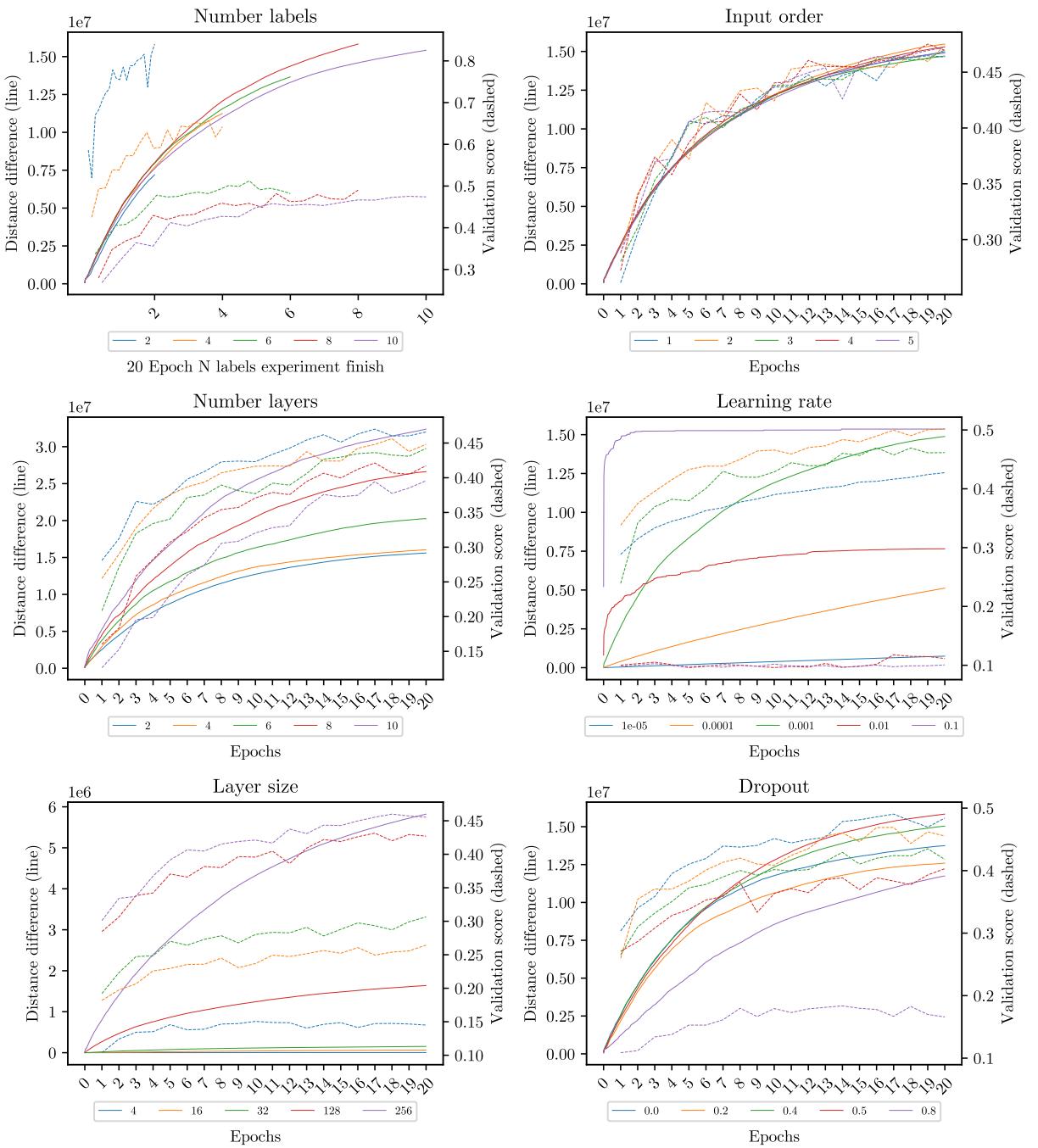
Figure 24: CIFAR-10 MLP cumulative no normalized using Heat discretization.
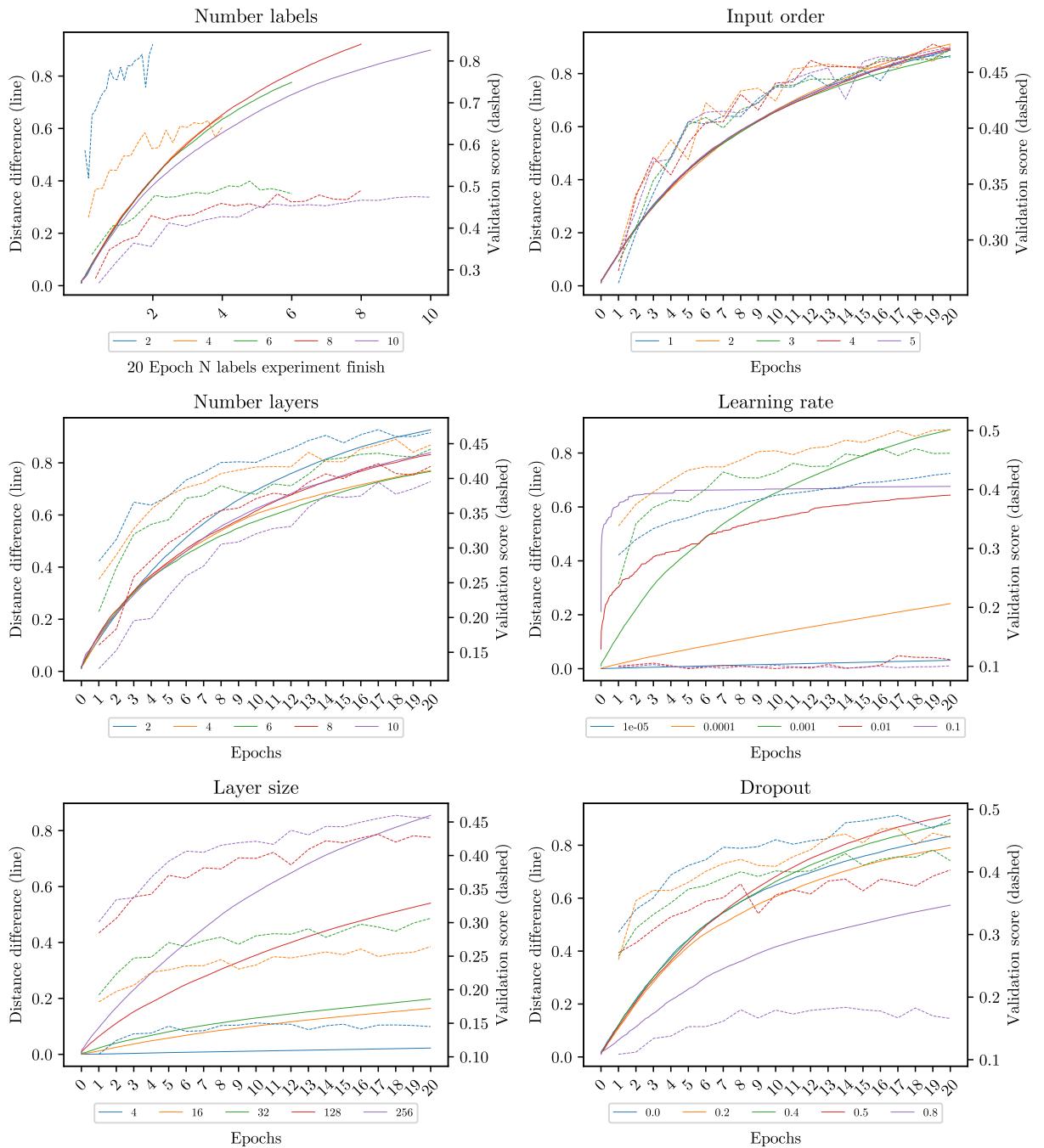
Figure 25: CIFAR-10 MLP cumulative no normalized using Silhouette discretization.
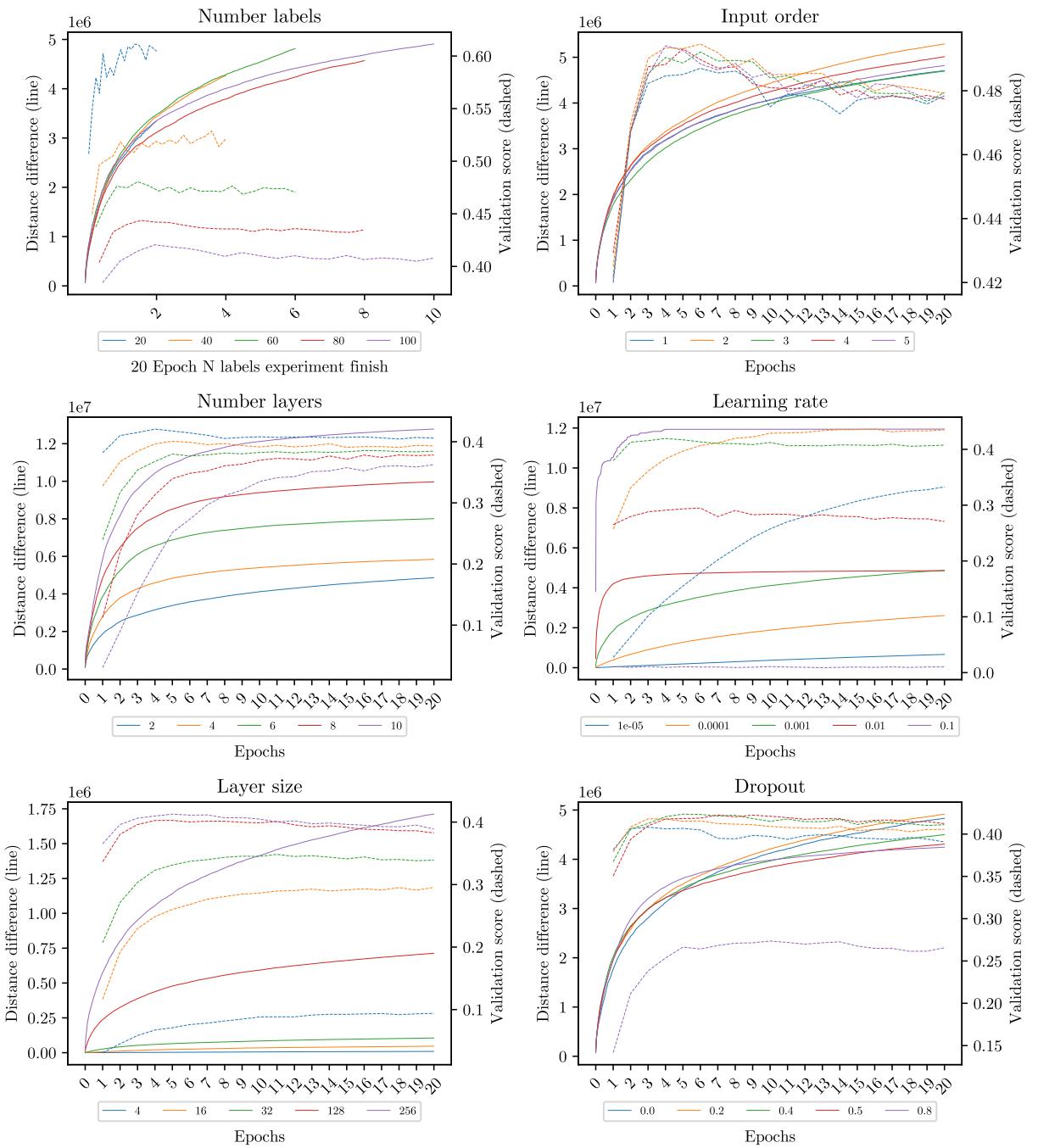
Figure 26: CIFAR-100 CNN cumulative no normalized using Heat discretization.
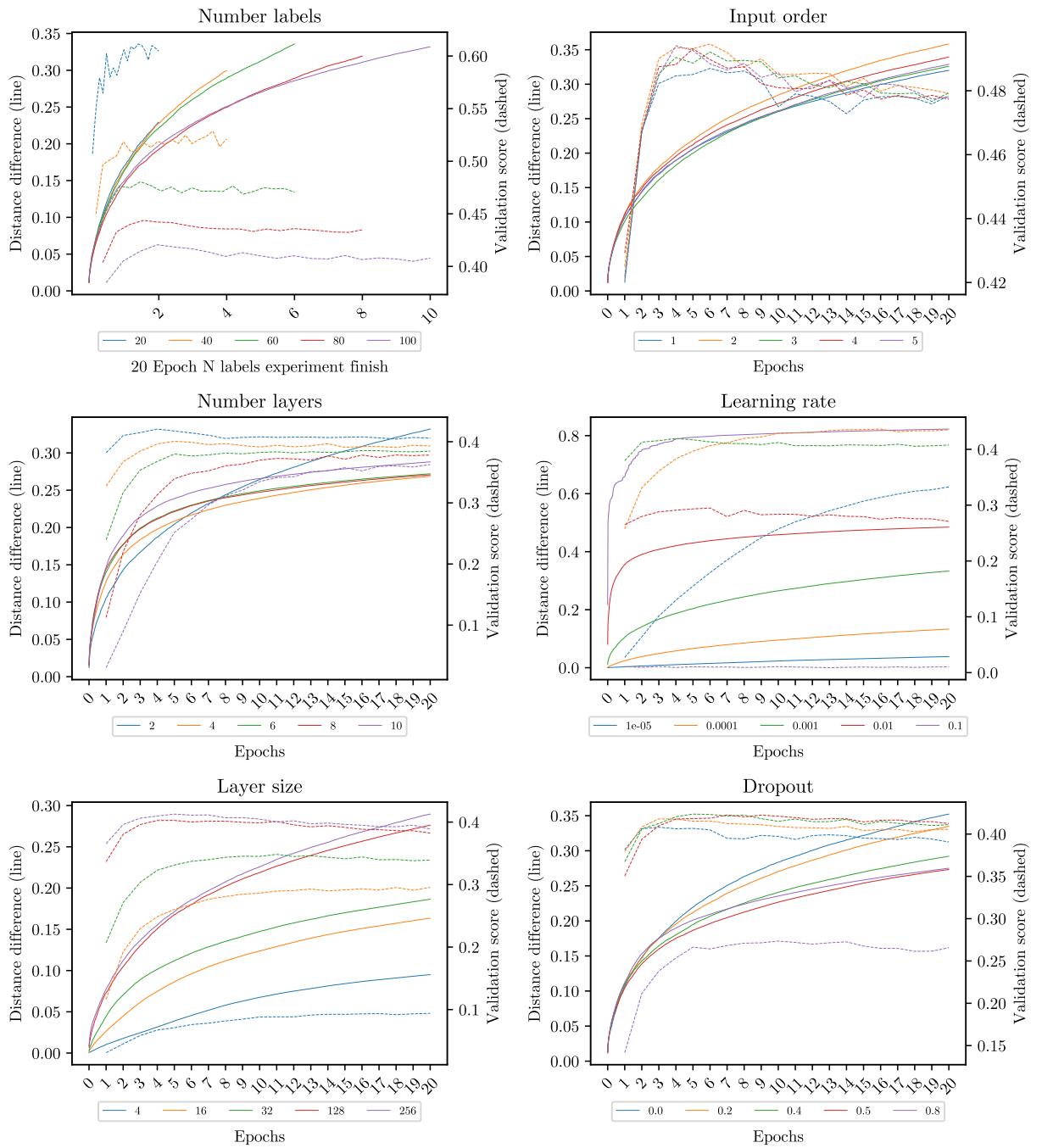
Figure 27: CIFAR-100 CNN cumulative no normalized using Silhouette discretization.
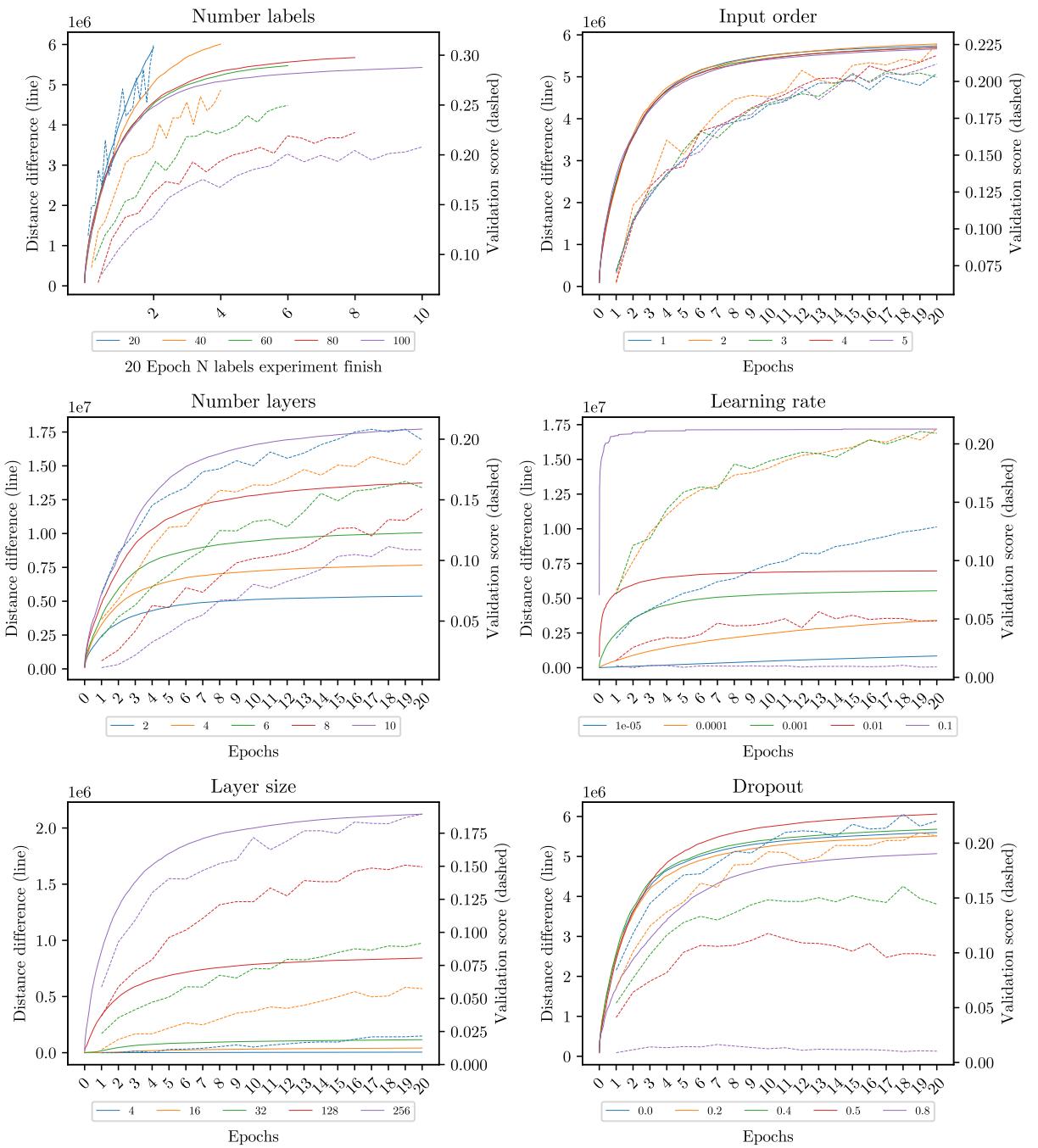
Figure 28: CIFAR-100 MLP cumulative no normalized using Heat discretization.
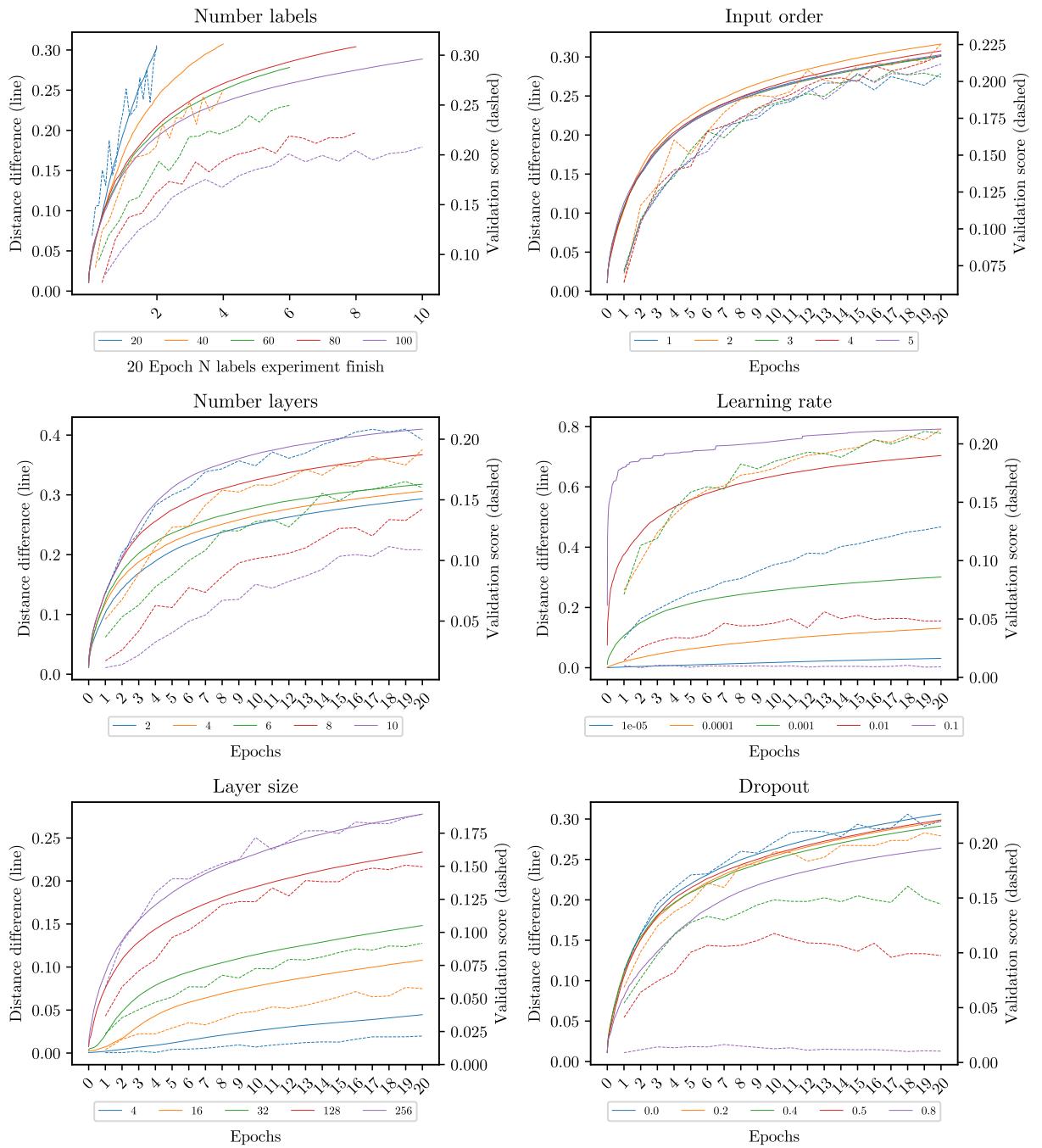
Figure 29: CIFAR-100 MLP cumulative no normalized using Silhouette discretization.
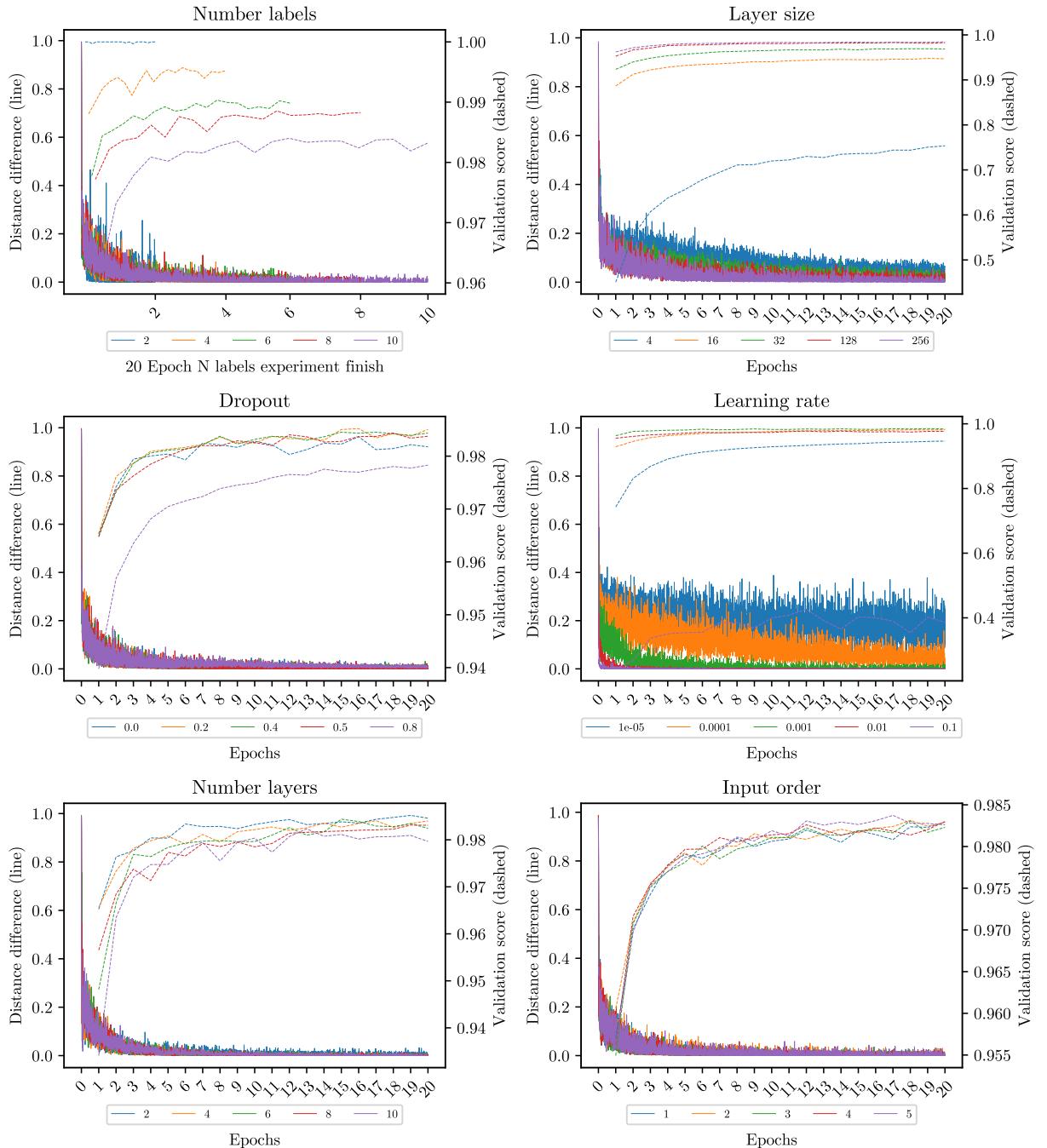
Figure 30: MNIST non-cumulative using Heat discretization.
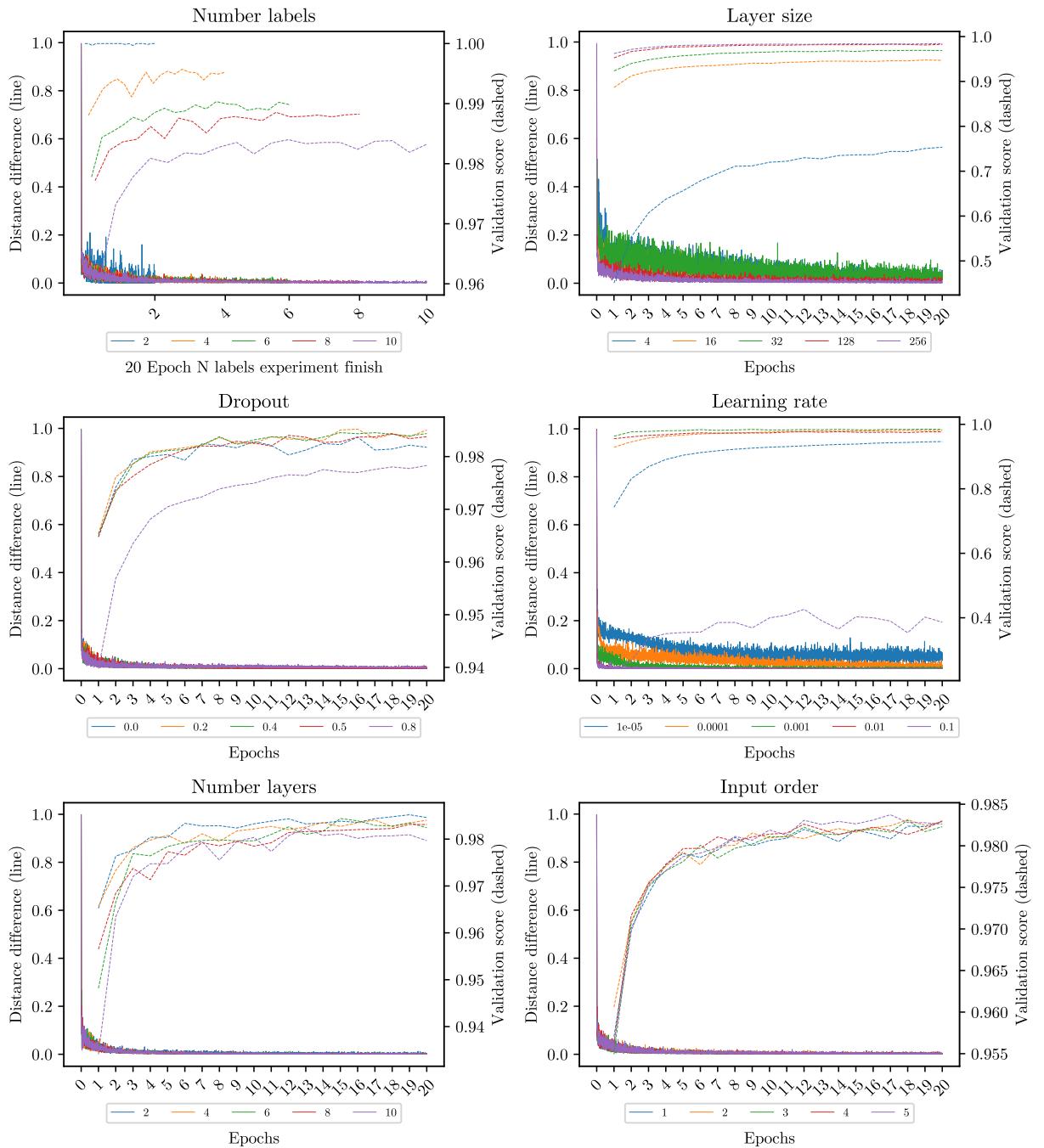
Figure 31: MNIST non-cumulative using Silhouette discretization.
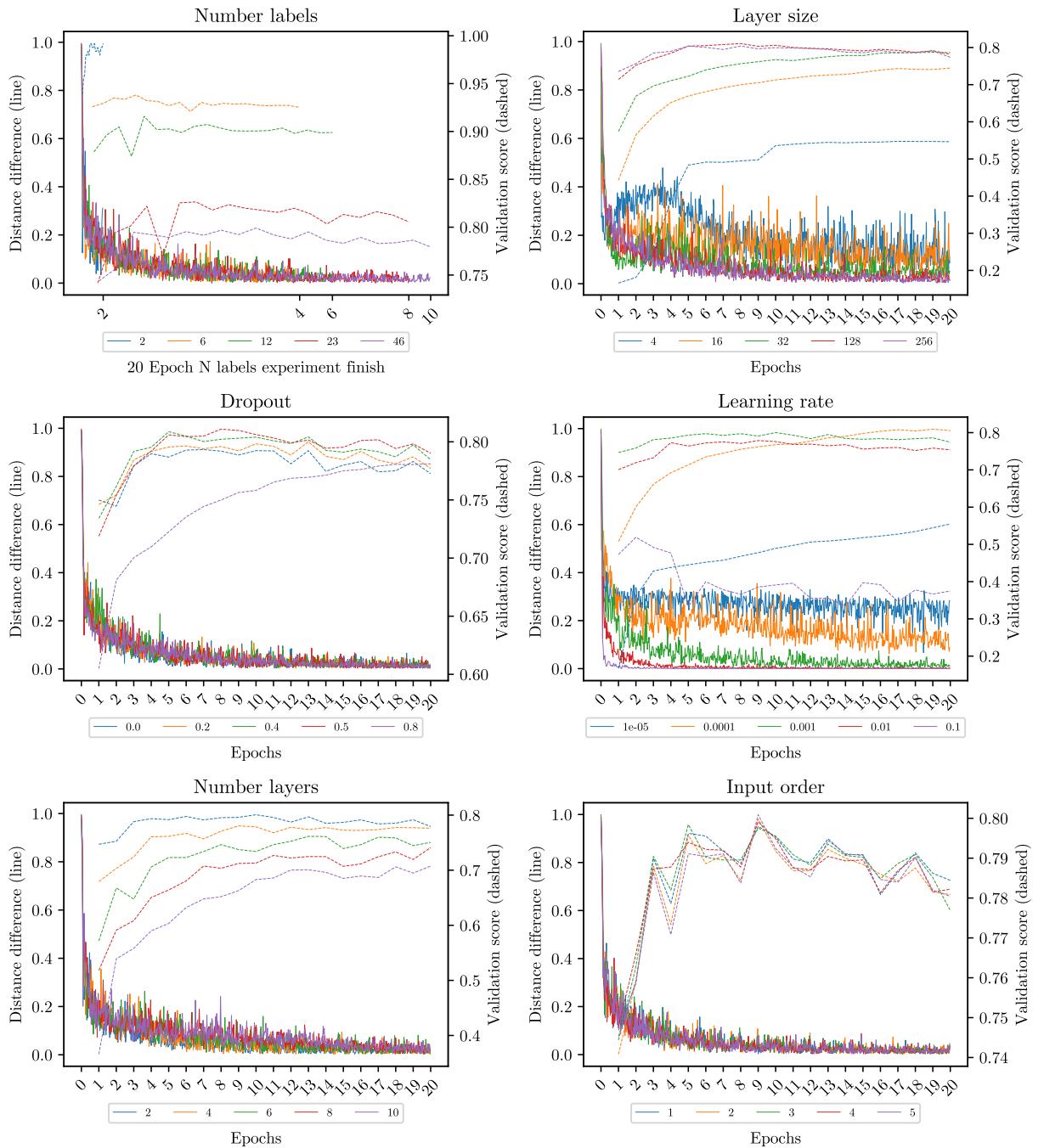
Figure 32: Reuters non-cumulative using Heat discretization.
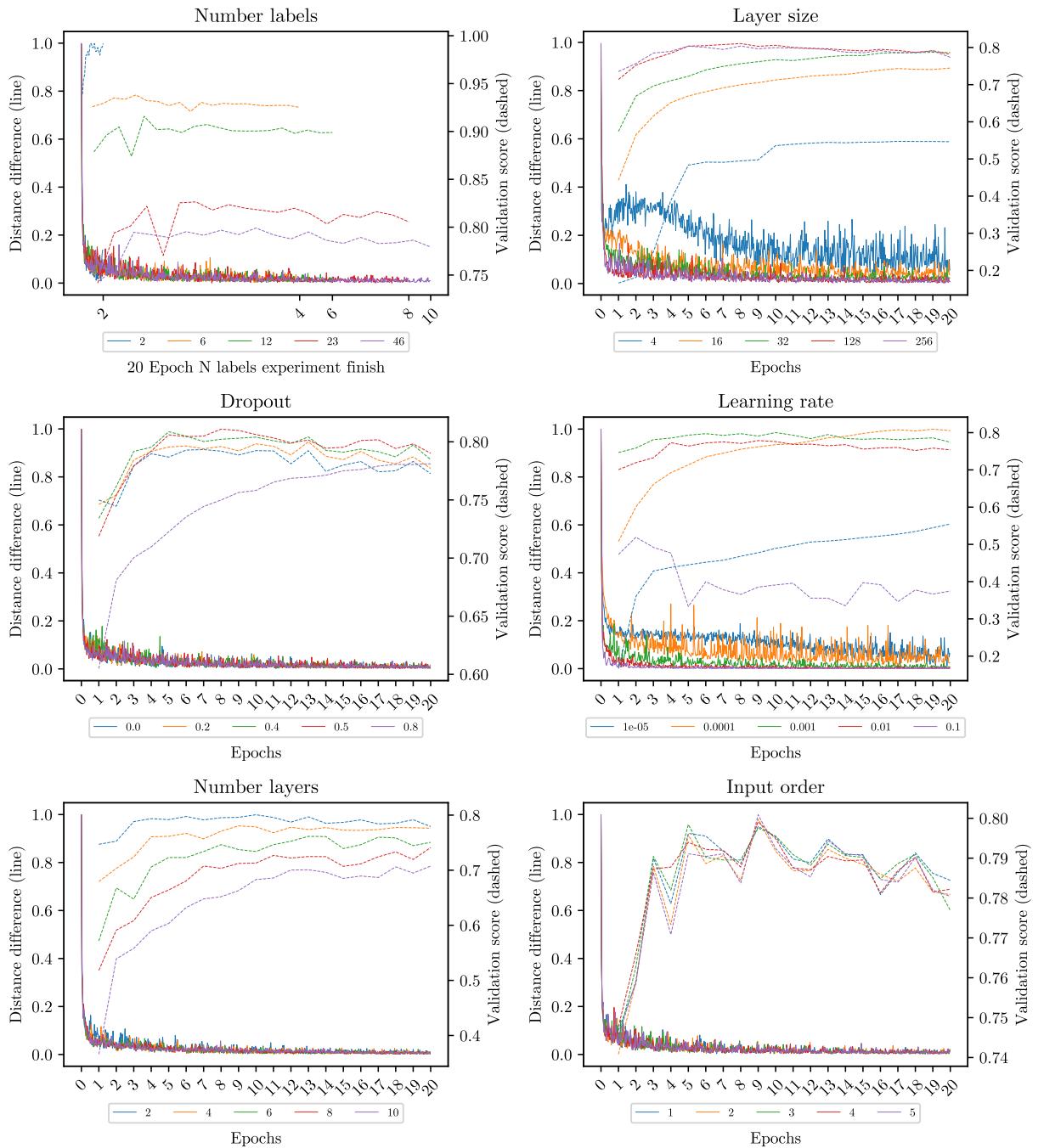
Figure 33: Reuters non-cumulative using Silhouette discretization.
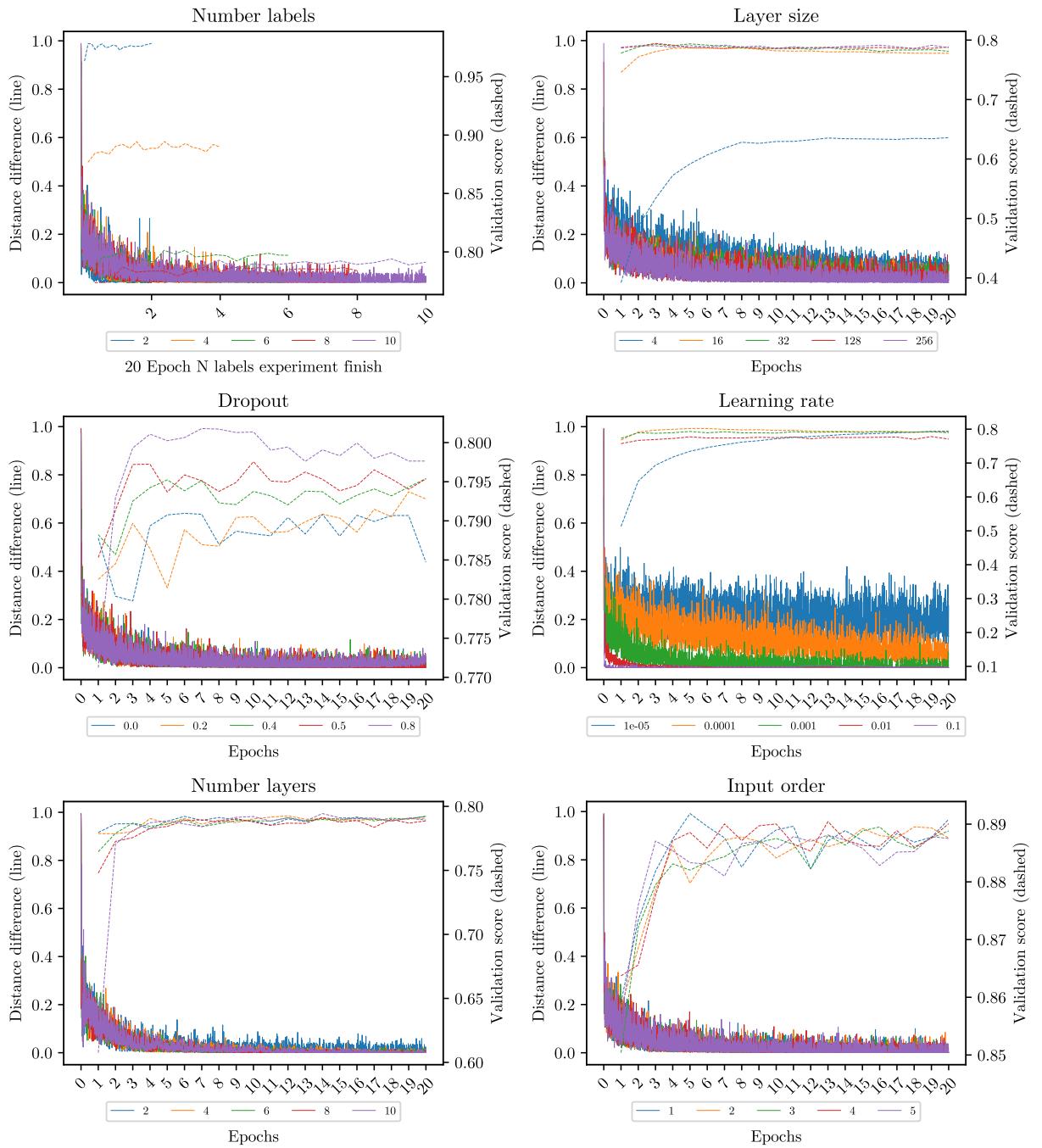
Figure 34: CIFAR-10 CNN non-cumulative using Heat discretization.
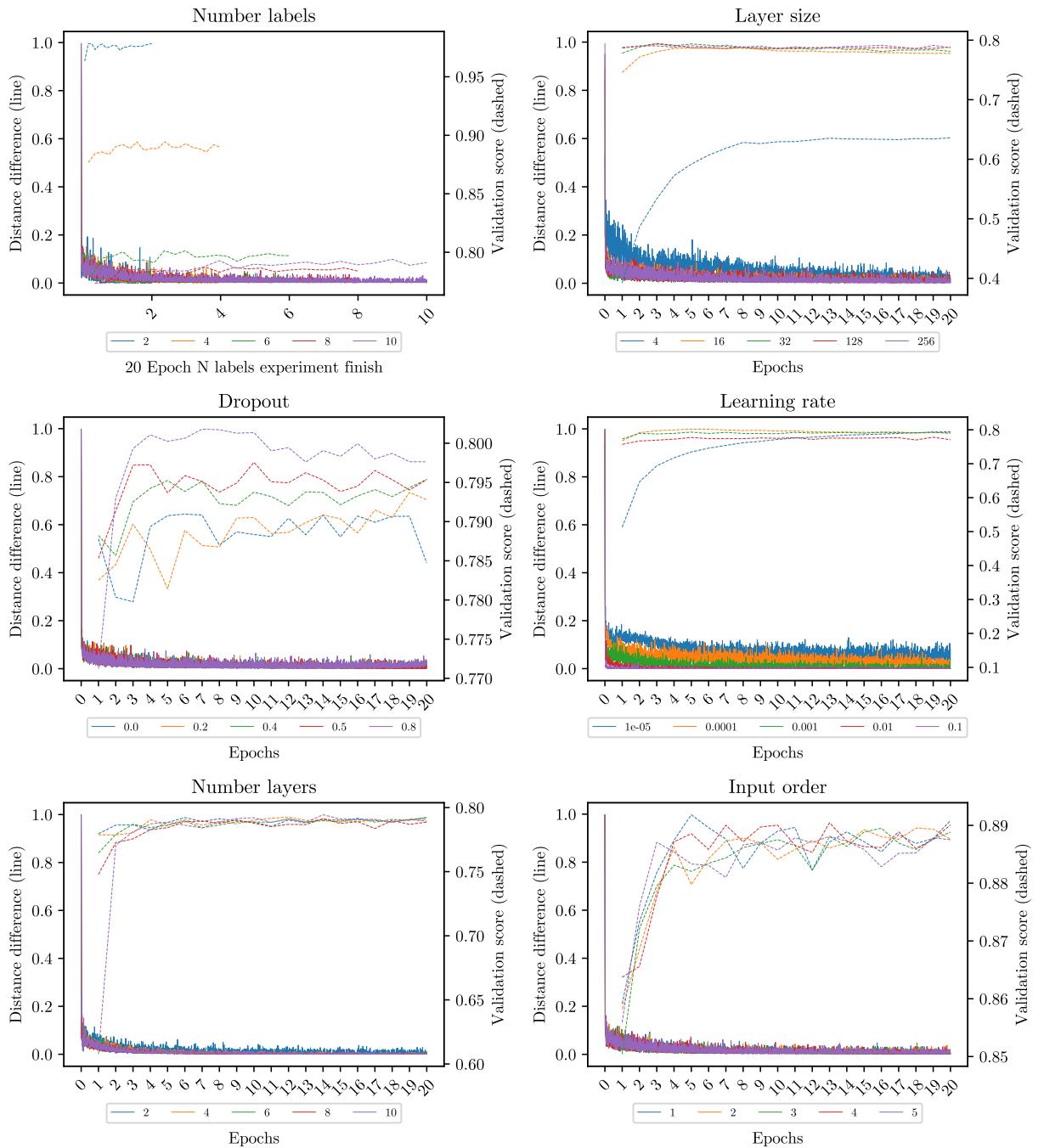
Figure 35: CIFAR-10 CNN non-cumulative using Silhouette discretization.
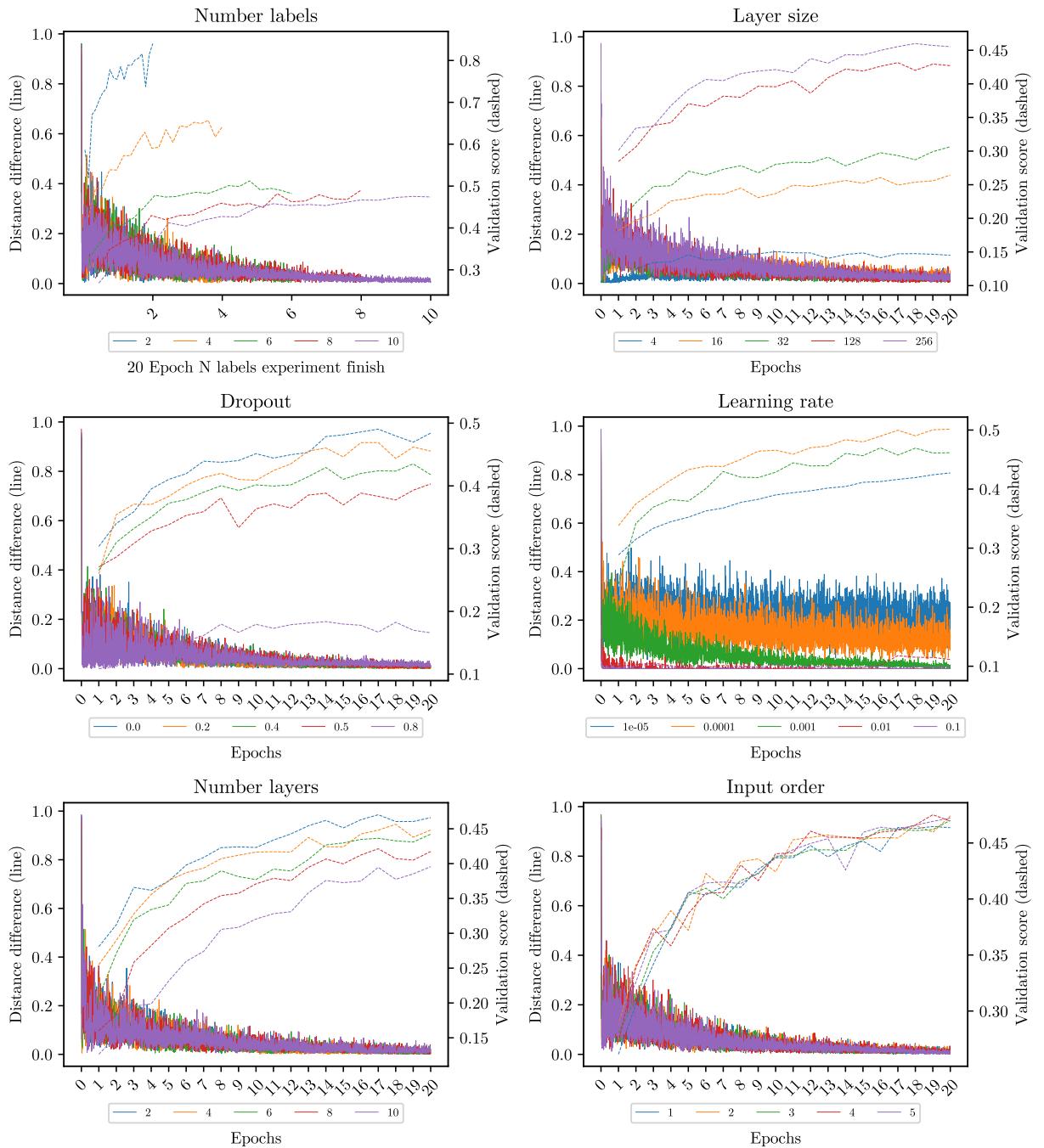
Figure 36: CIFAR-10 MLP non-cumulative using Heat discretization.
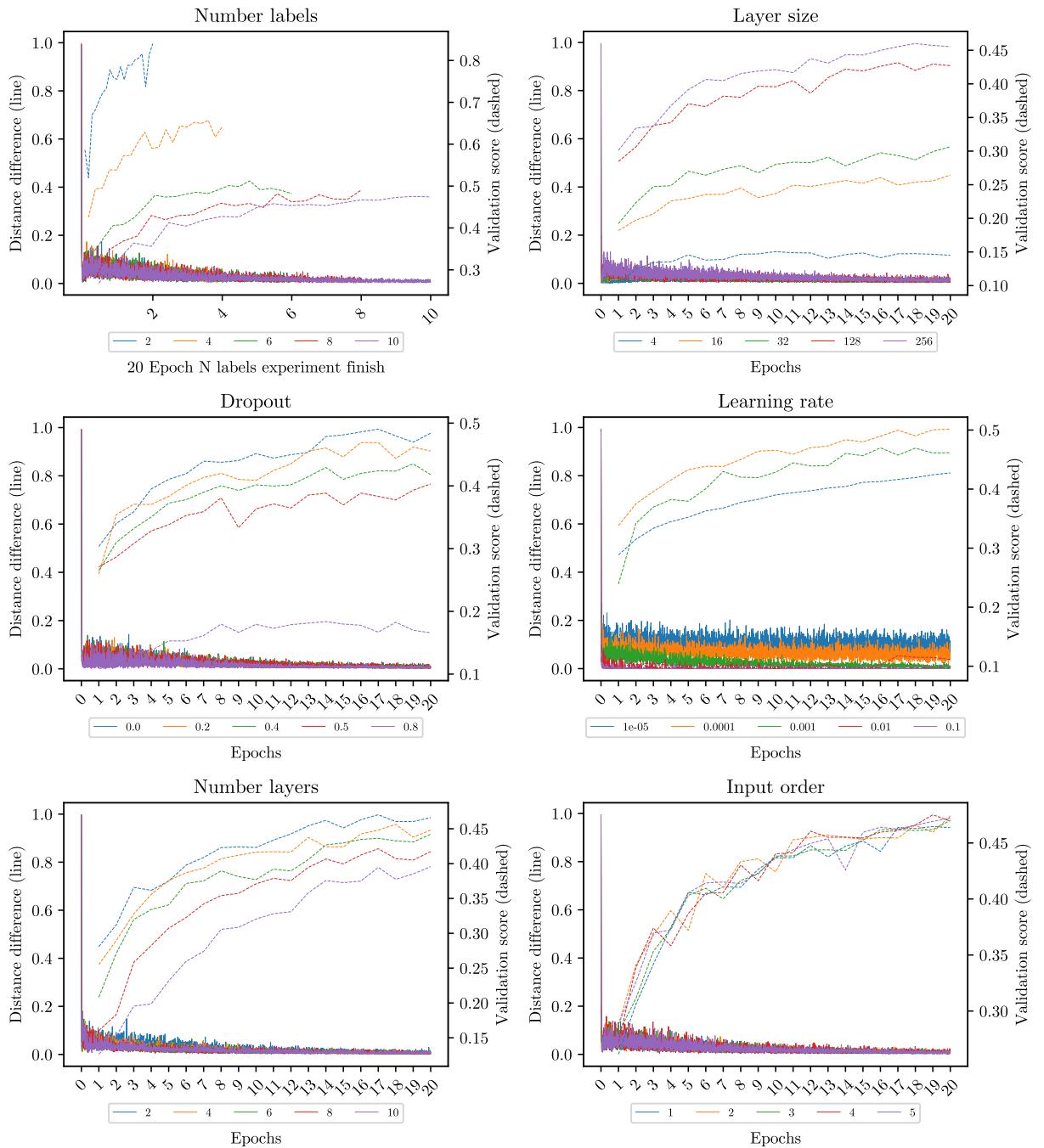
Figure 37: CIFAR-10 MLP non-cumulative using Silhouette discretization.
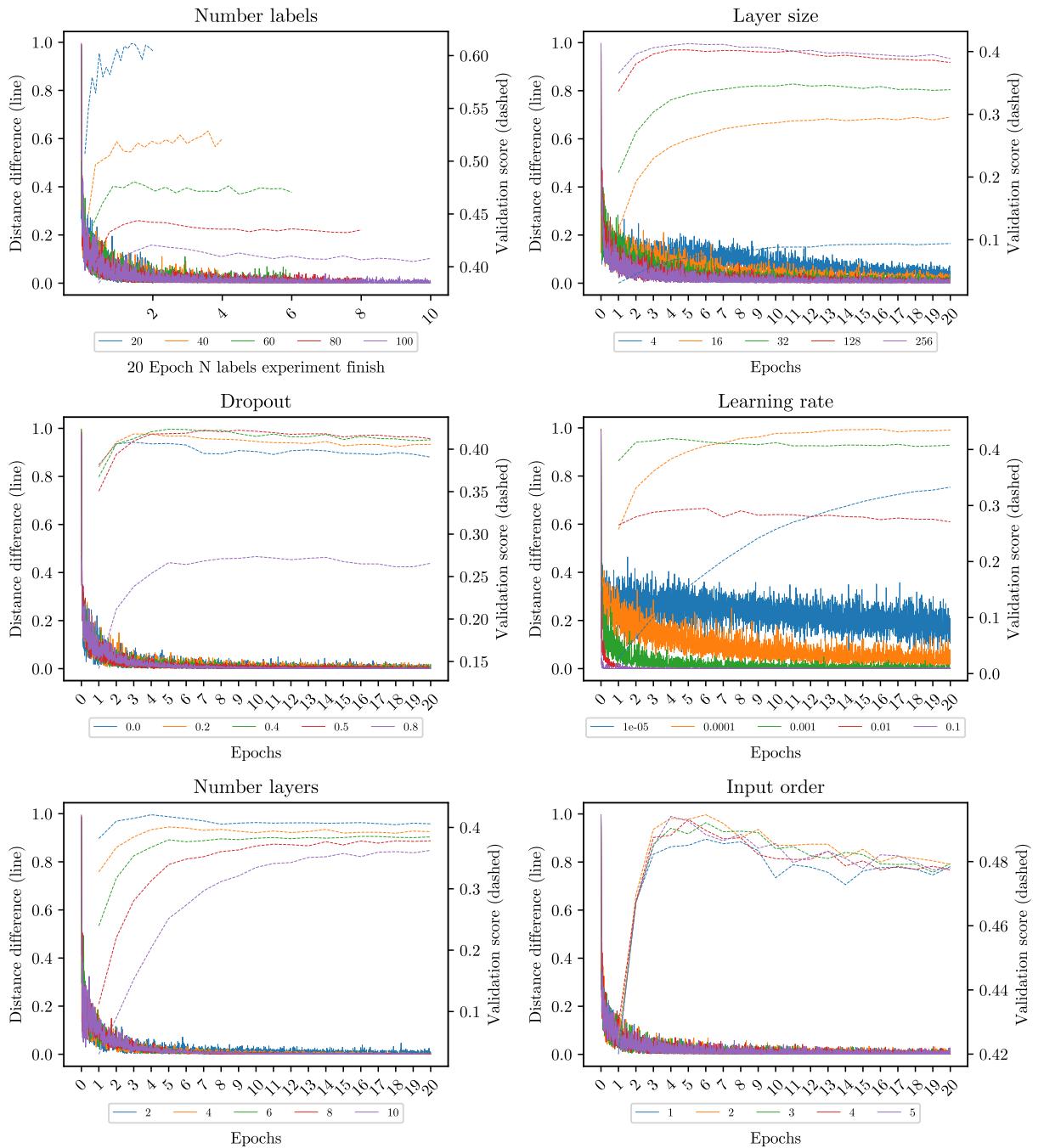
Figure 38: CIFAR-100 CNN non-cumulative using Heat discretization.
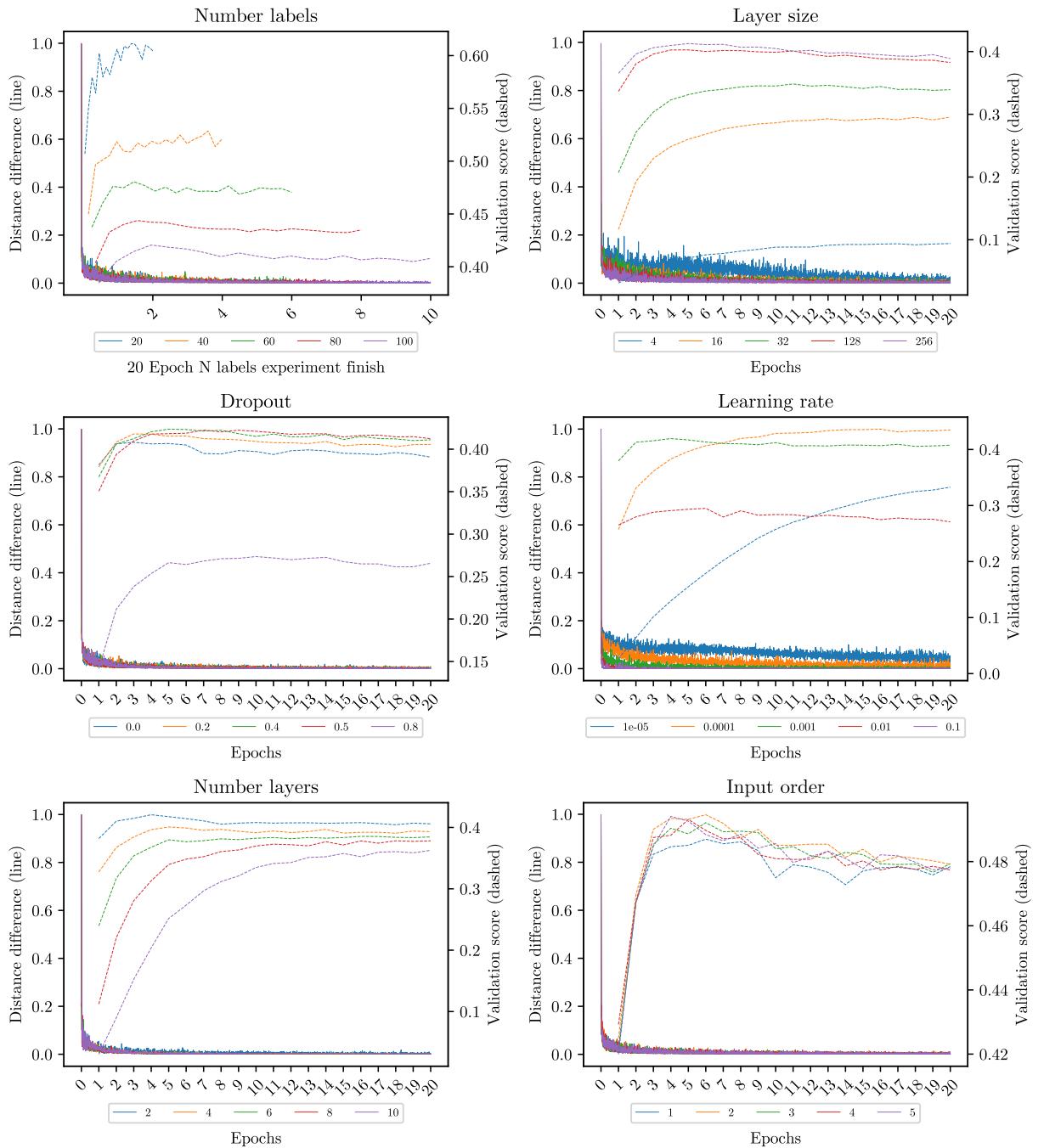
Figure 39: CIFAR-100 CNN non-cumulative using Silhouette discretization.
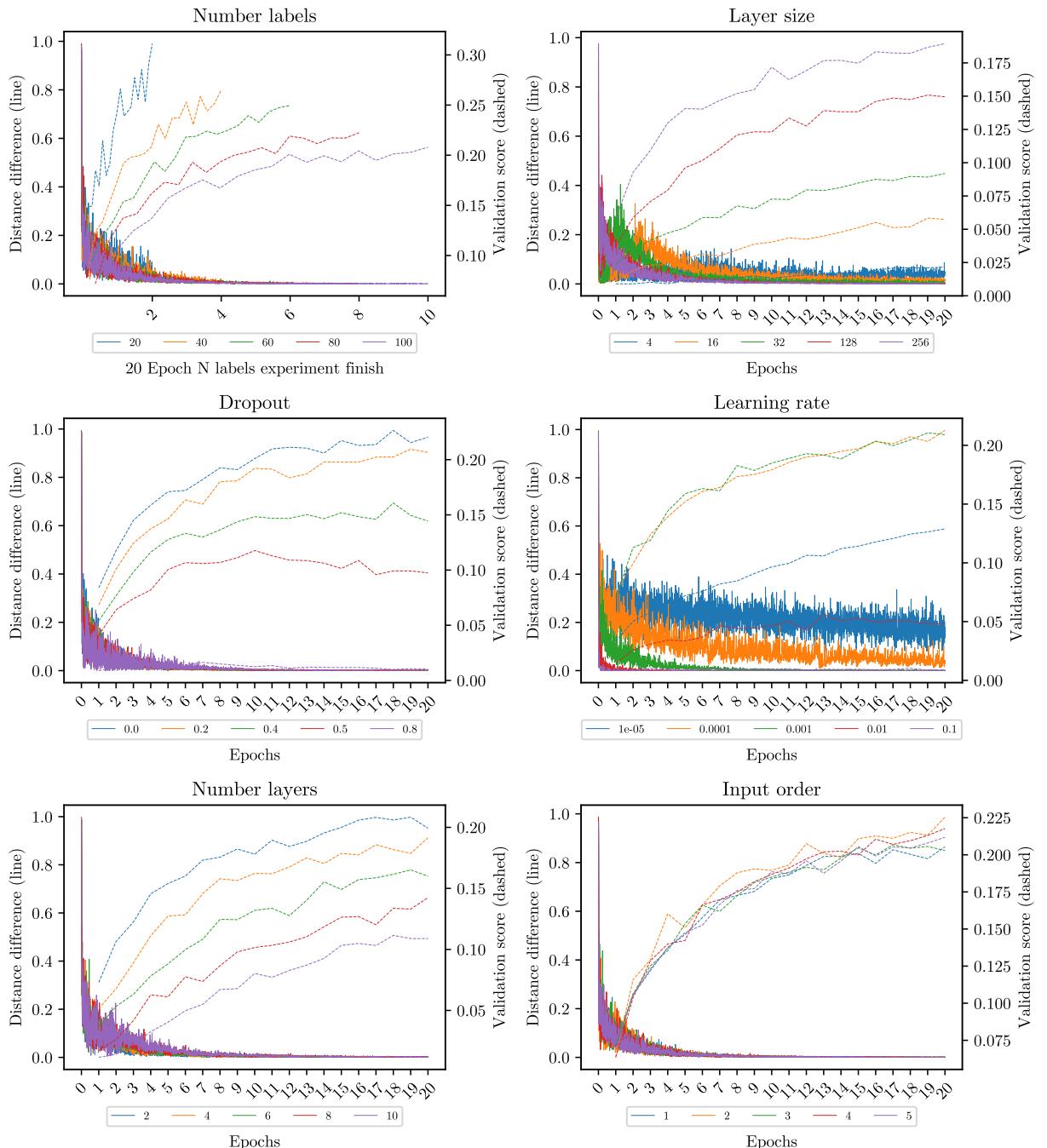
Figure 40: CIFAR-100 MLP non-cumulative using Heat discretization.
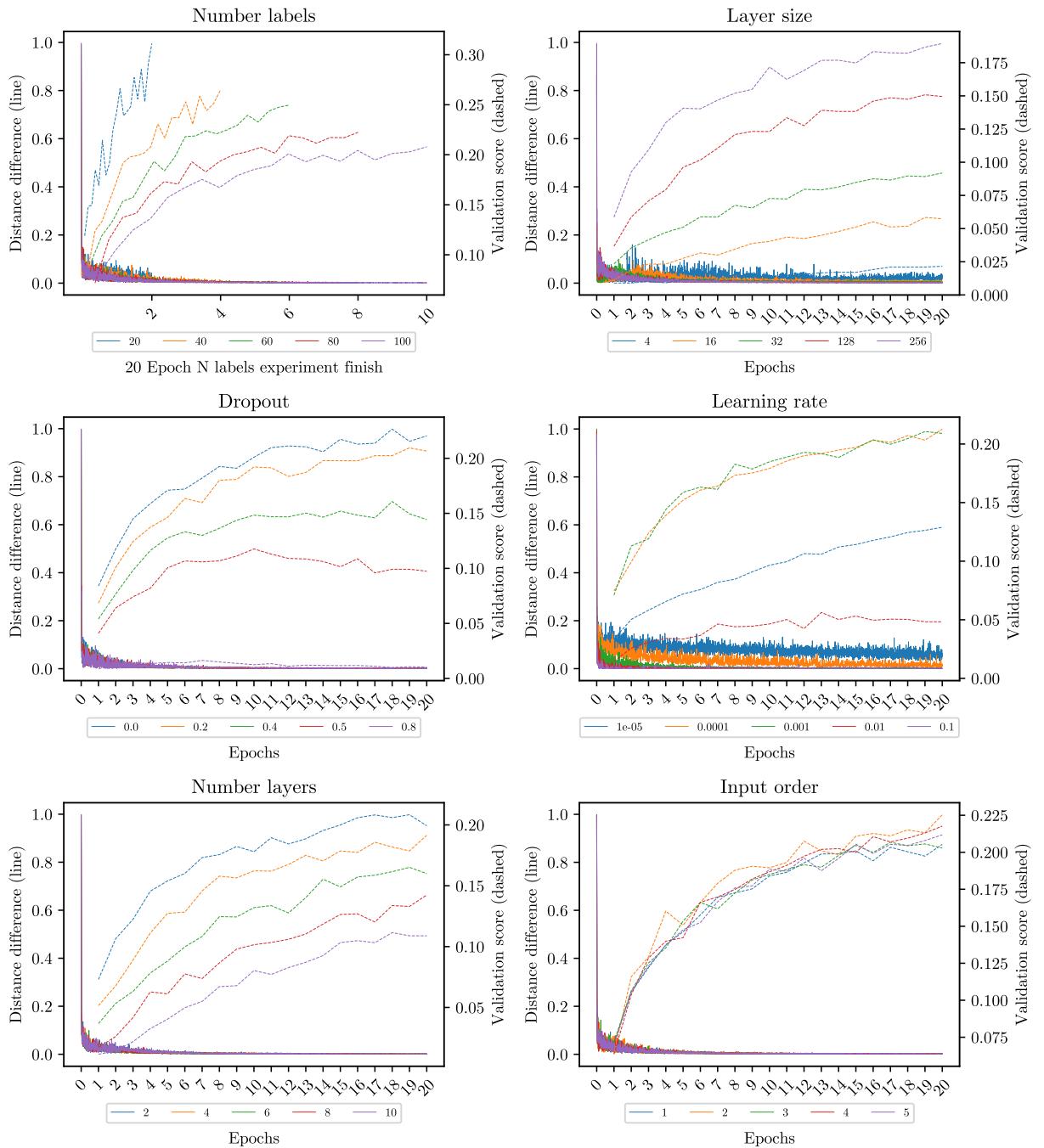
Figure 41: CIFAR-100 MLP non-cumulative using Silhouette discretization.

| Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8080	0.1270
Layer size	16	0.9120	0.0416
Layer size	32	0.9182	0.0327
Layer size	128	0.9224	0.0234
Layer size	256	0.9302	0.0328
Number labels	2	ERROR	ERROR
Number labels	4	0.6805	0.1218
Number labels	6	0.8525	0.0394
Number labels	8	0.8081	0.0835
Number labels	10	0.9197	0.0235
Learning rate	1e-05	0.8035	0.0158
Learning rate	0.0001	0.9165	0.0136
Learning rate	0.001	0.9106	0.0283
Learning rate	0.01	0.8258	0.0615
Learning rate	0.1	0.6516	0.2004
Dropout	0.0	0.8855	0.0251
Dropout	0.2	0.9399	0.0344
Dropout	0.4	0.9567	0.0097
Dropout	0.5	0.9487	0.0057
Dropout	0.8	0.9085	0.0494
Input order	1	0.9416	0.0212
Input order	2	0.9376	0.0233
Input order	3	0.9461	0.0290
Input order	4	0.9401	0.0190
Input order	5	0.9517	0.0124
Number layers	2	0.9244	0.0307
Number layers	4	0.8943	0.0431
Number layers	6	0.9428	0.0131
Number layers	8	0.9322	0.0360
Number layers	10	0.9290	0.0313

Table 1: MNIST Pearson's correlation using Heat distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8232	0.1269
Layer size	16	0.8992	0.0428
Layer size	32	0.8976	0.0403
Layer size	128	0.8866	0.0246
Layer size	256	0.8882	0.0417
Number labels	2	ERROR	ERROR
Number labels	4	0.6691	0.1120
Number labels	6	0.8039	0.0325
Number labels	8	0.7982	0.0716
Number labels	10	0.8668	0.0179
Learning rate	1e-05	0.8542	0.0125
Learning rate	0.0001	0.9309	0.0117
Learning rate	0.001	0.8701	0.0496
Learning rate	0.01	0.8666	0.0283
Learning rate	0.1	0.6576	0.2068
Dropout	0.0	0.8172	0.0247
Dropout	0.2	0.9073	0.0483
Dropout	0.4	0.9274	0.0160
Dropout	0.5	0.9237	0.0148
Dropout	0.8	0.8574	0.0575
Input order	1	0.9039	0.0268
Input order	2	0.9000	0.0320
Input order	3	0.9018	0.0492
Input order	4	0.8907	0.0262
Input order	5	0.9159	0.0209
Number layers	2	0.8903	0.0445
Number layers	4	0.8872	0.0303
Number layers	6	0.9236	0.0169
Number layers	8	0.9384	0.0318
Number layers	10	0.9057	0.0348

Table 2: MNIST Pearson's correlation using Silhouette distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8322	0.0319
Layer size	16	0.8581	0.0788
Layer size	32	0.9122	0.0252
Layer size	128	0.6472	0.0350
Layer size	256	0.5612	0.0534
Number labels	2	0.7383	0.0774
Number labels	6	-0.2842	0.2590
Number labels	12	0.3378	0.0808
Number labels	23	0.5743	0.0610
Number labels	46	0.5203	0.0883
Learning rate	1e-05	0.7949	0.1165
Learning rate	0.0001	0.8973	0.0077
Learning rate	0.001	0.5292	0.1114
Learning rate	0.01	0.7052	0.0696
Learning rate	0.1	-0.4898	0.1535
Dropout	0.0	0.4508	0.0532
Dropout	0.2	0.5700	0.0740
Dropout	0.4	0.6080	0.0378
Dropout	0.5	0.7262	0.0242
Dropout	0.8	0.9752	0.0076
Input order	1	0.6251	0.1389
Input order	2	0.6323	0.0630
Input order	3	0.6109	0.1181
Input order	4	0.5771	0.1141
Input order	5	0.6703	0.0275
Number layers	2	0.5752	0.0677
Number layers	4	0.8567	0.0320
Number layers	6	0.8669	0.0235
Number layers	8	0.8962	0.0155
Number layers	10	0.8843	0.0548

Table 3: REUTERS Pearson's correlation using Heat distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8443	0.0372
Layer size	16	0.9011	0.0668
Layer size	32	0.9353	0.0222
Layer size	128	0.5695	0.0463
Layer size	256	0.4760	0.0581
Number labels	2	0.7434	0.0785
Number labels	6	-0.3016	0.2553
Number labels	12	0.3083	0.0873
Number labels	23	0.5370	0.0666
Number labels	46	0.4552	0.0830
Learning rate	1e-05	0.8071	0.1153
Learning rate	0.0001	0.9187	0.0073
Learning rate	0.001	0.4635	0.1232
Learning rate	0.01	0.6082	0.0751
Learning rate	0.1	-0.5837	0.1366
Dropout	0.0	0.3854	0.0561
Dropout	0.2	0.5108	0.0767
Dropout	0.4	0.5380	0.0508
Dropout	0.5	0.6617	0.0143
Dropout	0.8	0.9554	0.0112
Input order	1	0.5739	0.1468
Input order	2	0.5711	0.0738
Input order	3	0.5591	0.1282
Input order	4	0.5220	0.1145
Input order	5	0.6216	0.0352
Number layers	2	0.5207	0.0712
Number layers	4	0.8503	0.0348
Number layers	6	0.8823	0.0234
Number layers	8	0.9220	0.0110
Number layers	10	0.9068	0.0409

Table 4: REUTERS Pearson's correlation using Silhouette distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8066	0.0480
Layer size	16	0.2882	0.3050
Layer size	32	-0.3713	0.2989
Layer size	128	-0.2872	0.1714
Layer size	256	-0.1035	0.2905
Number labels	2	0.2517	0.2946
Number labels	4	0.3307	0.3197
Number labels	6	0.2712	0.3099
Number labels	8	0.5009	0.2813
Number labels	10	0.3501	0.1339
Learning rate	1e-05	0.7876	0.0099
Learning rate	0.0001	0.1131	0.0572
Learning rate	0.001	0.5346	0.2875
Learning rate	0.01	0.4672	0.1593
Learning rate	0.1	-0.1287	0.3295
Dropout	0.0	0.2901	0.2357
Dropout	0.2	0.5089	0.2655
Dropout	0.4	0.3915	0.2252
Dropout	0.5	0.3815	0.2564
Dropout	0.8	0.5608	0.1693
Input order	1	0.6134	0.1403
Input order	2	0.7551	0.1070
Input order	3	0.7949	0.0424
Input order	4	0.7134	0.1470
Input order	5	0.6552	0.1582
Number layers	2	0.3803	0.2913
Number layers	4	0.5976	0.0911
Number layers	6	0.6576	0.1912
Number layers	8	0.8081	0.0917
Number layers	10	0.8050	0.0364

Table 5: CIFAR10CNN Pearson's correlation using Heat distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8365	0.0410
Layer size	16	0.2859	0.3022
Layer size	32	-0.3900	0.2913
Layer size	128	-0.2940	0.1659
Layer size	256	-0.1073	0.2823
Number labels	2	0.2326	0.2877
Number labels	4	0.2942	0.3251
Number labels	6	0.2448	0.2918
Number labels	8	0.4721	0.2600
Number labels	10	0.3736	0.0926
Learning rate	1e-05	0.8300	0.0094
Learning rate	0.0001	0.1090	0.0606
Learning rate	0.001	0.4838	0.2909
Learning rate	0.01	0.4343	0.1867
Learning rate	0.1	-0.1287	0.3343
Dropout	0.0	0.2839	0.2095
Dropout	0.2	0.5217	0.2303
Dropout	0.4	0.3686	0.2129
Dropout	0.5	0.3344	0.2336
Dropout	0.8	0.4527	0.1747
Input order	1	0.5628	0.1471
Input order	2	0.7108	0.1276
Input order	3	0.7470	0.0569
Input order	4	0.6606	0.1529
Input order	5	0.5970	0.1521
Number layers	2	0.3759	0.2634
Number layers	4	0.5952	0.0978
Number layers	6	0.6525	0.1865
Number layers	8	0.7839	0.0942
Number layers	10	0.7738	0.0317

Table 6: CIFAR10CNN Pearson's correlation using Silhouette distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.2950	0.3612
Layer size	16	0.7703	0.2399
Layer size	32	0.8268	0.1422
Layer size	128	0.9307	0.0079
Layer size	256	0.9245	0.0157
Number labels	2	0.7327	0.0660
Number labels	4	0.8480	0.0572
Number labels	6	0.8443	0.0502
Number labels	8	0.8792	0.0411
Number labels	10	0.9536	0.0217
Learning rate	1e-05	0.9596	0.0051
Learning rate	0.0001	0.9317	0.0198
Learning rate	0.001	0.9270	0.0250
Learning rate	0.01	-0.0373	0.2588
Learning rate	0.1	0.0838	0.2118
Dropout	0.0	0.9411	0.0264
Dropout	0.2	0.9144	0.0229
Dropout	0.4	0.9365	0.0154
Dropout	0.5	0.8564	0.0743
Dropout	0.8	0.7389	0.0910
Input order	1	0.9468	0.0101
Input order	2	0.9232	0.0301
Input order	3	0.9482	0.0099
Input order	4	0.9168	0.0415
Input order	5	0.9327	0.0282
Number layers	2	0.9425	0.0178
Number layers	4	0.9511	0.0151
Number layers	6	0.9466	0.0150
Number layers	8	0.9613	0.0139
Number layers	10	0.9741	0.0119

Table 7: CIFAR10MLP Pearson's correlation using Heat distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.3302	0.3375
Layer size	16	0.7520	0.2474
Layer size	32	0.8016	0.1655
Layer size	128	0.9163	0.0208
Layer size	256	0.9124	0.0205
Number labels	2	0.7272	0.0622
Number labels	4	0.8307	0.0630
Number labels	6	0.8269	0.0517
Number labels	8	0.8627	0.0425
Number labels	10	0.9403	0.0262
Learning rate	1e-05	0.9611	0.0041
Learning rate	0.0001	0.9235	0.0213
Learning rate	0.001	0.9088	0.0284
Learning rate	0.01	-0.0310	0.2815
Learning rate	0.1	0.0566	0.2259
Dropout	0.0	0.9342	0.0274
Dropout	0.2	0.9082	0.0165
Dropout	0.4	0.9213	0.0247
Dropout	0.5	0.8486	0.0679
Dropout	0.8	0.7399	0.0910
Input order	1	0.9312	0.0117
Input order	2	0.9080	0.0294
Input order	3	0.9395	0.0071
Input order	4	0.9077	0.0343
Input order	5	0.9267	0.0272
Number layers	2	0.9351	0.0172
Number layers	4	0.9427	0.0178
Number layers	6	0.9419	0.0147
Number layers	8	0.9593	0.0135
Number layers	10	0.9788	0.0069

Table 8: CIFAR10MLP Pearson's correlation using Silhouette distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.8908	0.0707
Layer size	16	0.8926	0.0230
Layer size	32	0.8653	0.0332
Layer size	128	0.5020	0.1325
Layer size	256	0.2174	0.1214
Number labels	20	0.7607	0.0742
Number labels	40	0.7704	0.0761
Number labels	60	0.5372	0.2043
Number labels	80	0.4041	0.1317
Number labels	100	0.3615	0.1028
Learning rate	1e-05	0.9694	0.0067
Learning rate	0.0001	0.8980	0.0166
Learning rate	0.001	0.3175	0.0993
Learning rate	0.01	0.3461	0.2098
Learning rate	0.1	-0.1318	0.1987
Dropout	0.0	-0.0528	0.1971
Dropout	0.2	0.3034	0.0576
Dropout	0.4	0.6674	0.0835
Dropout	0.5	0.8462	0.0514
Dropout	0.8	0.9509	0.0168
Input order	1	0.6445	0.0834
Input order	2	0.5699	0.0830
Input order	3	0.6100	0.0477
Input order	4	0.5651	0.0729
Input order	5	0.6246	0.0409
Number layers	2	0.2756	0.1057
Number layers	4	0.8624	0.0419
Number layers	6	0.9706	0.0089
Number layers	8	0.9931	0.0020
Number layers	10	0.9572	0.0063

Table 9: CIFAR100CNN Pearson's correlation using Heat distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.9129	0.0642
Layer size	16	0.8940	0.0274
Layer size	32	0.8530	0.0325
Layer size	128	0.4209	0.1335
Layer size	256	0.1129	0.1202
Number labels	20	0.7450	0.0777
Number labels	40	0.7361	0.0734
Number labels	60	0.4762	0.2142
Number labels	80	0.3103	0.1319
Number labels	100	0.2614	0.0929
Learning rate	1e-05	0.9806	0.0052
Learning rate	0.0001	0.9015	0.0188
Learning rate	0.001	0.2212	0.1084
Learning rate	0.01	0.2032	0.2316
Learning rate	0.1	-0.1386	0.2254
Dropout	0.0	-0.1262	0.1812
Dropout	0.2	0.1994	0.0553
Dropout	0.4	0.5566	0.0944
Dropout	0.5	0.7549	0.0621
Dropout	0.8	0.8843	0.0347
Input order	1	0.5514	0.0865
Input order	2	0.4804	0.0842
Input order	3	0.5161	0.0472
Input order	4	0.4686	0.0772
Input order	5	0.5282	0.0374
Number layers	2	0.1789	0.1113
Number layers	4	0.8221	0.0507
Number layers	6	0.9577	0.0124
Number layers	8	0.9850	0.0032
Number layers	10	0.9531	0.0066

Table 10: CIFAR100CNN Pearson's correlation using Silhouette distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.5640	0.6287
Layer size	16	0.9096	0.0500
Layer size	32	0.9275	0.0152
Layer size	128	0.9253	0.0191
Layer size	256	0.9348	0.0208
Number labels	20	0.8612	0.0283
Number labels	40	0.9361	0.0188
Number labels	60	0.9233	0.0178
Number labels	80	0.9515	0.0120
Number labels	100	0.9314	0.0204
Learning rate	1e-05	0.9864	0.0056
Learning rate	0.0001	0.9813	0.0046
Learning rate	0.001	0.9385	0.0084
Learning rate	0.01	0.7182	0.0230
Learning rate	0.1	-0.1033	0.1915
Dropout	0.0	0.9298	0.0307
Dropout	0.2	0.9274	0.0101
Dropout	0.4	0.9372	0.0133
Dropout	0.5	0.8675	0.0518
Dropout	0.8	0.0635	0.3428
Input order	1	0.9278	0.0135
Input order	2	0.9425	0.0066
Input order	3	0.9315	0.0251
Input order	4	0.9221	0.0259
Input order	5	0.9172	0.0133
Number layers	2	0.9346	0.0112
Number layers	4	0.9135	0.0315
Number layers	6	0.8876	0.0143
Number layers	8	0.8980	0.0140
Number layers	10	0.8733	0.0297

Table 11: CIFAR100MLP Pearson's correlation using Heat distance.

Analysis type	Analysis Value	Pearson's r mean	Pearson's r standard deviation
Layer size	4	0.5423	0.6368
Layer size	16	0.9261	0.0409
Layer size	32	0.9602	0.0092
Layer size	128	0.9689	0.0162
Layer size	256	0.9669	0.0103
Number labels	20	0.8617	0.0284
Number labels	40	0.9375	0.0194
Number labels	60	0.9484	0.0095
Number labels	80	0.9697	0.0048
Number labels	100	0.9706	0.0076
Learning rate	1e-05	0.9882	0.0053
Learning rate	0.0001	0.9792	0.0047
Learning rate	0.001	0.9754	0.0065
Learning rate	0.01	0.7968	0.0622
Learning rate	0.1	-0.0751	0.2187
Dropout	0.0	0.9635	0.0155
Dropout	0.2	0.9689	0.0056
Dropout	0.4	0.9405	0.0117
Dropout	0.5	0.8099	0.0619
Dropout	0.8	-0.0034	0.3426
Input order	1	0.9659	0.0147
Input order	2	0.9740	0.0051
Input order	3	0.9722	0.0131
Input order	4	0.9652	0.0135
Input order	5	0.9656	0.0068
Number layers	2	0.9708	0.0090
Number layers	4	0.9512	0.0235
Number layers	6	0.9330	0.0079
Number layers	8	0.9272	0.0122
Number layers	10	0.8917	0.0236

Table 12: CIFAR100MLP Pearson's correlation using Silhouette distance.

0.1   Correlation Points
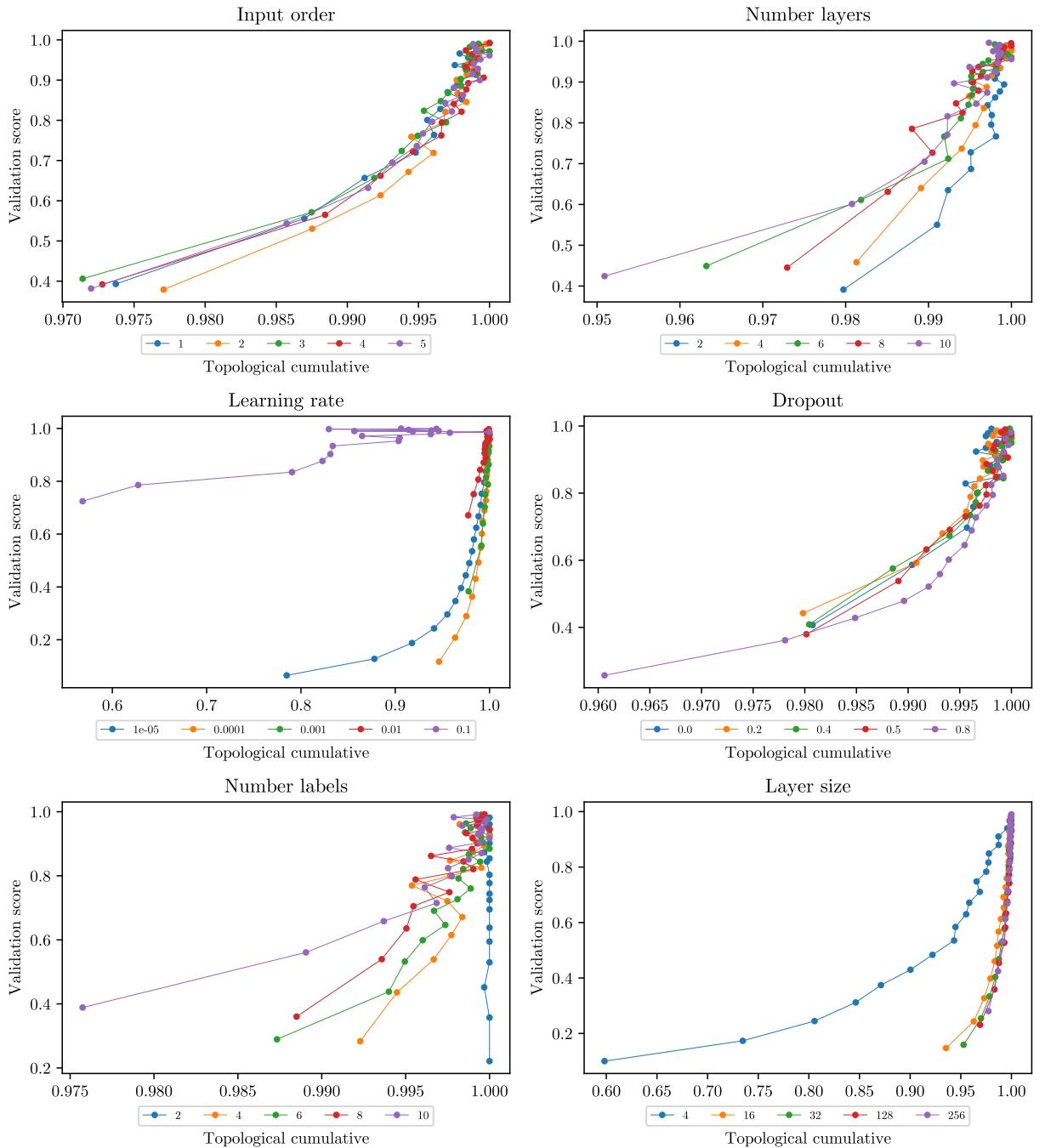
Figure 42: MNIST points relationship across epochs using Heat discretization.
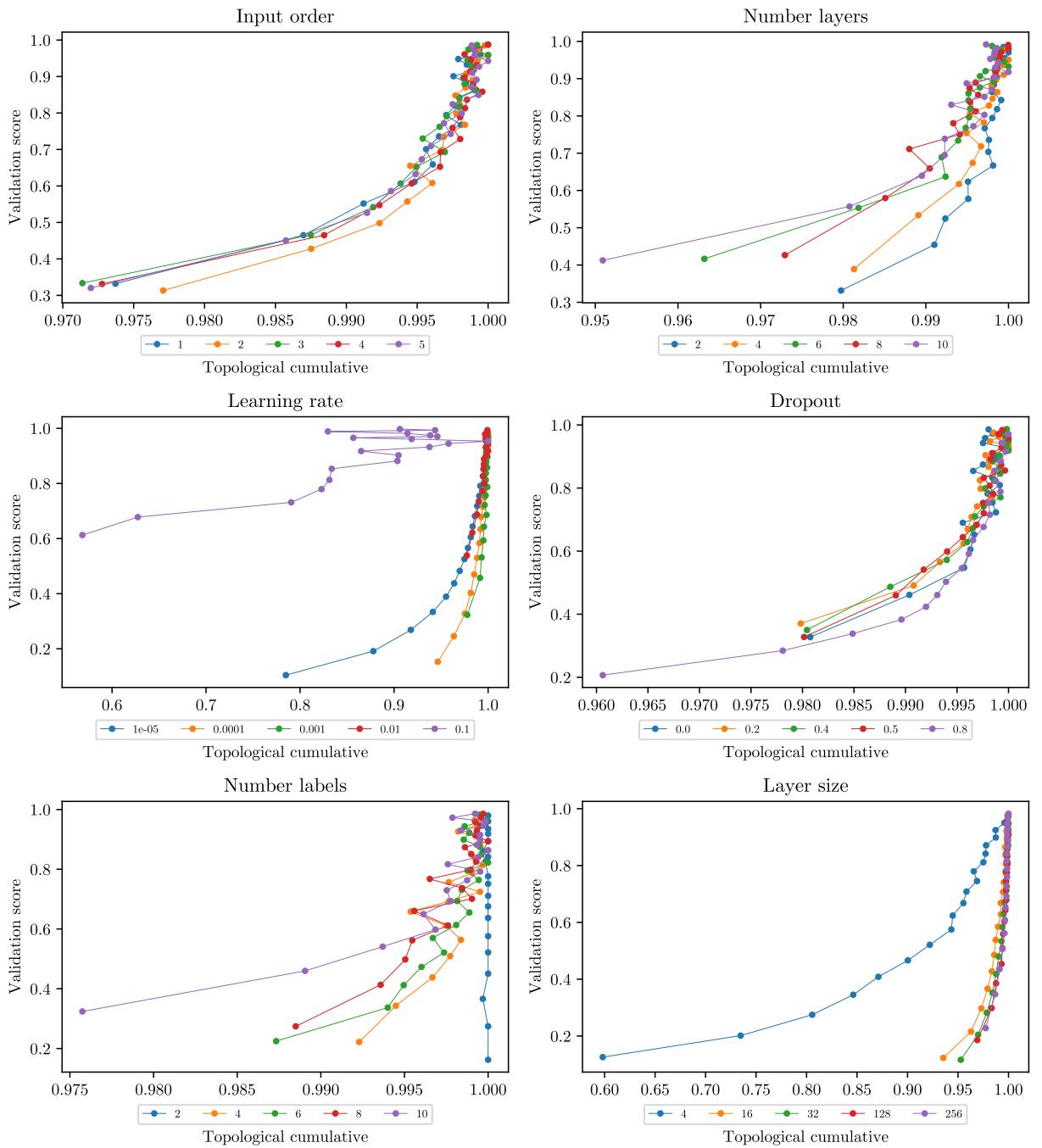
Figure 43: MNIST points relationship across epochs using Silhouette discretization.
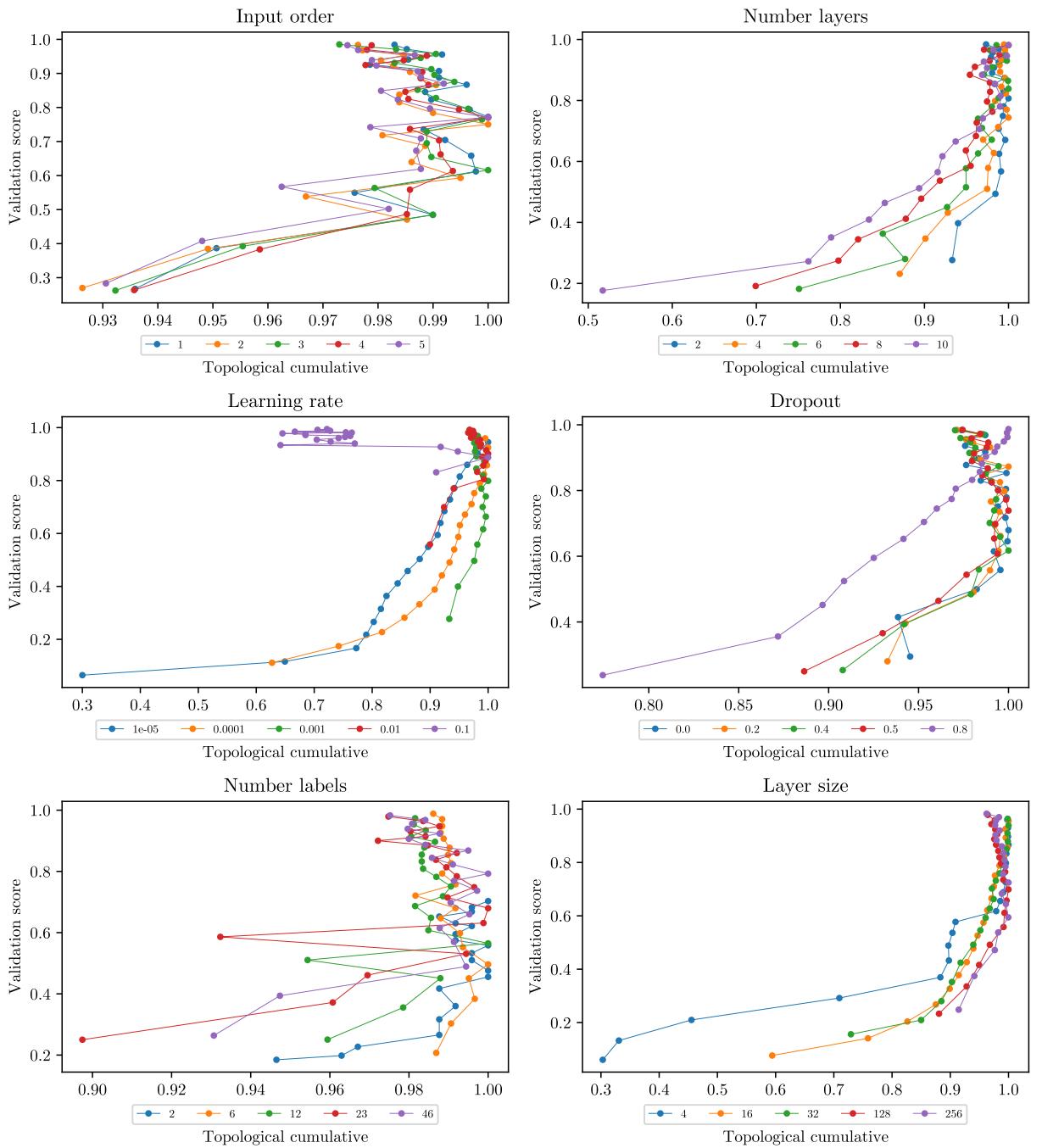
Figure 44: Reuters points relationship across epochs using Heat discretization.
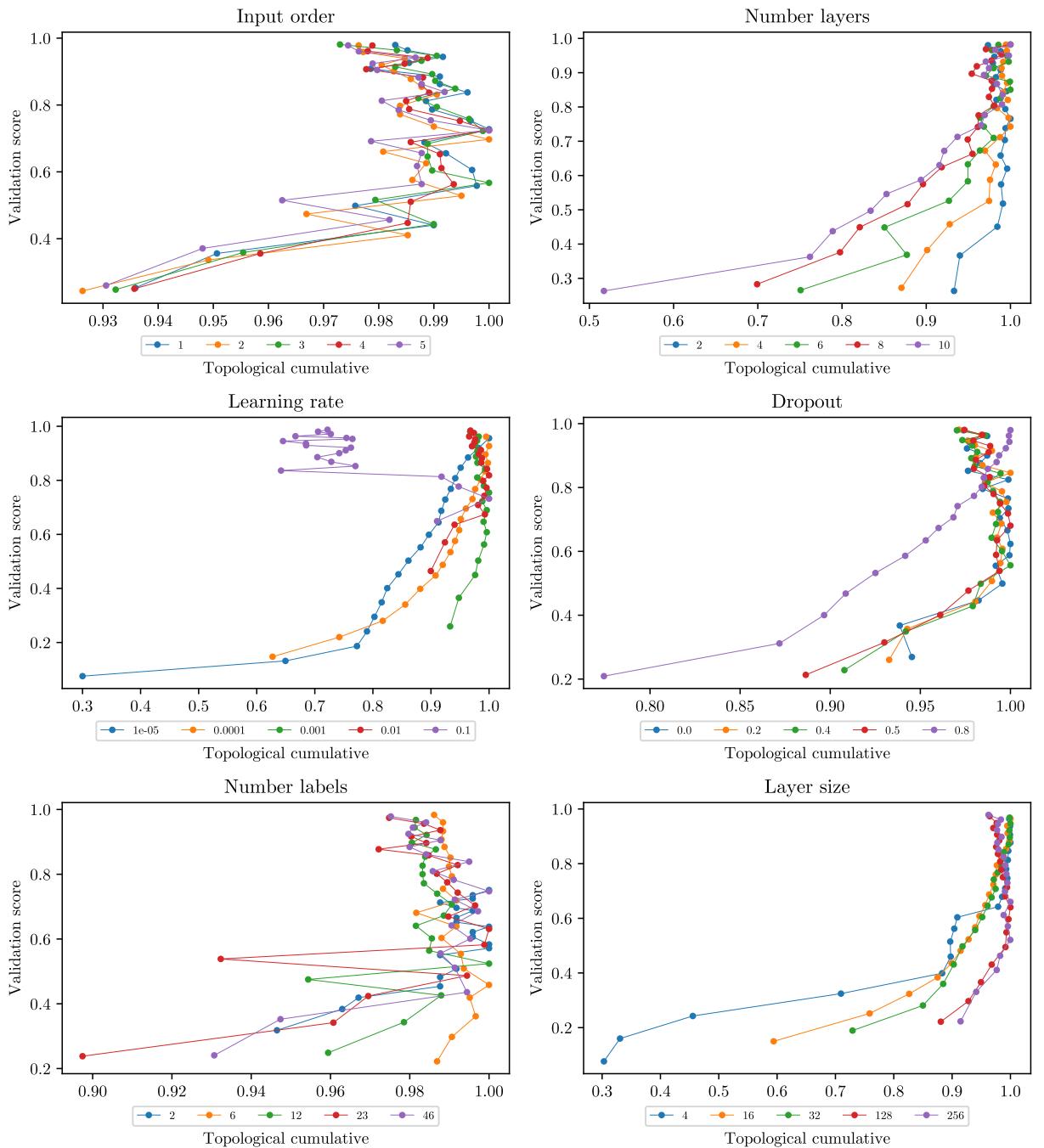
Figure 45: Reuters points relationship across epochs using Silhouette discretization.
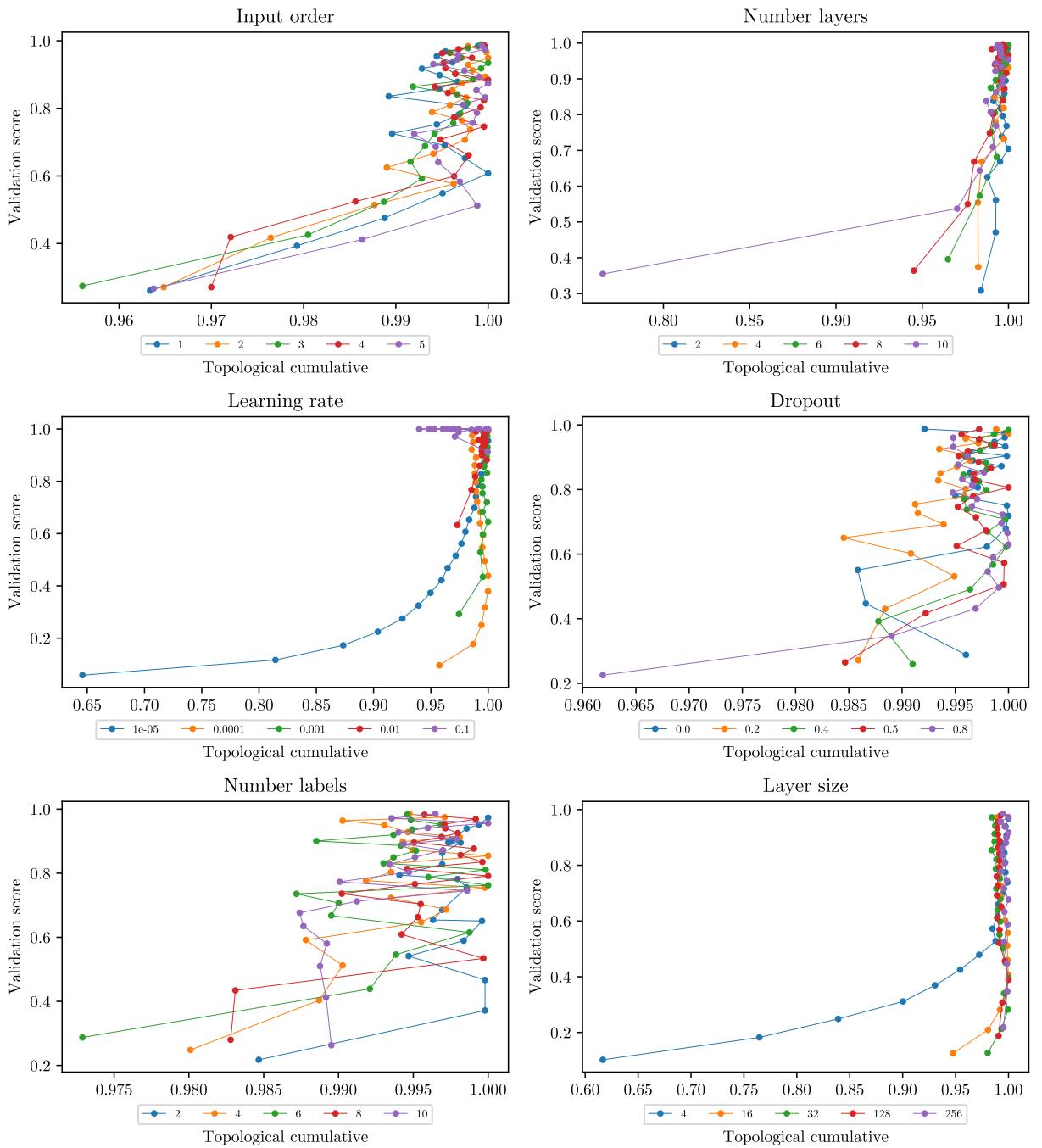
Figure 46: CIFAR-10 CNN points relationship across epochs using Heat discretization.
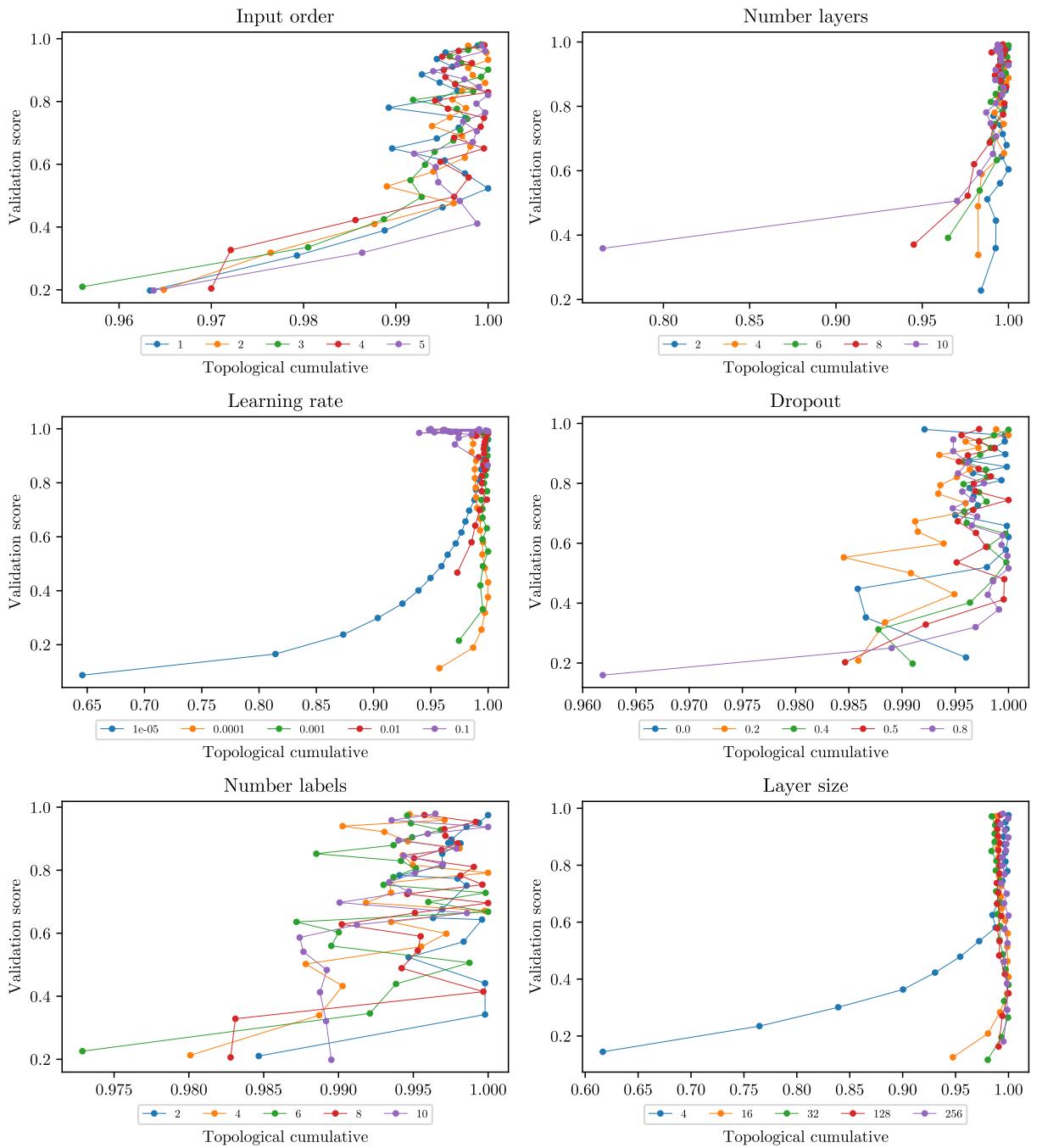
Figure 47: CIFAR-10 CNN points relationship across epochs using Silhouette discretization.
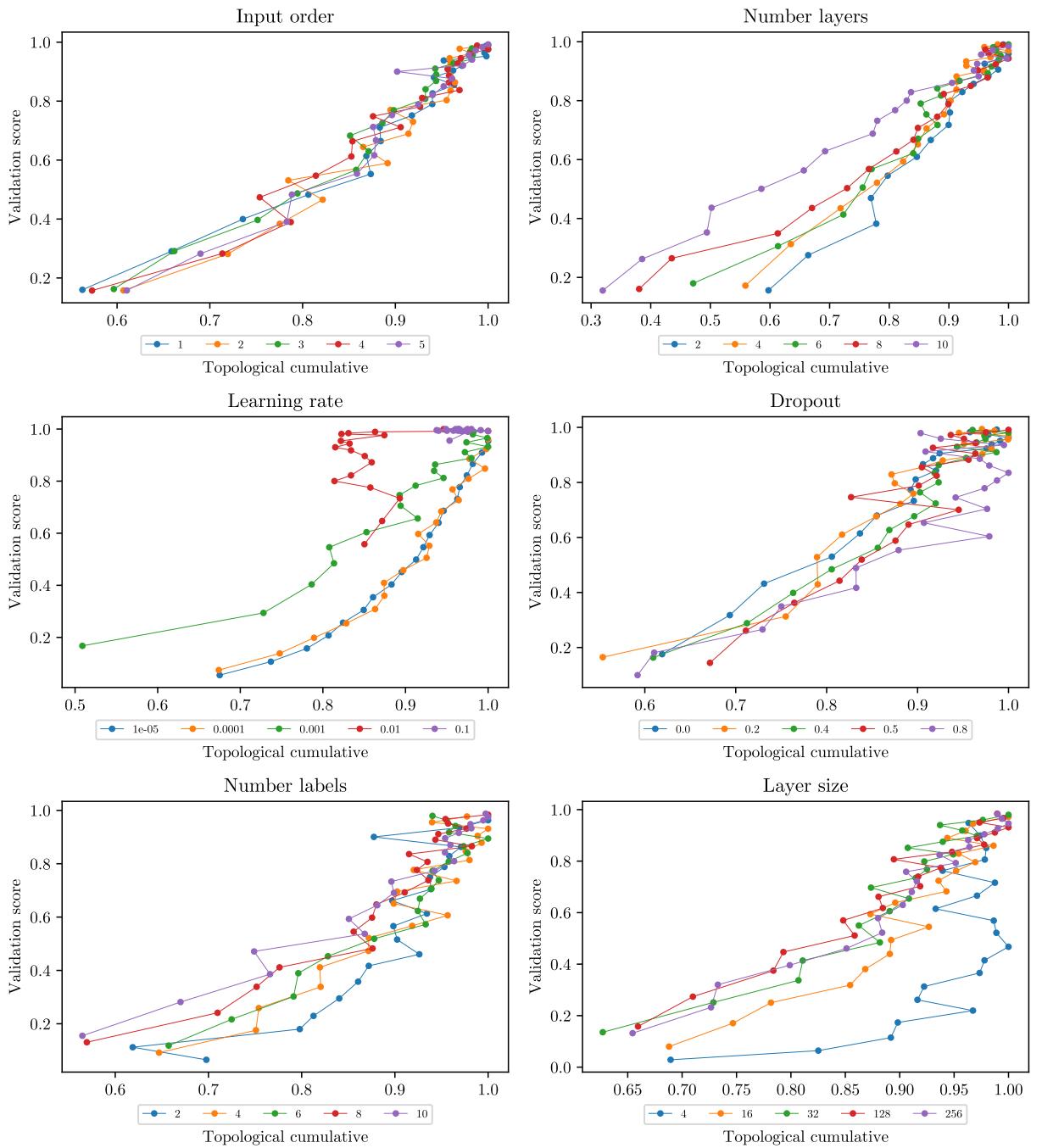
Figure 48: CIFAR-10 MLP points relationship across epochs using Heat discretization.
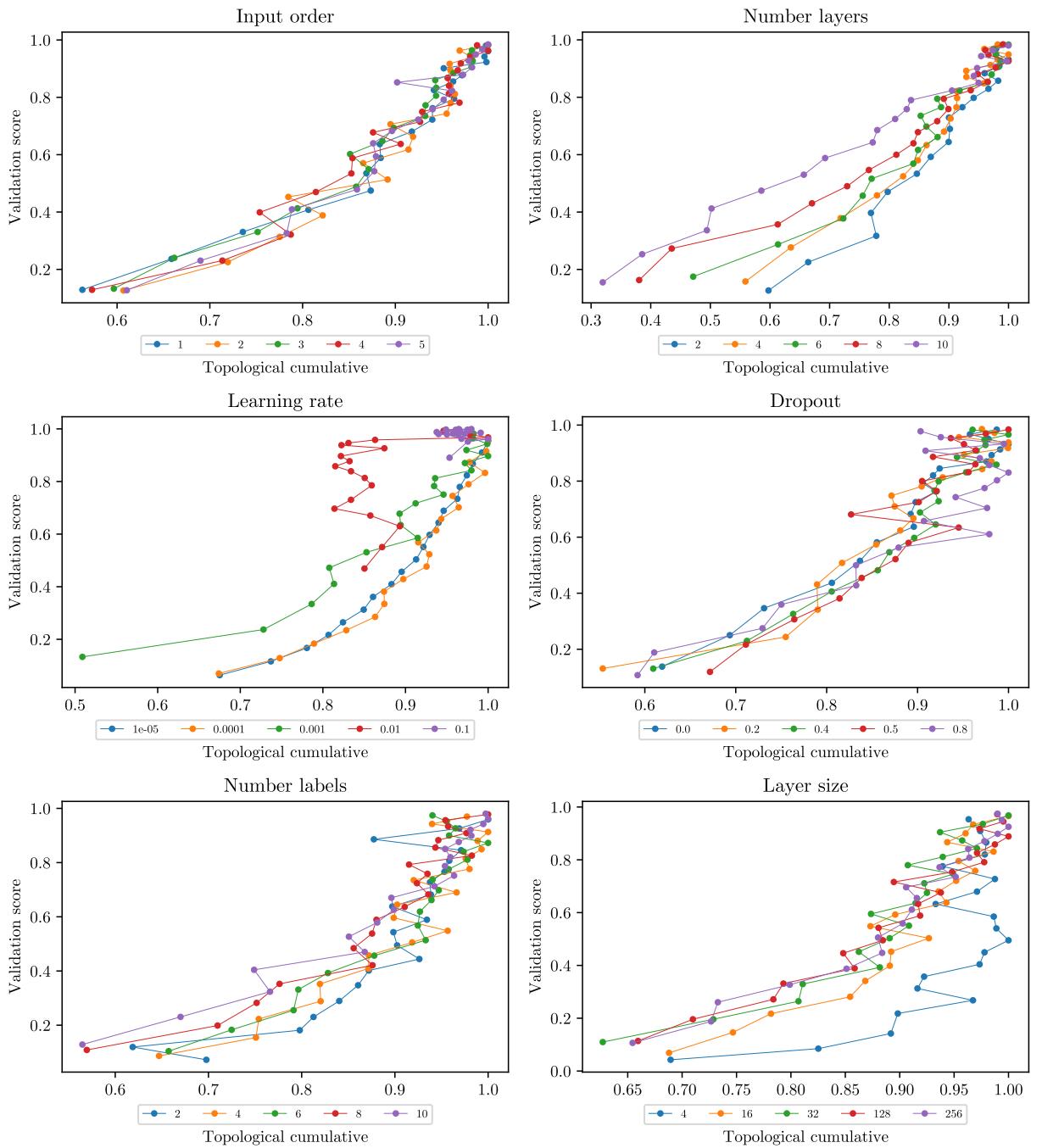
Figure 49: CIFAR-10 MLP points relationship across epochs using Silhouette discretization.
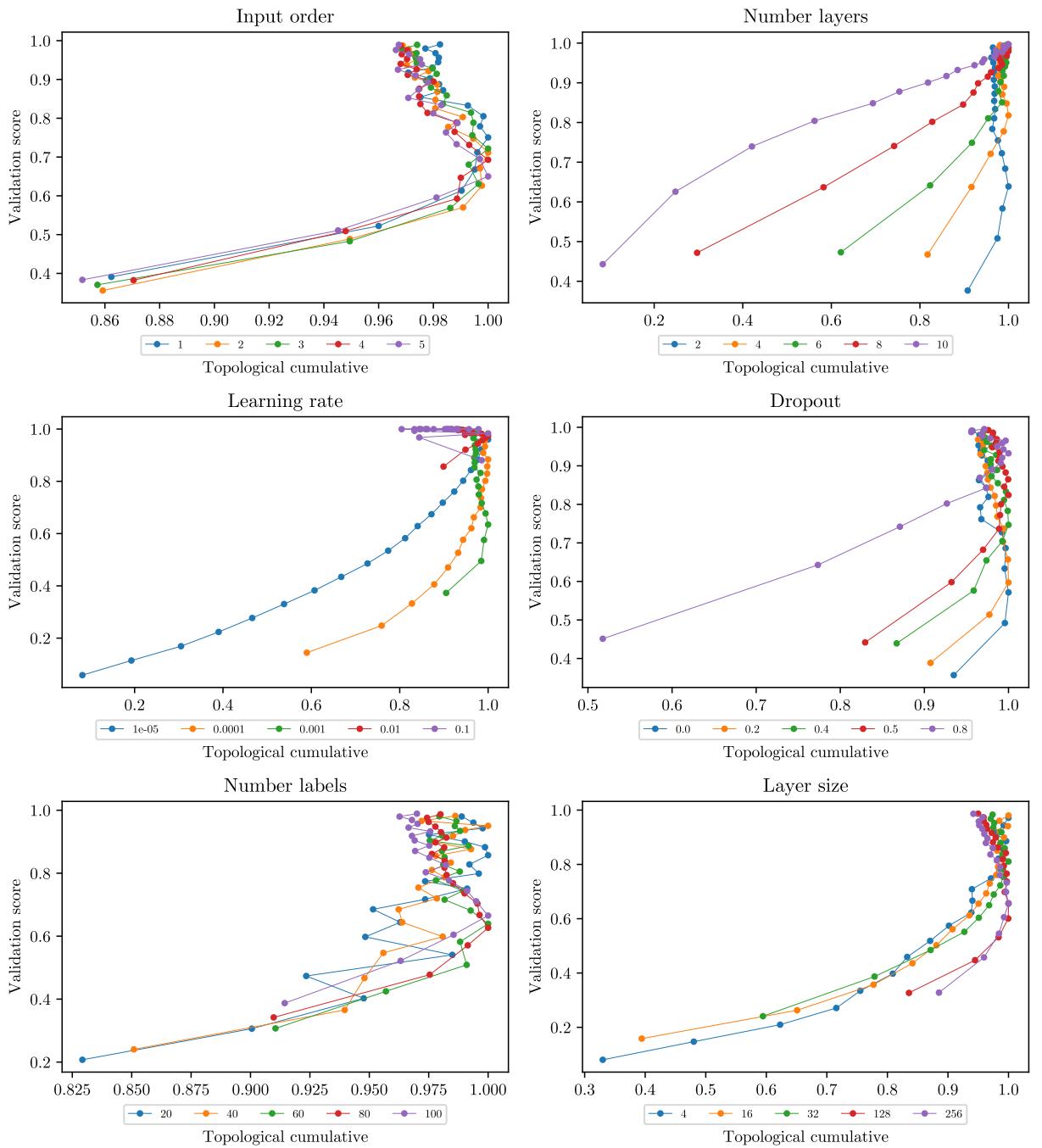
Figure 50: CIFAR-100 CNN points relationship across epochs using Heat discretization.
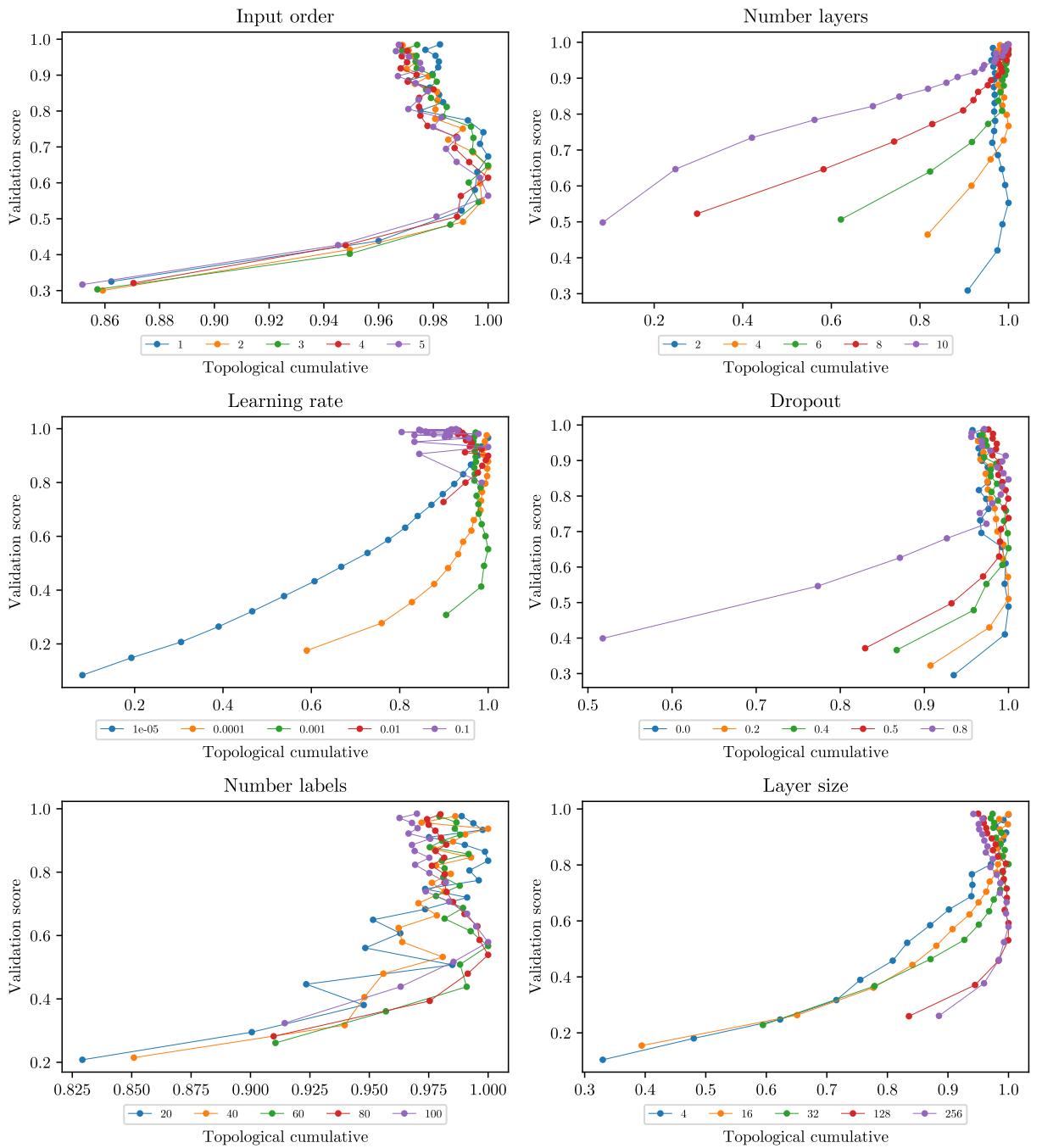
Figure 51: CIFAR-100 CNN points relationship across epochs using Silhouette discretization.
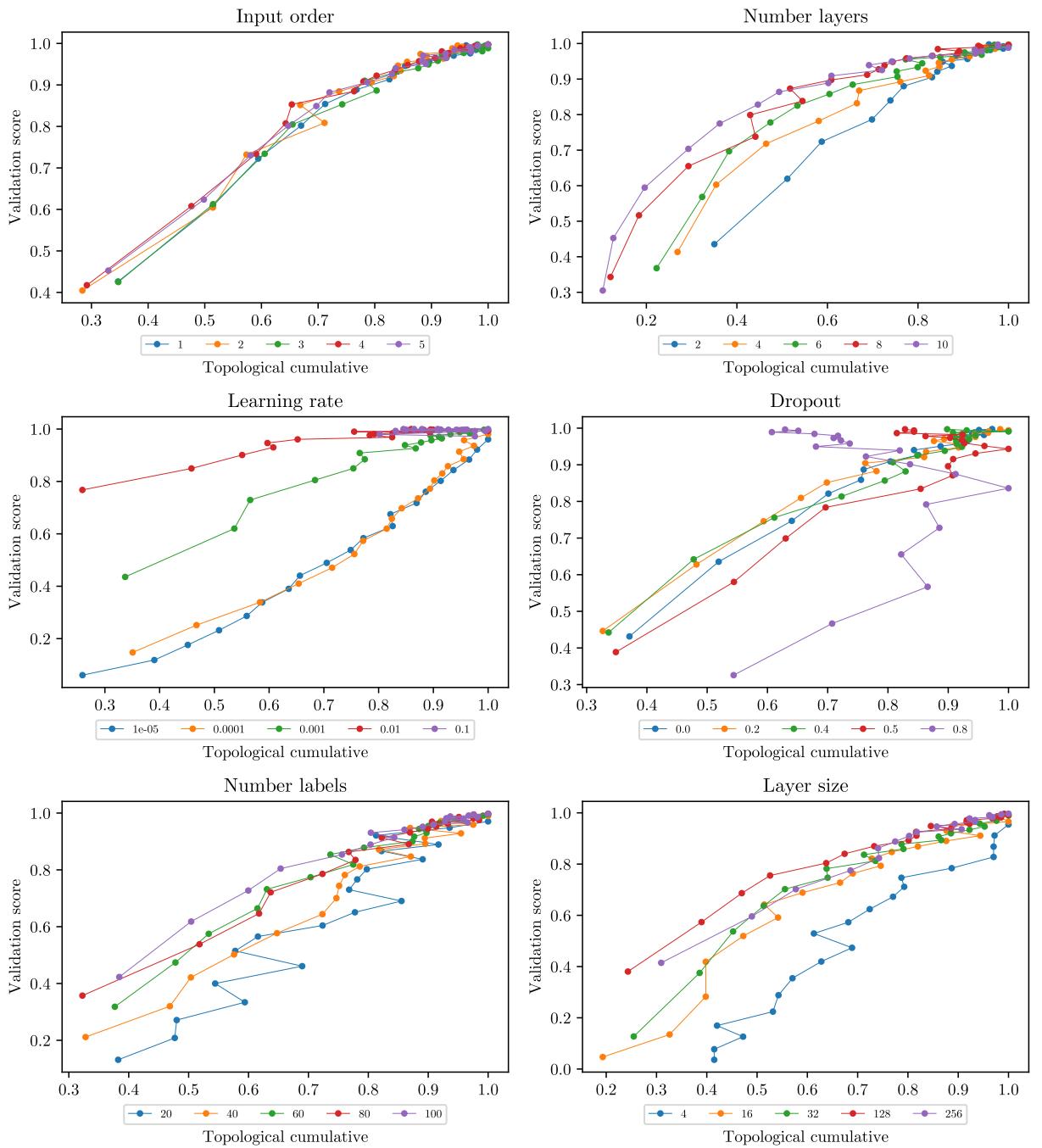
Figure 52: CIFAR-100 MLP points relationship across epochs using Heat discretization.
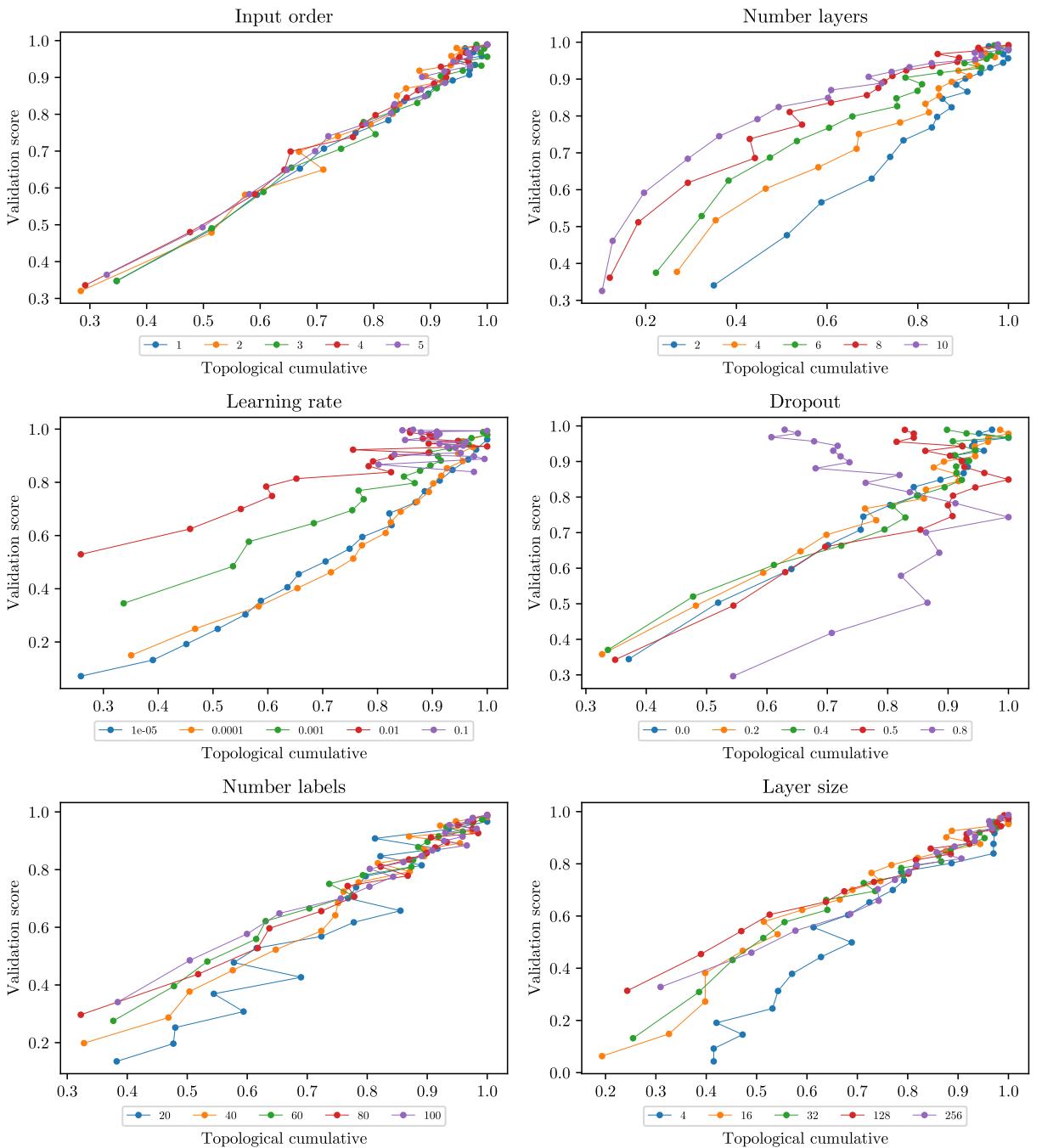
Figure 53: CIFAR-100 MLP points relationship across epochs using Silhouette discretization.

References

[1] E. Berry, Y.-C. Chen, J. Cisewski-Kehe, and B. T. Fasy. Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, 4:211–262, 2020.

[2] G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and K. Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration, 2020.