

深層学習における ドロップアウトデザインの適用

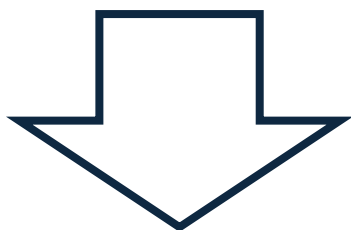
6321120 横溝 尚也

本日の発表内容

1. 研究背景
2. 研究目的
3. ドロップアウト法
4. ドロップアウトデザイン
5. ドロップアウトデザインの適用
6. 実験概要
7. 実験結果
8. まとめ

深層学習での課題

- 学習コストの高さ
- 過学習への正確な対処



今までの歴史の中でも直行表による実験計画法は計画的な実験により実験のコストを削減してきた

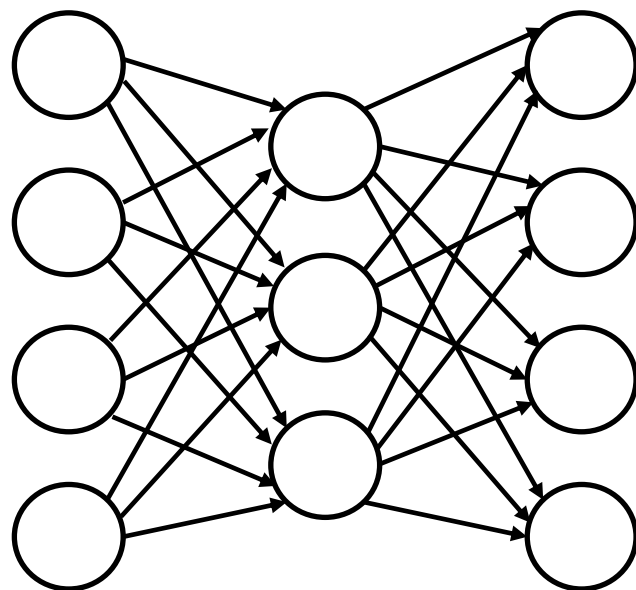
正則化手法の一種であるドロップアウトに着目し、
ドロップアウトに組み合わせ最適化を導入した新たな手法を検討する

既存のドロップアウト法とドロップアウトデザインとの正則化効果を考察し、二つのモデルを比較してモデルの収束速度に違いがないか検証する

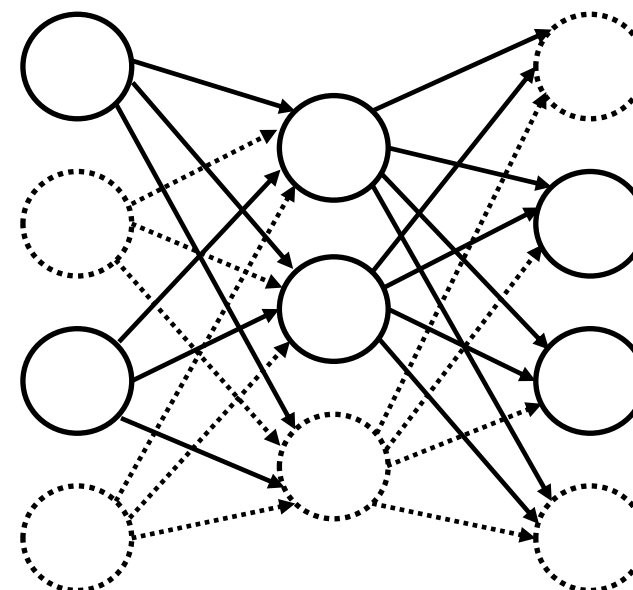
ドロップアウト法

ドロップアウト法とは

ニューラルネットワークにおいて一定確率でノードを不活性にすることで訓練モデルを簡潔化し、過学習を抑制する手法



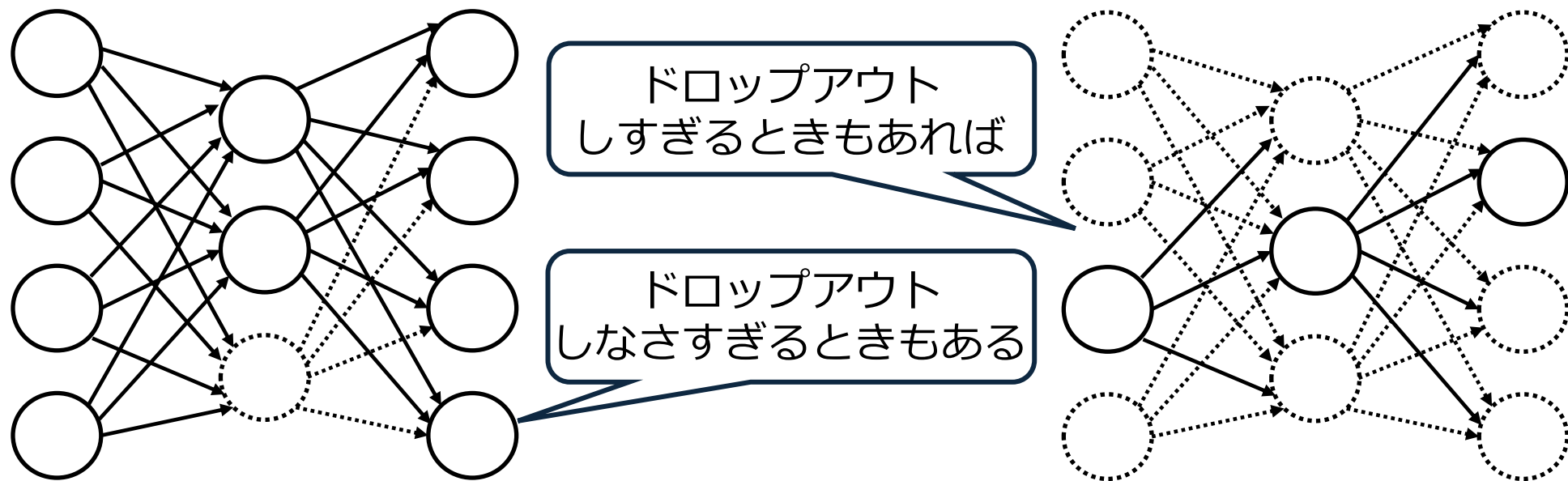
ドロップアウト後



ドロップアウト法

ドロップアウト法の問題点

ハイパパラメータによって指定された確率に従って不活性にするノードを選ぶため、ドロップアウトの**安定性に欠ける**



ドロップアウトデザインとは

- 要素全体の中からブロックと呼ばれる単位で要素を抽出する
ブロックデザイン的一种
- 以下のような正則性を持つようにブロックを構成
 - ある要素がブロックに選ばれる回数が一定
 - 任意の要素のペアが同時に選ばれる回数が一定

ブロックとして要素を抽出する回数が均一であるため
ノード選出に偏りがなくなる

定義 (ドロップアウトデザイン)

V_1, V_2, \dots, V_n をそれぞれ要素数 v の異なる点集合とし, ブロック集合を

$$B = \{ \{C_1|C_2|\dots|C_n\} \mid C_i \in V_i, |C_i| = k, 1 \leq i \leq n \}$$

とする. B をブロック集合, 各 $\{C_1|C_2|\dots|C_n\}$ をブロック, 各 C_i をサブブロックと呼ぶ.

B のうち, 連続する t 個の点集合におけるサブブロックの集合を

$$B|_{V_i, V_{i+1}, \dots, V_{i+t-1}} = \{ \{C_i|C_{i+1}|\dots|C_{i+t-1}\} \mid C_j \subset V_j, i = 1, \dots, n-t+1 \}$$

とする. このとき, 次の条件を満たす $(V_1, \dots, V_n; B)$ を (d_1, \dots, d_t) 型 $(v, k, \lambda; n)$ -DDという.

(条件)

t 個の連続する点集合 $V_{i+1}, V_{i+2}, \dots, V_{i+t}$ ($0 \leq i \leq n-t$)において, 各 V_{i+j} , $j = 1, 2, \dots, t$ の中からそれぞれ任意に d_j 個ずつ取り出した点集合を同時に含む $B|_{V_{i+1}, V_{i+2}, \dots, V_{i+t}}$ の中のブロック数は λ 個存在する.

要素数 v の n 個の集合から k 個ずつ選んだとき、
任意の (d_1, \dots, d_t) 個の組み合わせが λ 回出現する組み合わせ構造のこと

具体例

点集合をそれぞれ

$V_1 = \{0, 1, 2, 3\}, V_2 = \{a, b, c, d\}, V_3 = \{\alpha, \beta, \gamma, \delta\}$ とする.

以下のブロック集合は(1,2)型(4,2,1;3)-DDをなす.

$$\begin{aligned} \mathcal{B} = \{ & \{0, 1|a, b|\alpha, \beta\}, \{0, 1|b, c|\alpha, \beta\}, \\ & \{0, 2|b, d|\alpha, \gamma\}, \{0, 2|a, c|\alpha, \gamma\}, \\ & \{0, 3|b, c|\alpha, \delta\}, \{0, 3|a, d|\alpha, \gamma\}, \\ & \{1, 2|b, d|\beta, \gamma\}, \{1, 2|a, d|\beta, \gamma\}, \\ & \{1, 3|b, d|\beta, \delta\}, \{1, 3|a, c|\beta, \gamma\}, \\ & \{2, 3|a, b|\gamma, \delta\}, \{2, 3|c, d|\gamma, \delta\} \} \end{aligned}$$

具体例

点集合をそれぞれ

$V_1 = \{0, 1, 2, 3\}, V_2 = \{a, b, c, d\}, V_3 = \{\alpha, \beta, \gamma, \delta\}$ とする.
以下のブロック集合は(1,2)型(4,2,1;3)-DDをなす.

$$\mathcal{B} = \{\{0, 1|a, b|\alpha, \beta\}, \{0, 1|b, c|\alpha, \beta\}, \\ \{0, 2|b, d|\alpha, \gamma\}, \{0, 2|a, c|\alpha, \gamma\}, \\ \{0, 3|b, c|\alpha, \delta\}, \{0, 3|a, d|\alpha, \gamma\}, \\ \{1, 2|b, d|\beta, \gamma\}, \{1, 2|a, d|\beta, \gamma\}, \\ \{1, 3|b, d|\beta, \delta\}, \{1, 3|a, c|\beta, \gamma\}, \\ \{2, 3|a, b|\gamma, \delta\}, \{2, 3|c, d|\gamma, \delta\}\}$$

0, a, bが同じブロック
に含まれるのは1回

このドロップアウトデザインは
以下の性質を持つ

- 点集合は3つ
↳ 点集合の個数 $n = 3$
- 各点集合の要素数は一定
↳ 点集合の要素数 $v = 4$
- 各ブロックは各点集合から一定個の要素を含んだ集合
↳ サブブロックサイズ $k = 2$
- 連続する2集合における任意の(1,2)個ずつ選んだ要素の組み合わせは一定個のブロックに含まれる
↳ 会合数 $\lambda = 1$

定理 (大規模デザインの構成)

d を3以上の整数とし, q を素数べきとする. このとき, $(2, 1)$ 型 $(v, k, \lambda; n)$ -DDが存在する.ただし,

$$v = q^t, \quad k = q^{t-1}, \quad \lambda = \frac{q^{d-2} - q^{d-t-1}}{q-1}, \quad n = q^{d-t}$$

であり, t は $2 \leq t \leq d-1$ となる整数である.

位数 q , d 次元のアフィン空間における
 t 次元部分空間を考えることで構成

具体例

$d = 8, q = 2, t = 7$ とすると $(2, 1)$ 型 $(128, 64, 63; 2)$ -DD
という大規模パラメータを持つデザインを構成できる

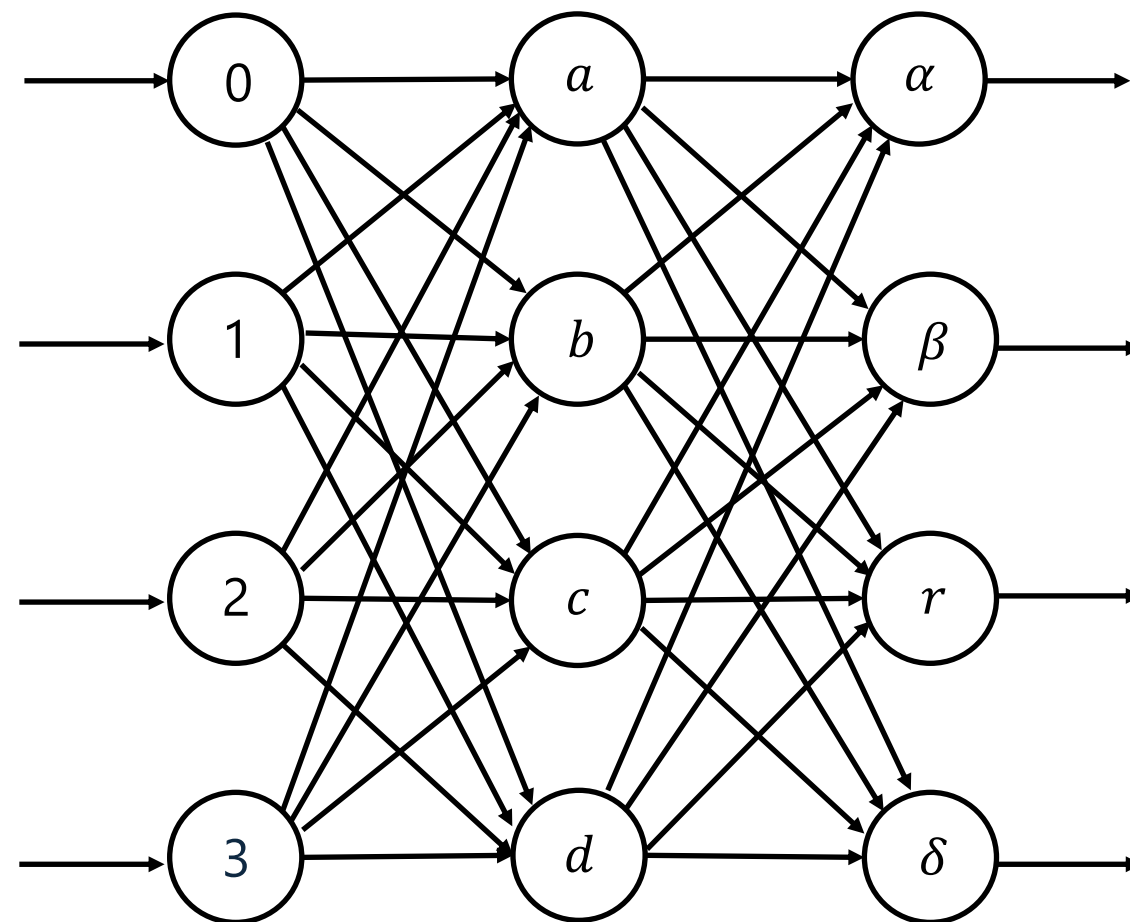
具体例

点集合をそれぞれ

$V_1 = \{0, 1, 2, 3\}, V_2 = \{a, b, c, d\}, V_3 = \{\alpha, \beta, \gamma, \delta\}$
とする.

以下のブロック集合は(1, 2)型(4, 2, 1; 3)-DDをなす.

$$\begin{aligned} \mathcal{B} = \{ & \{0, 1|a, b|\alpha, \beta\}, \{0, 1|b, c|\alpha, \beta\}, \\ & \{0, 2|b, d|\alpha, \gamma\}, \{0, 2|a, c|\alpha, \gamma\}, \\ & \{0, 3|b, c|\alpha, \delta\}, \{0, 3|a, d|\alpha, \gamma\}, \\ & \{1, 2|b, d|\beta, \gamma\}, \{1, 2|a, d|\beta, \gamma\}, \\ & \{1, 3|b, d|\beta, \delta\}, \{1, 3|a, c|\beta, \gamma\}, \\ & \{2, 3|a, b|\gamma, \delta\}, \{2, 3|c, d|\gamma, \delta\} \end{aligned}$$



ネットワーク構造

具体例

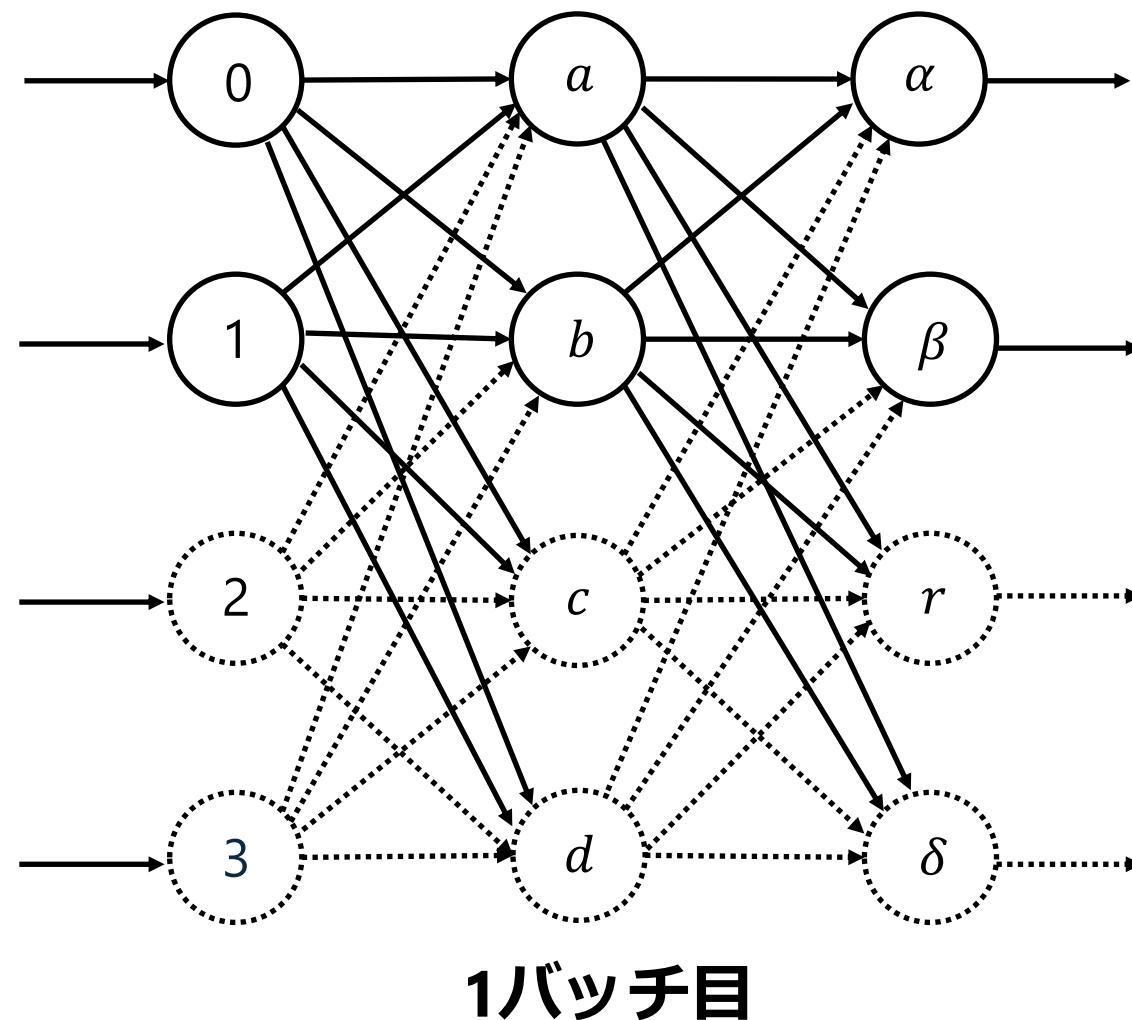
点集合をそれぞれ

$V_1 = \{0, 1, 2, 3\}, V_2 = \{a, b, c, d\}, V_3 = \{\alpha, \beta, \gamma, \delta\}$
とする.

以下のブロック集合は(1, 2)型(4, 2, 1; 3)-DDをなす.

1バッチ目使用ブロック

$\mathcal{B} = \{\{0, 1|a, b|\alpha, \beta\}, \{0, 1|b, c|\alpha, \beta\},$
 $\{0, 2|b, d|\alpha, \gamma\}, \{0, 2|a, c|\alpha, \gamma\},$
 $\{0, 3|b, c|\alpha, \delta\}, \{0, 3|a, d|\alpha, \gamma\},$
 $\{1, 2|b, d|\beta, \gamma\}, \{1, 2|a, d|\beta, \gamma\},$
 $\{1, 3|b, d|\beta, \delta\}, \{1, 3|a, c|\beta, \gamma\},$
 $\{2, 3|a, b|\gamma, \delta\}, \{2, 3|c, d|\gamma, \delta\}\}$



具体例

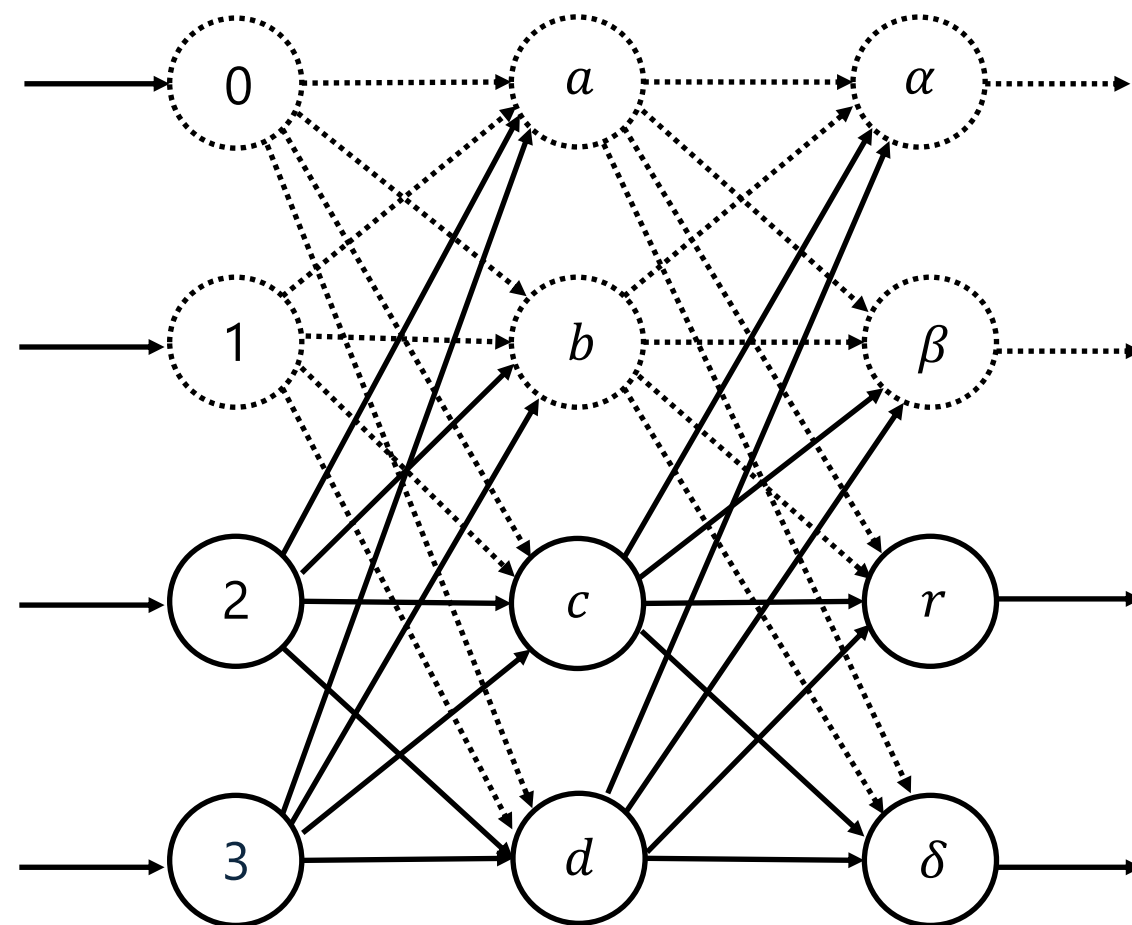
点集合をそれぞれ

$V_1 = \{0, 1, 2, 3\}, V_2 = \{a, b, c, d\}, V_3 = \{\alpha, \beta, \gamma, \delta\}$
とする.

以下のブロック集合は(1, 2)型(4, 2, 1; 3)-DDをなす.

$$\mathcal{B} = \{\{0, 1|a, b|\alpha, \beta\}, \{0, 1|b, c|\alpha, \beta\}, \\ \{0, 2|b, d|\alpha, \gamma\}, \{0, 2|a, c|\alpha, \gamma\}, \\ \{0, 3|b, c|\alpha, \delta\}, \{0, 3|a, d|\alpha, \gamma\}, \\ \{1, 2|b, d|\beta, \gamma\}, \{1, 2|a, d|\beta, \gamma\}, \\ \{1, 3|b, d|\beta, \delta\}, \{1, 3|a, c|\beta, \gamma\}, \\ \{2, 3|a, b|\gamma, \delta\}, \boxed{\{2, 3|c, d|\gamma, \delta\}}\}$$

12バッチ目使用ブロック



12バッチ目

比較モデル

1. 正則化なし
2. ドロップアウト法
3. ドロップアウトデザイン(shift)
4. ドロップアウトデザイン(non shift)

以上4モデルの正則化効果を検証する

ニューラルネットワーク構造

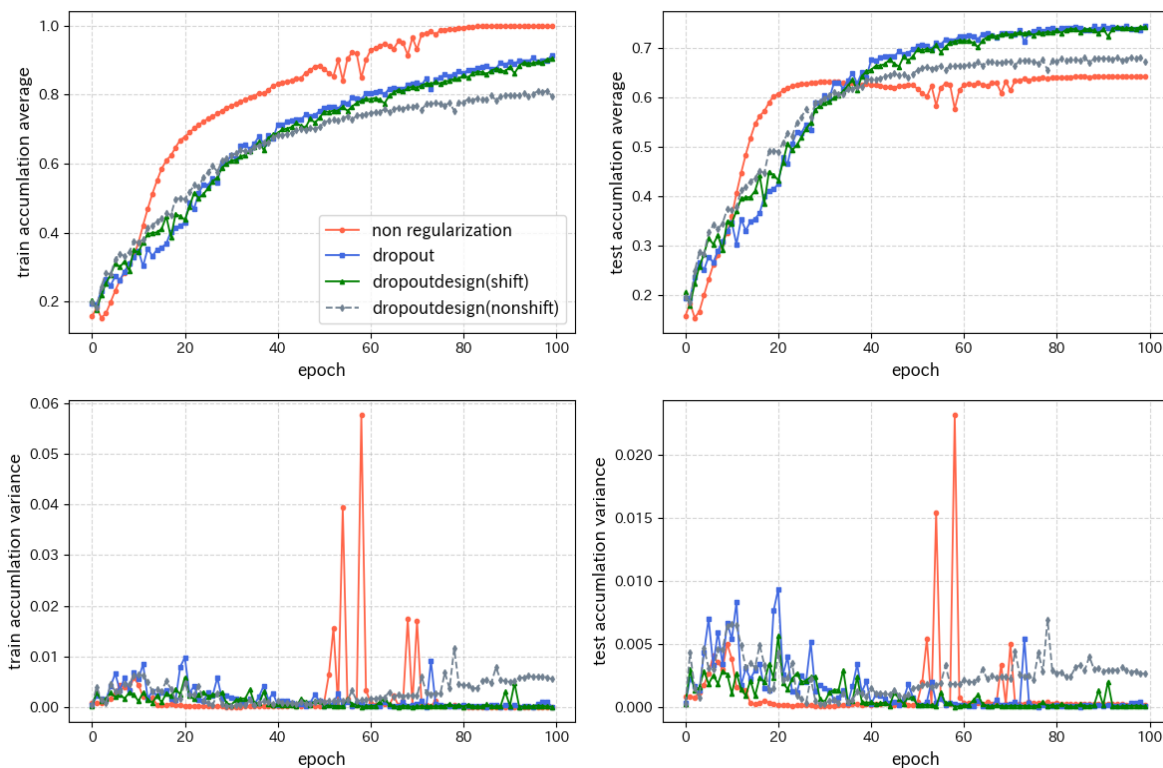
- ニューラルネットワーク構造は使用するドロップアウトデザインから自動的に決定される

実験	ドロップアウトデザイン					ニューラルネットワーク				
	サブブロック数	点集合サイズ	サブブロックサイズ	ブロック数	会合数	層数	1層あたりのノード数	1層あたりの活性化ノード数	バッチ数	ドロップアウト率
1	2	128	64	508	63	2	128	64	508	0.5
2	3	729	243	3276	121	3	729	243	3276	0.67

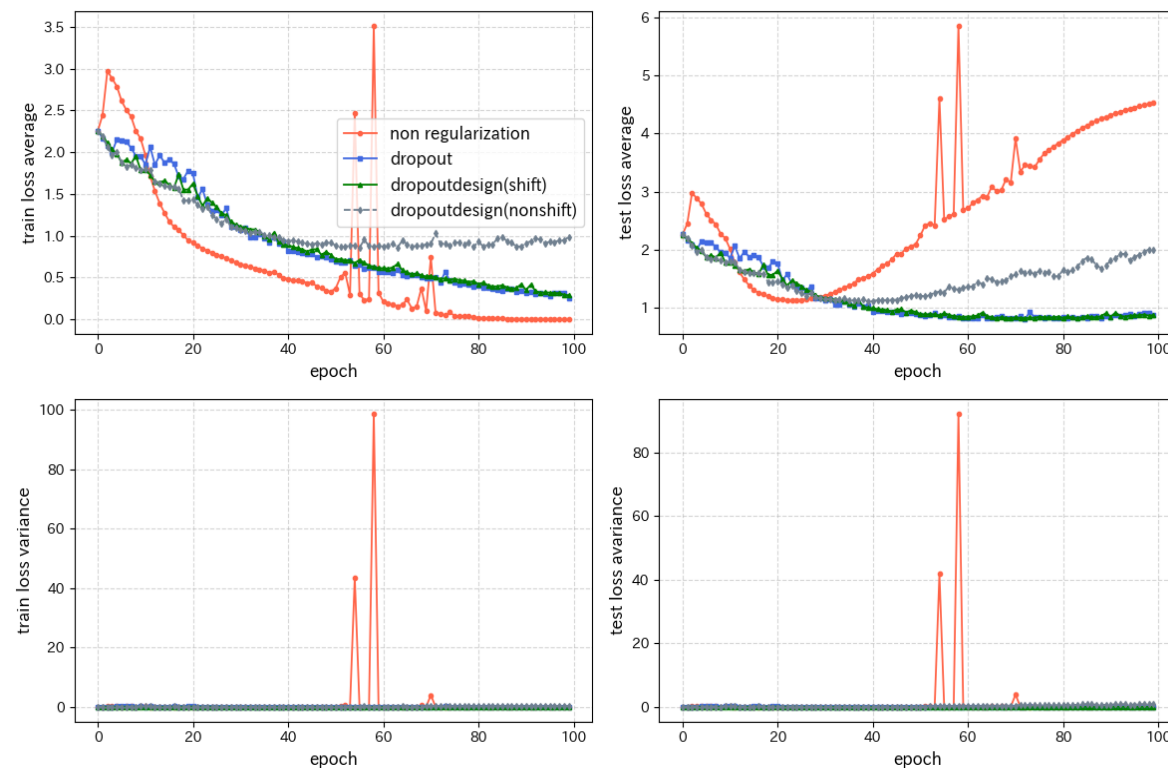
- CNNニューラルネットワークを用いたcifar10データセットにおける画像分類を扱う
- 100エポックの訓練データ・テストデータの精度と損失を計測
- 実験を10回ずつ繰り返し、各指標の平均・分散などの統計量を算出
- ドロップアウトアルゴリズム以外のその他パラメータは統一して比較実験を行う

128ノード × 2層 dropout rate=0.5 のニューラルネットワーク

精度

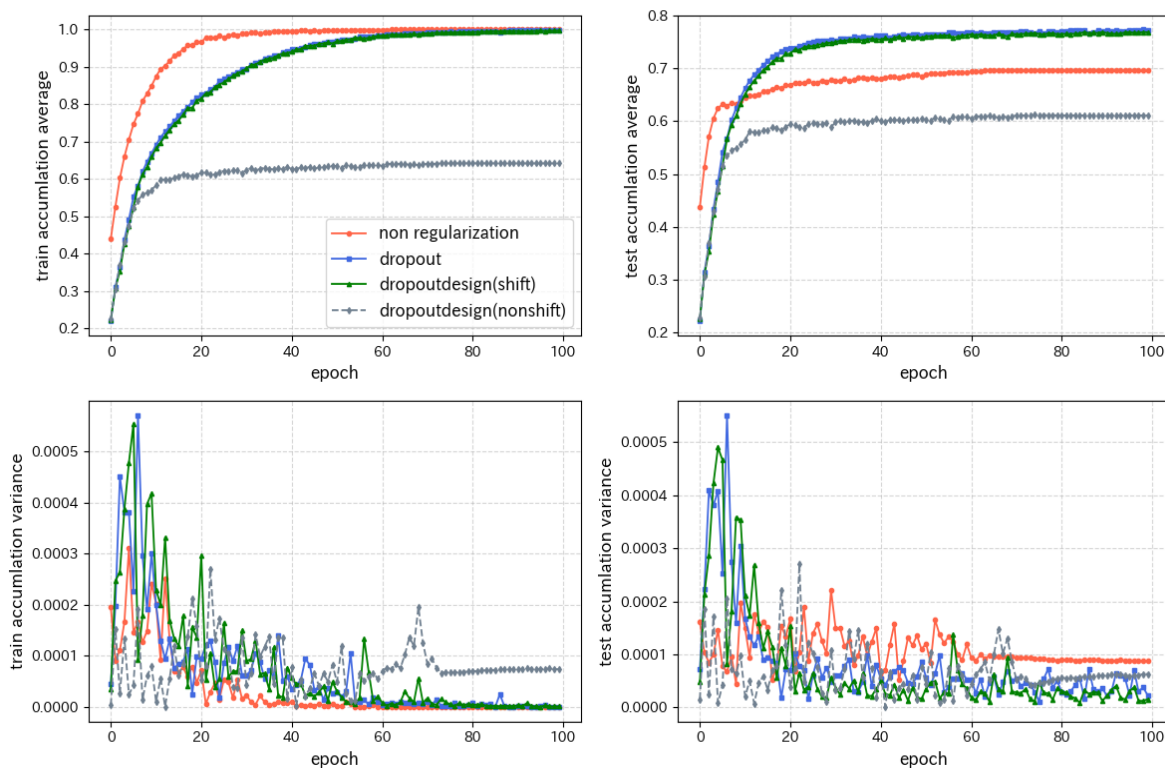


損失

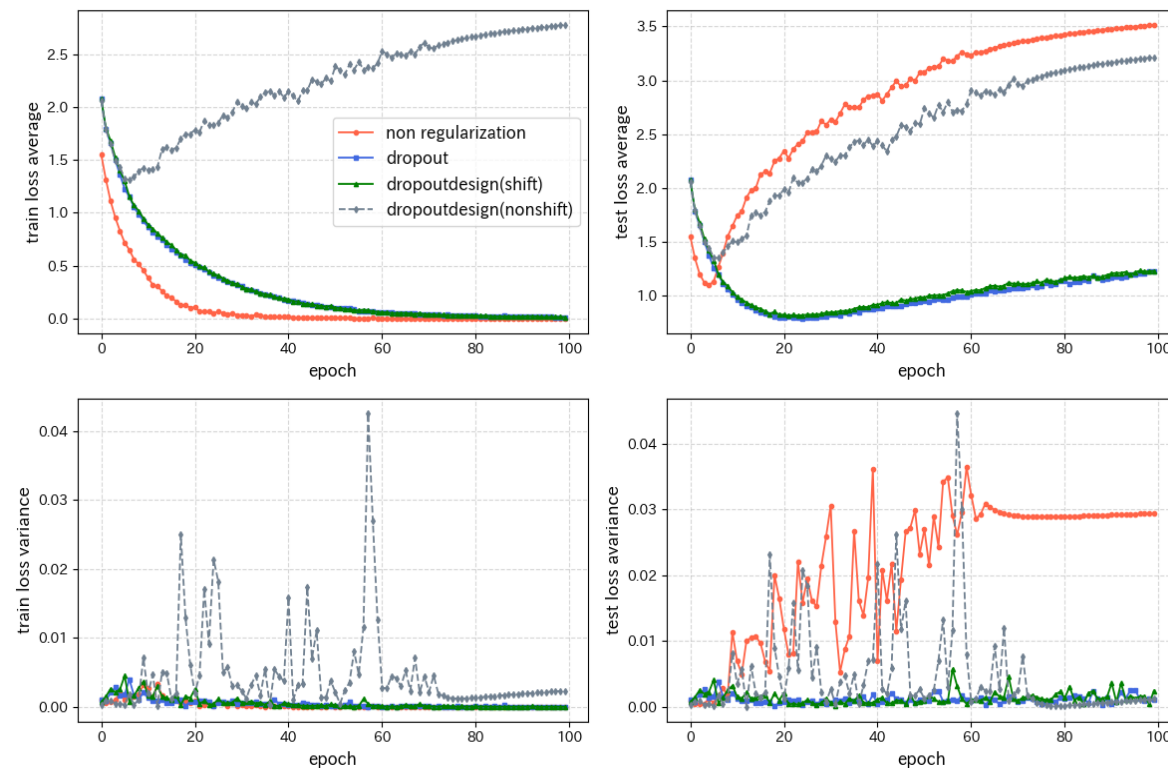


729ノード × 3層 dropout rate=0.67 のニューラルネットワーク

精度



損失



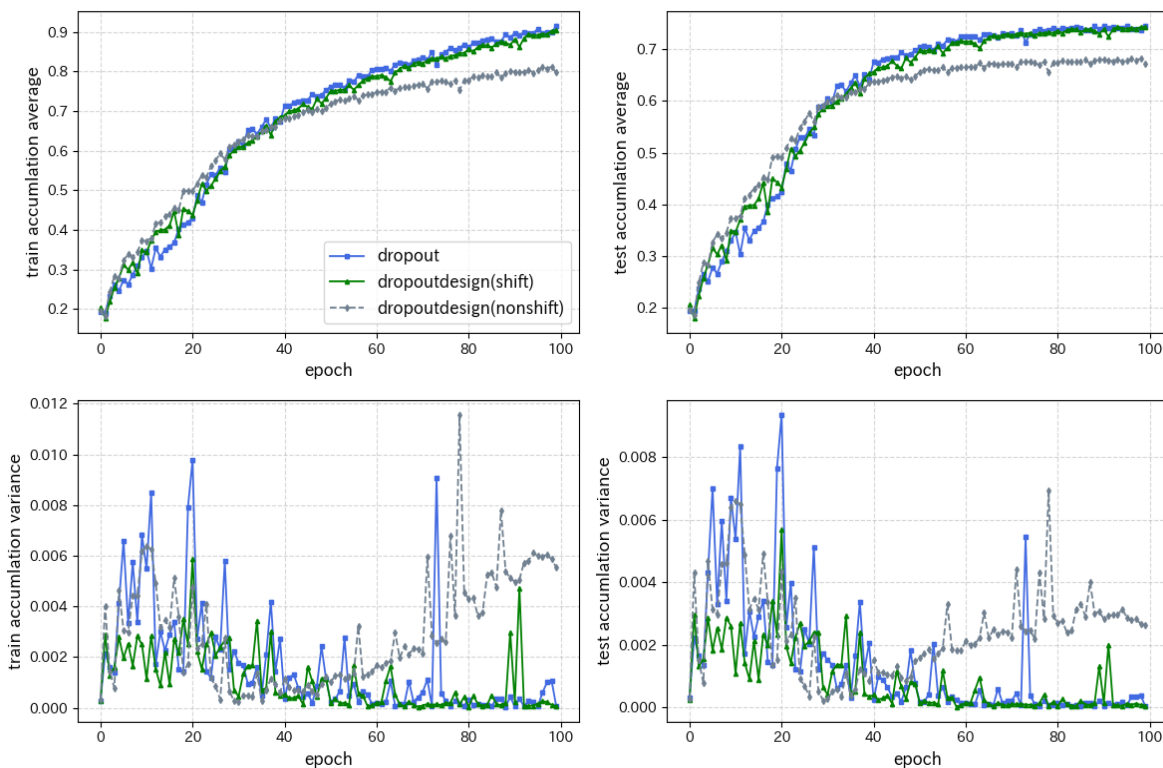
- 正則化なしに比べて、ドロップアウトとドロップアウトデザイン(shift) は過学習をある程度抑制できている
- ドロップアウトとドロップアウトデザイン (shift) は同じような挙動をしている
- ドロップアウトデザイン (nonshift) は正則化なしよりも精度が悪いため、ドロップアウトを行う時にランダム性が仮定されていないとモデルの学習に偏りが生じ、過学習を正確に抑えられていない

- ドロップアウトデザイン特有の挙動は見られなかった
- モデルのハイパパラメータを自由に決められるように操作性の高いドロップアウトデザインを利用していく必要がある
- 既存のドロップアウト法との学習過程の違いをさらに分析していく

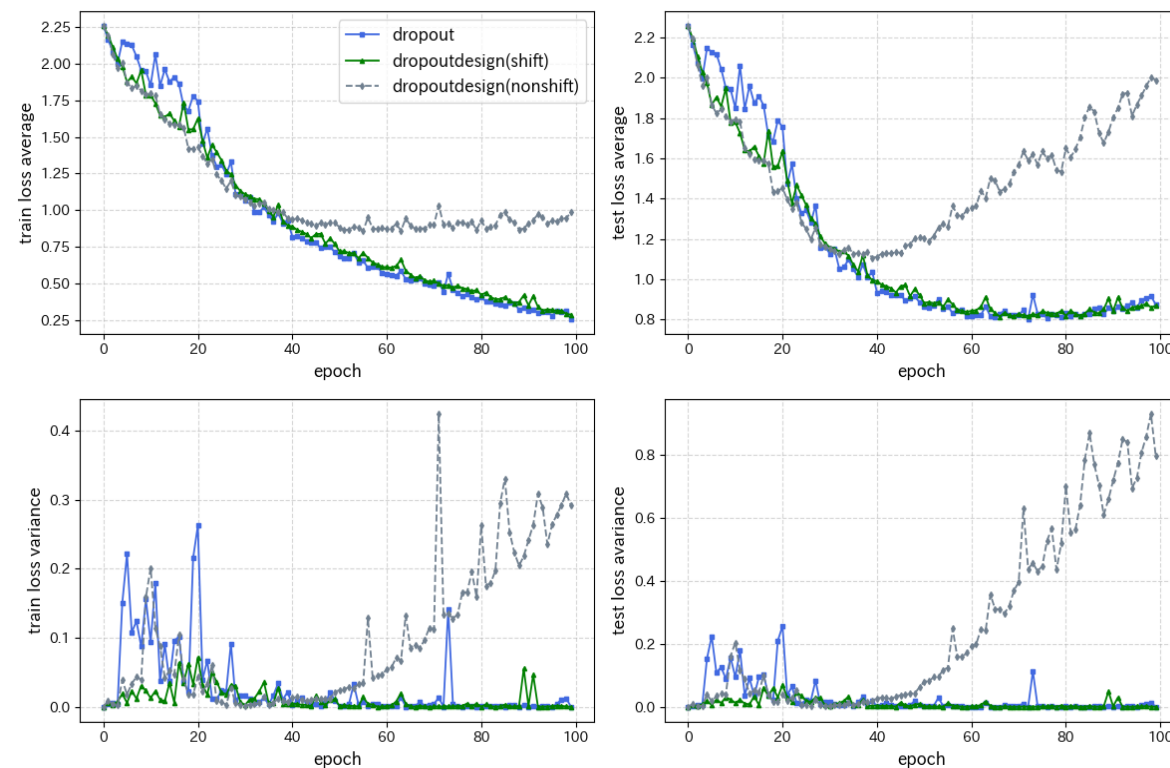
ご清聴ありがとうございました

128ノード × 2層 dropout rate=0.5 のニューラルネットワーク

精度



損失



ブロック集合 B を1エポックの訓練時活性化ノードの集合とみなし、ドロップアウトデザインを参照しながら計画的にドロップアウトを行うノードを選択する

ドロップアウトデザイン	ニューラルネットワーク
各点集合	各層のノード
ブロック集合全体	1エポック全体の活性化ノード集合
i 番目のブロック	i バッチ目の活性化ノード集合
i バッチ目の第 j サブブロック	i バッチ目 j 層目の活性化ノード集合
ブロック集合全体	1エポック全体の活性化ノード集合