Deep neural networks for choice analysis: Enhancing behavioral regularity with gradient regularization

Siqi Feng, Rui Yao, Stephane Hess, Ricardo A. Daziano, Timothy Brathwaite, Joan Walker, Shenhao Wang. (2024). *Transportation Research Part C: Emerging Technologies*, vol.166, 104767.

M1 門坂 佑真

Abstract

- DNNは高い予測性能から近年移動需要モデリングでの応用が進んでいる
- 一方で行動的に不規則なパターンを示すことがある
- そこで需要関数の単調性（≒需要の法則、**行動的規則性**）を評価するための指標として、**strong and weak regularities**を提案する
- この行動的規則性強化のために**6つの勾配正則化項**を用いたフレームワークを設計
- **大サンプルシナリオと小サンプルシナリオ**、**領域内汎化と領域外汎化**における**予測力と行動的規則性のトレードオフ**を検証

Proposed model

Conventional DNN

- high predictive performance
- present behaviorally irregular patterns

＋ **six gradient regularizers**

Verification

- large vs small sample scenarios
- in-domain vs out-of-domain generalizations

Proposed metrics : **strong and weak regularities**

Abstract

新規性

- DNNにおいて需要の法則のような説明変数と選択確率の単調な明示関係を保証するために、事前知識を組み込むことが可能な**勾配正則化**を導入することで行動的規則性を改善
- **strong and weak regularities**という行動的規則性に関する指標を導入

有用性

- sum-baseの勾配正則化は予測性能を犠牲にせずに、行動的規則性を改善
- 小サンプルや領域外汎化における検証では行動的規則性の改善に加え、予測性能も向上
- 勾配正則化という単純な制約により、あらゆるアーキテクチャに導入することが容易であり、深層学習の計算能力と一般的な行動的規則性の過程を統合可能

信頼性

- 2つの旅行調査データに対して、3つのシナリオにおける予測力と行動的規則性のトレードオフを検証
- 5つの評価指標（提案指標含む）でモデルを分析

1.Introduction

2. Literature review

3. Methodology

4. Setup of experiments

5. Results

6. Conclusions

1. Introduction

従来のDiscrete choice modeling (DCM)

- ランダム効用最大化(RUM)に基づく
- **最適な効用関数を見つけるのに時間を要し、その決定も主観的**
- **この探索過程により事前仮定に反する結果を導くモデルは防ぐことができる**

Deep neural network (DNN)

- 自動的な特徴学習が可能 → **探索過程を回避し、主観性を下げることができる**
- 高い予測精度で複雑な行動関係を捉えることができる
- **行動的に不規則なパターンを示す（∵ 非線形構造）**

共通の課題

- **限られたサンプルで勾配方向を決定することが難しい**
- **未知の分布をもつテストデータに学習済みモデルを適用すると、その問題はより悪化する**

解決策

- **strong and weak regularities**という指標を導入
- 勾配の方向を明示的に制約する**勾配正則化を行う制約付き最適化フレームワークを設計**
- サンプルサイズによる性能の違い、領域内汎化と領域外汎化における予測力と行動的規則性の性能を調べる
- ベンチマークとしてMNLとTasteNetsも用いる

1. Introduction

2.Literature review

3. Methodology

4. Setup of experiments

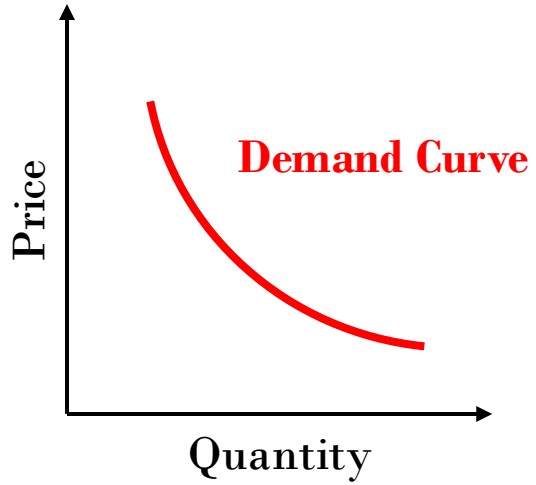
5. Results

6. Conclusions

2. Literature review

需要の法則 : 価格と需要の間に逆相関がある

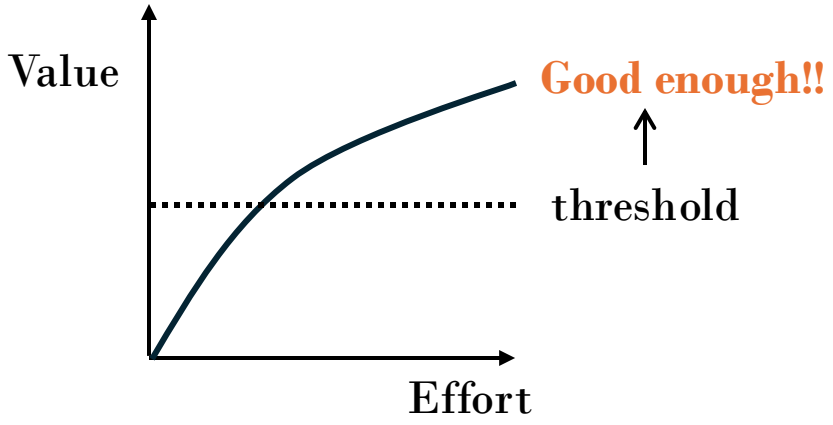
- 消費者の購買力（価格や所得）の変化に伴い、市場の需要が**単調に**変化
- 交通でも移動コストに関して同じことが言える



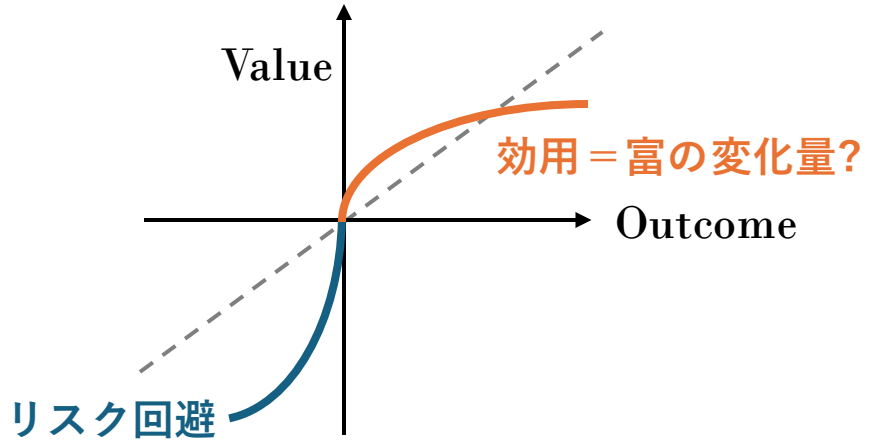
参考 : 厳密な単調性の緩和をする研究

- 限定合理性 : 認識能力の限界等により、限られた合理性しか持ち得ない
- 満足化仮説 : 目的関数（e.g.効用）最大化ではなく、ある水準以上の値の選択肢があれば代案を模索しない
- プロスペクト理論 : 不確実性下において現実の損得と人間の心理的な損得が一致しない

満足化仮説



プロスペクト理論



2. Literature review

RUMにおける需要の法則

- MNLでは、移動コスト増→効用低下→選択確率減少
- 直感に反するパラメータ推定値→問題あり！（e.g.モデルの設計、データ）

↓

DNNにおける需要の法則

- Xia et al. (2023)はコスト増加に対して非単調な需要関数を観測
- 浅いNNは非単調のリスクを減らすが、柔軟性と普遍近似性の低下を伴う

参考：普遍近似性（Universal Approximation Theorem）

- 任意の連続関数は適切な構造を持つNNで任意の精度で近似できる
- DNNではこれにより、少ないノードで複雑な関数を近似しやすい（≒局所的特徴を捉えやすい）

2. Literature review

解決策の例

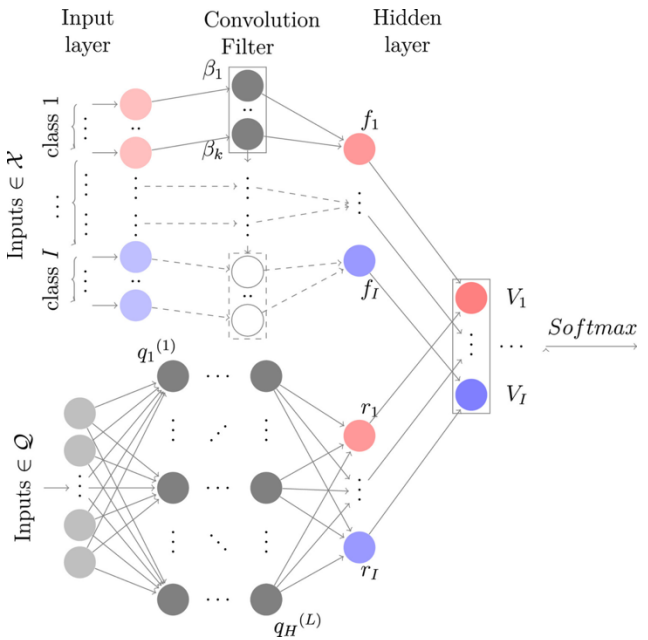
1. ハイブリッドモデル

- Sifringer et al. (2020) の**Learning-MNL**
  - 知識のあるデータはMNLを用い、潜在表現の学習にDNNに頼る手法
  - 理論談話会2024三上さん回
- Han et al. (2022) の**TasteNet**
  - NN部分で異質性を考慮し、MNLのパラメータ値を出力するモデル
  - 理論談話会2023倉澤さん回
- 結局主観的プロセスが必要
- 未だ行動的に不規則な予測が存在 (Wang et al., 2021)

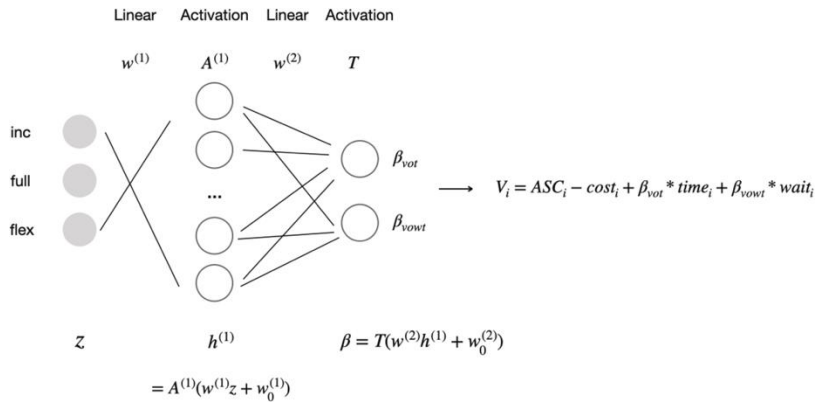
2. モデルアンサンブル

- 複数の学習モデルを組み合わせることでより優れた予測精度を実現する手法
- 同様に不規則なパターンが予測される

Learning-MNL



TasteNet



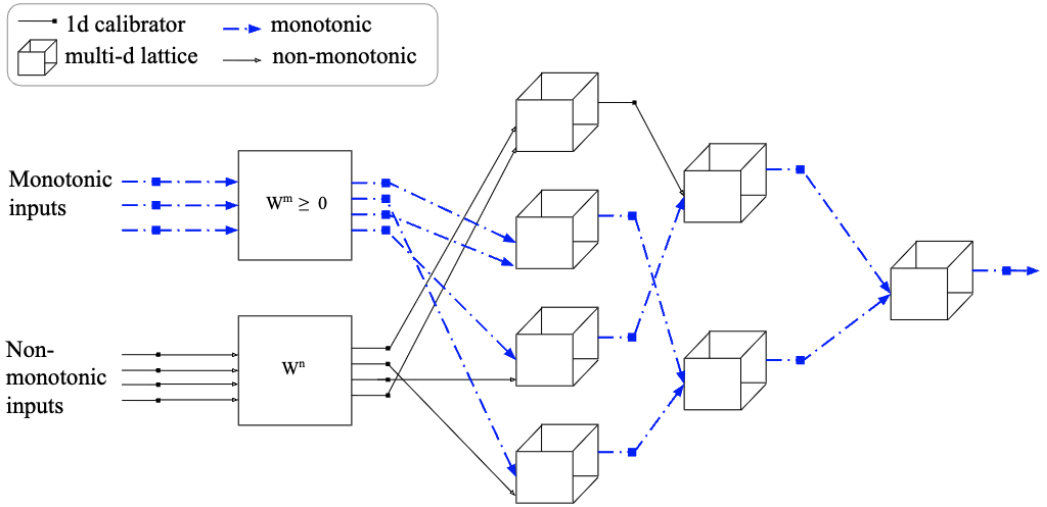
2. Literature review

3. DNNの規則性制約

3-1. モデル構造によるもの

- 非負性制約により隠れ層の重みを制約
- 単調性に反するサンプルの重みを減らす
- You et al. (2017)によるDeep lattice networkの導入

Deep lattice network



3-2. 正則化項によるもの＝制約付き最適化

- 入力変数の仮想ペアに対して単調性の二乗偏差にペナルティを与える
- 単調性に関する事前知識を埋め込んだポイントワイズ損失関数の導入
- 勾配正則化
  - 勾配ノルム正則化：**勾配の大きさ**を小さくするように損失関数にペナルティを加える
  - 勾配方向正則化：**勾配の方向**をあるべき方向と一致させるようにペナルティを加える

1. Introduction

2. Literature review

3.Methodology

4. Setup of experiments

5. Results

6. Conclusions

3-1. DNNs for choice analysis

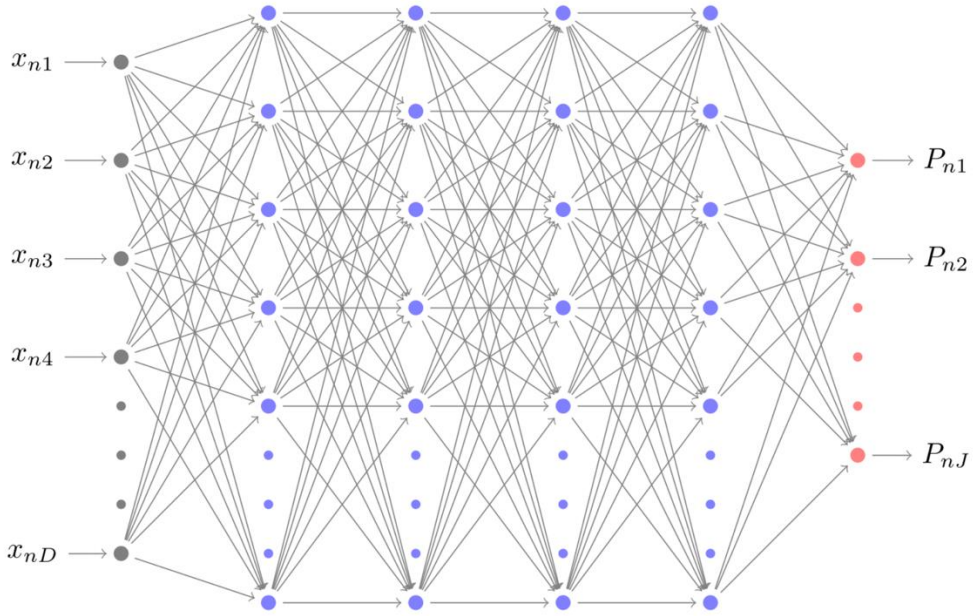
DCM部分の設定

- $\mathbf{x}_n$ ：個人nの説明変数ベクトル
- D ：説明変数の数
- P_{ni}：個人nがJ個の中から選択肢iを選ぶ確率
- $\mathbf{y}_n$ ：個人nの観測された選択ベクトル（$\in \{0,1\}^J$）
- $\mathbf{V}_n$ ：効用ベクトル

DNN部分の設定

- f_h ：h番目のレイヤー
- H ：レイヤーの数
- W_h ：f_hのパラメータ
- $\mathbf{b}_n$ ：バイアスベクトル
- $\varphi(\cdot)$：活性化関数

DNNの構造



- レイヤーf_hでの作業　$f_h(\mathbf{x}_n) = \varphi(W_h\mathbf{x}_n + \mathbf{b}_h)$
- 効用ベクトル　$\mathbf{V}_n = (f_H \circ f_{H-1} \circ \cdots \circ f_2 \circ f_1)(\mathbf{x}_n)$
- 確率　$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}$

3-2. Behavioral regularity metrics

行動的規則性の評価指標

選択肢iのx_dに対して　　$B_{id} = \iint \mathbb{1}\left\{\frac{\partial P_i(\mathbf{x}_z)}{\partial x_d} < \varepsilon\right\} \rho(x_d, z) dx_d dz$

- z　：社会経済属性（年齢、収入、性別等）
- $\mathbf{x}_z$　：属性がzである人の説明変数（個人属性・選択肢のコスト等）
- $P_i(\mathbf{x}_z)$　：$\mathbf{x}_z$に対するiの選択確率
- $\rho(x_d, z)$　：x_dとzの同時分布
- $\mathbb{1}\{\cdot\}$　：$\frac{\partial P_i(\mathbf{x}_z)}{\partial x_d} < \varepsilon$を満たすとき1、そうでないとき0をとる

- モデルと事前知識の間で単調性の一貫性があるかを説明変数による選択確率の偏微分値の符号で測定
- この枠組みにより、$\varepsilon \coloneqq \varepsilon_z$とし、属性グループ固有の閾値を設定することができる

Strong regularity

- $\varepsilon = 0$
- 全ての属性グループでコストが大きくなれば選択確率が下がることを想定

Weak regularity

- $\varepsilon > 0$
- x_dに反応しない属性グループを許容

3-2. Behavioral regularity metrics

行動的規則性の評価指標

近似解 $$B_{id} \approx \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\left\{\frac{\partial P_i(\mathbf{x}_z)}{\partial x_d} < \varepsilon\right\}$$

- Nをサンプルサイズとすると
- 偏微分は有限差分法で計算される
- Glivenko–Cantelliの定理により上の近似解は前ページの厳密解に確率的に収束する

ちなみに従来のMNLであれば…
パラメータが負 → 全ての個人において需要の単調性が成立 → $B_{id} = 1$

3-3. Achieving behavioral regularity by constrained optimization

Unconstrained likelihood maximization

$$\min_W L(W) = \min_W \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{J} -y_{ni} \log P_i(x_n; W)$$

- 従来のRUMならこれでOK
- 最適化は負の推定値をもたらし単調性が満たされる
- DNNではネットワークが深い場合は特に非単調な確率となる可能性があるため不十分

3-3. Achieving behavioral regularity by constrained optimization

Constrained likelihood maximization

1. ハード制約

$$\min_{W} L(W) \quad \text{s.t.} \quad R(x_n; W) \leq 0, \qquad n = 1, \ldots, N$$

- $R(x_n; W)$：行動的規則性の違反度合い（後述）
- 個人レベルでその制約を課すことでB_{id}を向上させる
- 最適化が難しい

2. ソフト制約

$$\min_{W} L(W) + \lambda \sum_{n=1}^{N} R(x_n; W)$$

- λは正則化の強さを調整するハイパーパラメータ
- また行動的規則性の仮定と実際の行動の間の一致度合いに洞察を与える
- 制約に違反する個人を許容する
- 同時に強い行動的規則性を保証するものでないことに注意されたい

3-4. Gradient regularization

$R(x_n; W)$導出の準備

ヤコビアン行列：選択確率の勾配ベクトル

$$\nabla \mathbf{P}(\mathbf{x}_n) = \begin{bmatrix} \frac{\partial P_1}{\partial x_1}(\mathbf{x}_n) & \cdots & \frac{\partial P_1}{\partial x_D}(\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \frac{\partial P_J}{\partial x_1}(\mathbf{x}_n) & \cdots & \frac{\partial P_J}{\partial x_D}(\mathbf{x}_n) \end{bmatrix}$$

- direct derivatives：その選択肢に関する説明変数で偏微分
- cross derivatives：他の選択肢に関する説明変数で偏微分
- sociodemographic derivatives：選択者の社会経済属性で偏微分

ソフト制約のおかげでどのタイプにも勾配制約を課せる

勾配制約を課すためのマスク行列

$$\Psi(\mathbf{x}_n) = \begin{bmatrix} \mathbb{1}\{\frac{\partial P_1}{\partial x_1}(\mathbf{x}_n) \notin \mathbb{S}_{11}\} & \cdots & \mathbb{1}\{\frac{\partial P_1}{\partial x_D}(\mathbf{x}_n) \notin \mathbb{S}_{1D}\} \\ \vdots & \ddots & \vdots \\ \mathbb{1}\{\frac{\partial P_J}{\partial x_1}(\mathbf{x}_n) \notin \mathbb{S}_{J1}\} & \cdots & \mathbb{1}\{\frac{\partial P_J}{\partial x_D}(\mathbf{x}_n) \notin \mathbb{S}_{JD}\} \end{bmatrix}$$

$\mathbb{1}\{\cdot\}$：$\frac{\partial P_i}{\partial x_d}(\mathbf{x}_n) \notin \mathbb{S}_{id}$を満たすとき1、そうでないとき0をとる

$\mathbb{S}_{id}$：事前知識より期待されている符号（$\in \{\mathbb{R}^-, \mathbb{R}^+, \mathbb{R}\}$）

3-4. Gradient regularization

sum-based gradient regulization (sum-XGR) : 勾配の方向を正則

目的 : 説明変数と確率の自明な単調関係を保証する

$$R_\sigma(\mathbf{x}_n) = \langle \Psi(\mathbf{x}_n), \nabla \mathbf{P}(\mathbf{x}_n) \rangle_F$$

- 微分値の符号に関する異なる事前仮定を複数許容
- 数値実験ではdirect derivativesに負の制約を課すのみとする

norm-based gradient regulization (norm-XGR) : 勾配の大きさを正則

目的 : 小さなコストの摂動で急激に確率が変化しないようにする（平滑性）

$$R_\nu(\mathbf{x}_n) = \|\nabla \mathbf{P}(\mathbf{x}_n)\|_F^2 = \langle \nabla \mathbf{P}(\mathbf{x}_n), \nabla \mathbf{P}(\mathbf{x}_n) \rangle_F$$

- コンピュータサイエンスでよくある手法
- 数値実験では正則化手法のベンチマークとする
- 行動的規則性の信念ではなく数学的観点の仮定に基づく

3-4. Gradient regularization

Probability gradient regularizers (PGRs)

前述

Utility gradient regularizers (UGRs)

- 効用が高くなれば選択確率も単調に増える
- $\mathbf{P}(\mathbf{x}_n)$を$\mathbf{V}(\mathbf{x}_n)$に置き換えることができる

Log-likelihood gradient regularizers (LGRs)

- ある個人のある選択肢の尤度 $l_i(\mathbf{x}_n) = y_{ni} \log P_i(\mathbf{x}_n)$
- 対数変換は単調性を保持する
- を$\mathbf{l}(\mathbf{x}_n)$に置き換えることができる

6種類の勾配正則化項を設計：sum-PGR, sum-UGR, sum-LGR, norm-PGR, norm-UGR, norm-LGR

1. Introduction

2. Literature review

3. Methodology

4.Setup of experiments

5. Results

6. Conclusions

4-1. Datasets

Chicago Metropolitan Agency for Planning (CMAP)

- 自動車、徒歩、電車、自転車による26,099件のトリップデータ
- 自動車が全トリップの70％を占める
- 自転車はわずかのため自転車と徒歩は統合しActive modeとする
- 連続的選択肢属性×2（所要時間、自動車と電車のコスト）
- 離散的社会属性×3（年齢、世帯規模、車の数）
- 社会属性のダミー変数×5（高学歴、男性、一人世帯、一台世帯、高所得世帯）

London Travel Demand Survey (LTDS)

- 自動車、徒歩、公共交通機関、自転車による81,086件のトリップデータ
- CMAPほど自動車が極端に多くない
- 自転車はわずかのため自転車と徒歩は統合しActive modeとする
- 連続的選択肢属性×2（所要時間、自動車と公共交通機関のコスト）
- 離散的社会属性×2（車の数、公共交通機関の乗り換え回数）
- 社会属性のダミー変数×5（若者、高齢者、男性、運転免許、一台世帯）

4-1. Datasets

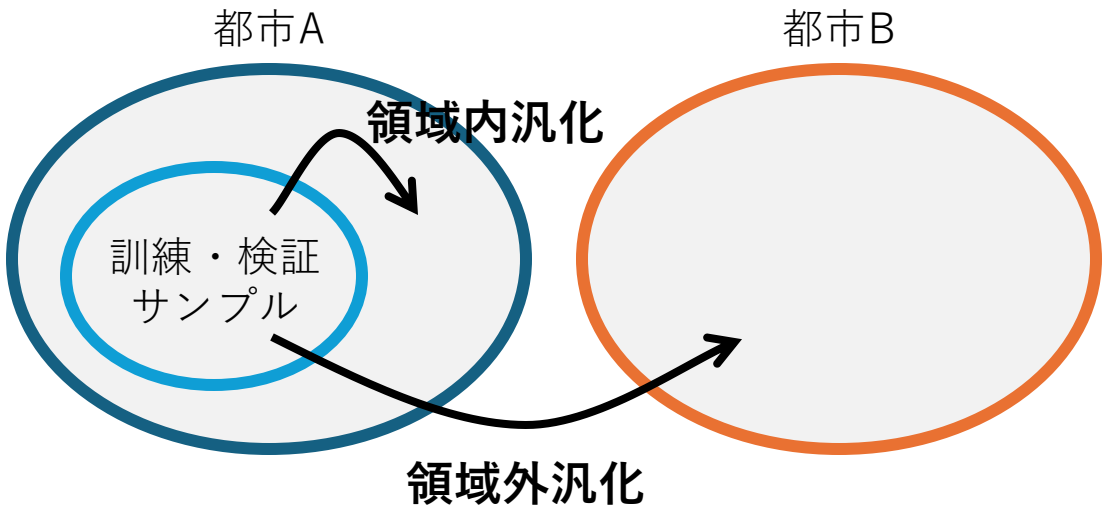
領域内汎化性（サンプル外汎化性）と領域外汎化性を検証したい！

10K-Random

- 10,000トリップをランダムにサンプリング
- 70%を訓練用、10%を検証用、20%をテスト用
- サンプルサイズとしても理想的なシナリオ

1K-Random

- 1,000トリップをランダムにサンプリング
- 80%を訓練用、20%を検証用
- テスト用に500トリップをランダムにサンプリング
- 限られたサンプルしか利用できないシナリオを想定



10K-Soreted

- 10,000トリップをランダムにサンプリングし、自動車のコストでソート
- 下位80%を訓練・検証用、上位20%をテスト用
- 高価（≒遠出）な旅行で実施されるテストを想定し、領域外汎化性を論じる

4-2. Experimental design

モデルの訓練

- 訓練データを用いる
- 10バッチに分割
- 学習率は10^{-3}
- 収束を確実なものにさせるため検証データでの損失が最適になるまで訓練

ハイパーパラメータの探索

- 検証データを用いる
- 4つの隠れ層と1層あたり100個のニューロンとする
- 勾配正則化の最適なλを$(10^{-4}, 100)$の範囲で探索

モデルの性能評価

- テストデータにより、10回のモデル反復の結果を平均したアンサンブル性能を分析する
- 対数尤度、accuracy、F1 score、**strong & weak regularity**で評価
- 前3つは予測力をとらえ、後2つは行動的規則性をとらえる
- 強い行動的規則性はεを小さな負の値とし、弱い行動的規則性は小さな正の値に設定（区別化のため）

4-2. Experimental design

機械学習の性能指標

二値問題において

	あり（Positive）	なし（Negative）
Positiveと予測	**TP**	**FP**
Negativeと予測	**FN**	**TN**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$ ：全体のうちどれだけ正解しているか

$$\text{F1 score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$ ：RecallとPrecisionのトレードオフを考える

$$\begin{cases} \text{Recall} = \dfrac{TP}{TP + FN} & \text{：「あり」をどれほど回収できているか} \\ \text{Precision} = \dfrac{TP}{TP + FP} & \text{：「あり」と予測したものの中での正答率} \end{cases}$$

4-3. Models

比較するモデル：MNL, DNN, TasteNets（後2つにXGRをつける）

MNL：行動的規則性を示すためのベンチマーク

$$\text{Driving}: V_{n1} = \beta_{t1} t_{n1} + \beta_{c1} c_{n1}$$
$$\text{Public transit}: V_{n2} = \alpha_2 + \boldsymbol{\gamma}_2 \mathbf{z}_n + \beta_{t2} t_{n2} + \beta_{c2} c_{n2}$$
$$\text{Active mode}: V_{n3} = \alpha_3 + \boldsymbol{\gamma}_3 \mathbf{z}_n + \beta_{t3} t_{n3}$$

t_{ni}：各選択肢の旅行時間
c_{ni}：各選択肢の旅行コスト
$\mathbf{w} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$：推定するパラメータ

TasteNets

$$V_{ni} = T(\mathbf{z}_n; W)^{\mathrm{T}} \mathbf{x}_{ni} = \beta_{ni}^{\mathrm{T}} \mathbf{x}_{ni}$$

T：NNのアーキテクチャ
W：重み
β_{ni}：個人に固有の選好パラメータベクトル

ここでb_{ni}が旅行コストに対するパラメータであるとすると、
$-\mathrm{ReLU}(-b_{ni})$や$-\exp(-b_{ni})$とすることで、行動的規則性を保証できる

$\Vert$

ハード制約($\because b_{ni} > 0$となることを許さない)

1. Introduction

2. Literature review

3. Methodology

4. Setup of experiments

5.Results

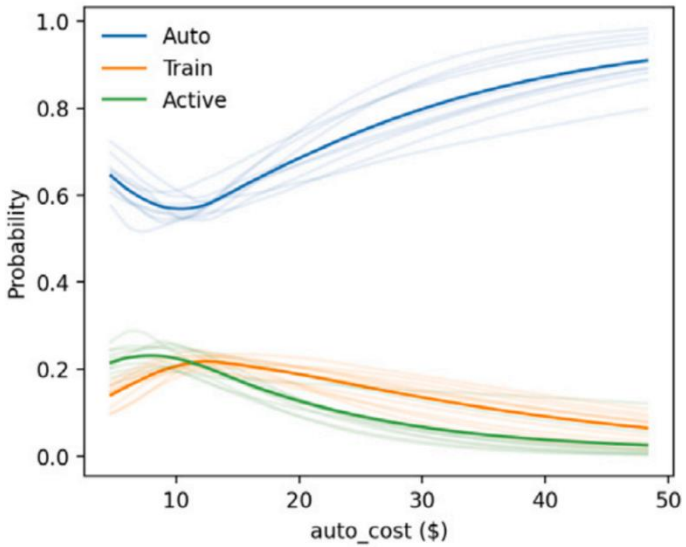
6. Conclusions

5-1. Enhancing model performance with gradient regularization

1. Large sample scenario (10K-Random)

Metric:	DNN	TasteNet	RUM
Mean (SD)	No GR	No GR	MNL
Panel 1: CMAP data, sum-XGR			
Log-likelihood	−1351.9	−1438.3	−1426.3
	(4.697)	(5.834)	(0)
Accuracy	0.729	0.713	0.718
	(0.003)	(0.003)	(0)
F_1 score	0.691	0.654	0.669
	(0.005)	(0.006)	(0)
Strong regularity	0.888	0.998	0.998
	(0.066)	(0.002)	(0)
Weak regularity	0.922	0.999	**1.000**
	(0.061)	(0.002)	(0)
Panel 3: LTDS data, sum-XGR			
Log-likelihood	−1292.0	−1305.8	−1366.9
	(7.894)	(2.226)	(0)
Accuracy	0.729	0.732	0.730
	(0.005)	(0.003)	(0)
F_1 score	0.727	0.728	0.726
	(0.004)	(0.003)	(0)
Strong regularity	0.950	0.942	0.993
	(0.034)	(0.021)	(0)
Weak regularity	0.969	0.966	**1.000**
	(0.027)	(0.019)	(0)

- 予測力：DNN (no GR) > MNL
  規則性：DNN (no GR) < MNL
- 尤度は予測性能に敏感
- accuracy&F1 scoreの差はデータの偏りを表現
- TasteNetは規則性は良いが予測力で劣る



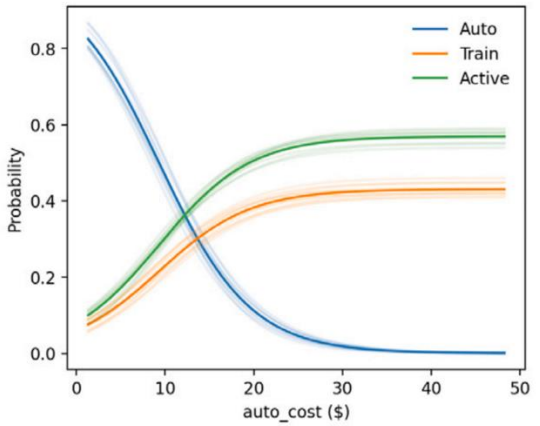
(a) DNN

5-1. Enhancing model performance with gradient regularization

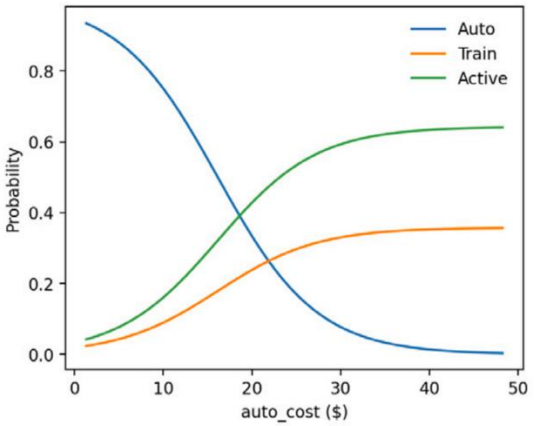
1. Large sample scenario (10K-Random)

Metric:	DNN				TasteNet						RUM
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR	ReLU	Exp	MNL
Panel 1: CMAP data, sum-XGR											
Log-likelihood	−1351.9	**−1344.3**	−1350.2	−1347.4	−1438.3	−1438.4	−1439.0	−1438.4	−1463.0	−1439.7	−1426.3
	(4.697)	(5.521)	(5.803)	(5.110)	(5.834)	(6.027)	(5.992)	(6.099)	(20.05)	(6.073)	(0)
Accuracy	0.729	**0.730**	0.729	0.729	0.713	0.713	0.713	0.713	0.712	0.717	0.718
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0)
F_1 score	0.691	**0.698**	0.694	0.696	0.654	0.654	0.654	0.654	0.650	0.664	0.669
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.002)	(0.004)	(0)
Strong regularity	0.888	0.990	0.982	0.991	0.998	**0.999**	**0.999**	**0.999**	0.426	**0.999**	0.998
	(0.066)	(0.003)	(0.012)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.320)	(0.001)	(0)
Weak regularity	0.922	0.999	0.996	0.999	0.999	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.061)	(0.001)	(0.006)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0)

- 予測力：DNN (soft) > TasteNet (soft / hard)
- hard constrainはweak regularityは必ず満たすが、strong regularityはExpしか満たさない
- MNLとTasteNet (Exp)は指標が似通っている



(e) TasteNet, Exp



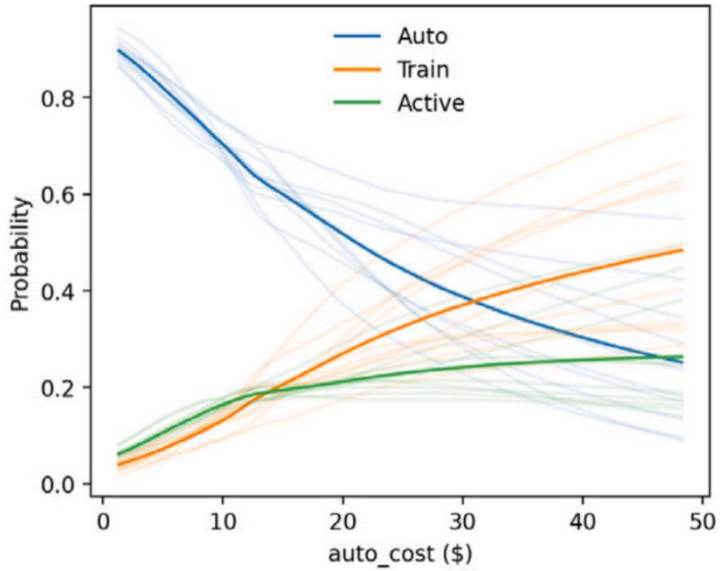
(f) MNL

5-1. Enhancing model performance with gradient regularization

1. Large sample scenario (10K-Random)

Metric:	DNN				TasteNet			
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR
Panel 1: CMAP data, sum-XGR								
Log-likelihood	−1351.9	**−1344.3**	−1350.2	−1347.4	−1438.3	−1438.4	−1439.0	−1438.4
	(4.697)	(5.521)	(5.803)	(5.110)	(5.834)	(6.027)	(5.992)	(6.099)
Accuracy	0.729	**0.730**	0.729	0.729	0.713	0.713	0.713	0.713
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
F_1 score	0.691	**0.698**	0.694	0.696	0.654	0.654	0.654	0.654
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)
Strong regularity	0.888	0.990	0.982	0.991	0.998	**0.999**	**0.999**	**0.999**
	(0.066)	(0.003)	(0.012)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
Weak regularity	0.922	0.999	0.996	0.999	0.999	**1.000**	**1.000**	**1.000**
	(0.061)	(0.001)	(0.006)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Panel 2: CMAP data, norm-XGR								
Log-likelihood	**−1351.9**	−1353.9	−1362.0	−1354.0	−1438.3	−1439.1	−1469.5	−1440.8
	(4.697)	(5.018)	(3.110)	(4.451)	(5.834)	(5.836)	(6.287)	(5.804)
Accuracy	**0.729**	**0.729**	0.725	0.727	0.713	0.713	0.710	0.712
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
F_1 score	**0.691**	0.688	0.677	0.683	0.654	0.653	0.643	0.652
	(0.005)	(0.007)	(0.009)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)
Strong regularity	0.888	0.857	0.706	0.815	0.998	0.998	0.929	0.997
	(0.066)	(0.069)	(0.051)	(0.068)	(0.002)	(0.002)	(0.030)	(0.004)
Weak regularity	0.922	0.893	0.756	0.851	0.999	0.999	0.941	0.998
	(0.061)	(0.067)	(0.051)	(0.069)	(0.002)	(0.002)	(0.026)	(0.003)
Panel 3: LTDS data, sum-XGR								
Log-likelihood	−1292.0	**−1288.6**	**−1288.6**	−1305.0	−1305.8	−1308.6	−1317.6	−1308.3
	(7.894)	(9.668)	(3.765)	(7.316)	(2.226)	(2.795)	(4.280)	(2.670)
Accuracy	0.729	0.729	0.727	0.724	0.732	0.730	0.728	0.730
	(0.005)	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)
F_1 score	0.727	0.728	0.725	0.721	0.728	0.726	0.723	0.726
	(0.004)	(0.004)	(0.002)	(0.006)	(0.003)	(0.002)	(0.004)	(0.002)
Strong regularity	0.950	0.994	0.994	**0.997**	0.942	0.964	0.987	0.972
	(0.034)	(0.004)	(0.009)	(0.003)	(0.021)	(0.013)	(0.008)	(0.012)
Weak regularity	0.969	0.999	0.998	**1.000**	0.966	**1.000**	**1.000**	**1.000**
	(0.027)	(0.001)	(0.004)	(0.000)	(0.019)	(0.000)	(0.000)	(0.000)

- sum-XGRは予測力犠牲にせずに規則性向上
- DNNにおいては予測力向上も見られる
- sum-PGR > sum-UGR, sum-LGR



(b) DNN, sum-PGR

5-1. Enhancing model performance with gradient regularization

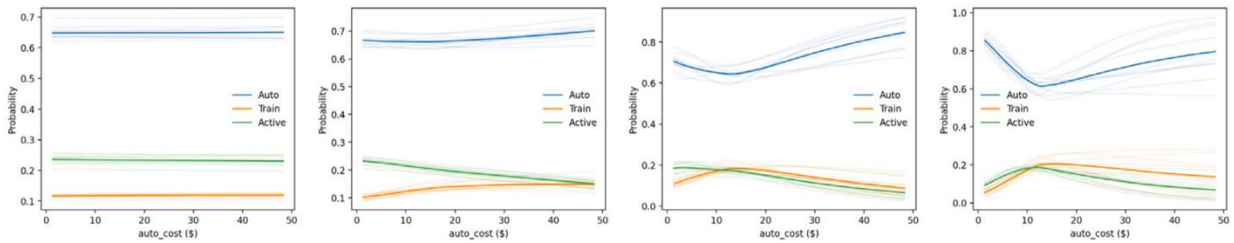
1. Large sample scenario (10K-Random)

Metric:	DNN				TasteNet			
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR
Panel 1: CMAP data, sum-XGR								
Log-likelihood	−1351.9	**−1344.3**	−1350.2	−1347.4	−1438.3	−1438.4	−1439.0	−1438.4
	(4.697)	(5.521)	(5.803)	(5.110)	(5.834)	(6.027)	(5.992)	(6.099)
Accuracy	0.729	**0.730**	0.729	0.729	0.713	0.713	0.713	0.713
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
F_1 score	0.691	**0.698**	0.694	0.696	0.654	0.654	0.654	0.654
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)
Strong regularity	0.888	0.990	0.982	0.991	0.998	**0.999**	**0.999**	**0.999**
	(0.066)	(0.003)	(0.012)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
Weak regularity	0.922	0.999	0.996	0.999	0.999	**1.000**	**1.000**	**1.000**
	(0.061)	(0.001)	(0.006)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Panel 2: CMAP data, norm-XGR								
Log-likelihood	**−1351.9**	−1353.9	−1362.0	−1354.0	−1438.3	−1439.1	−1469.5	−1440.8
	(4.697)	(5.018)	(3.110)	(4.451)	(5.834)	(5.836)	(6.287)	(5.804)
Accuracy	**0.729**	**0.729**	0.725	0.727	0.713	0.713	0.710	0.712
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
F_1 score	**0.691**	0.688	0.677	0.683	0.654	0.653	0.643	0.652
	(0.005)	(0.007)	(0.009)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)
Strong regularity	0.888	0.857	0.706	0.815	0.998	0.998	0.929	0.997
	(0.066)	(0.069)	(0.051)	(0.068)	(0.002)	(0.002)	(0.030)	(0.004)
Weak regularity	0.922	0.893	0.756	0.851	0.999	0.999	0.941	0.998
	(0.061)	(0.067)	(0.051)	(0.069)	(0.002)	(0.002)	(0.026)	(0.003)
Panel 3: LTDS data, sum-XGR								
Log-likelihood	−1292.0	**−1288.6**	**−1288.6**	−1305.0	−1305.8	−1308.6	−1317.6	−1308.3
	(7.894)	(9.668)	(3.765)	(7.316)	(2.226)	(2.795)	(4.280)	(2.670)
Accuracy	0.729	0.729	0.727	0.724	0.732	0.730	0.728	0.730
	(0.005)	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)
F_1 score	0.727	0.728	0.725	0.721	0.728	0.726	0.723	0.726
	(0.004)	(0.004)	(0.002)	(0.006)	(0.003)	(0.002)	(0.004)	(0.002)
Strong regularity	0.950	0.994	0.994	**0.997**	0.942	0.964	0.987	0.972
	(0.034)	(0.004)	(0.009)	(0.003)	(0.021)	(0.013)	(0.008)	(0.012)
Weak regularity	0.969	0.999	0.998	**1.000**	0.966	**1.000**	**1.000**	**1.000**
	(0.027)	(0.001)	(0.004)	(0.000)	(0.019)	(0.000)	(0.000)	(0.000)

- norm-XGRは予測力も規則性も向上なし
- 下図より読み取れるのは
  - 価格に対して単調でない
  - λが大きくなるにつれて曲線が平坦化する

λ

大 ← → 小



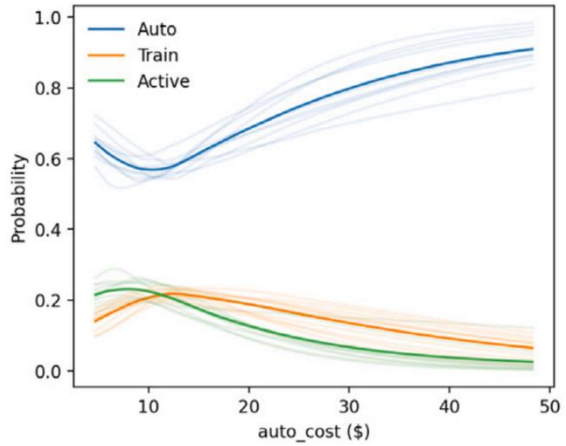
(e) $\lambda = 100$, norm-PGR (f) $\lambda = 1$, norm-PGR (g) $\lambda = 10^{-2}$, norm-PGR (h) $\lambda = 10^{-4}$, norm-PGR

norm-XGRは有効ではない
→次の比較からは省く

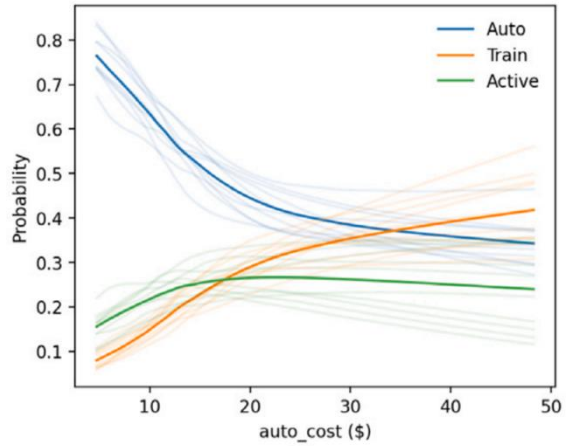
5-1. Enhancing model performance with gradient regularization

2. Small sample scenario (1K-Random)

Metric:	DNN				TasteNet						RUM
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR	ReLU	Exp	MNL
Panel 1: CMAP data, sum-XGR											
Log-likelihood	−375.1	**−367.9**	−374.2	−369.7	−389.3	−385.2	−383.1	−386.6	−386.3	−378.0	−380.2
	(3.096)	(4.181)	(3.298)	(3.304)	(2.838)	(3.351)	(3.139)	(3.447)	(2.506)	(4.123)	(0)
Accuracy	0.705	0.703	0.676	0.696	0.700	0.679	0.686	0.677	0.696	0.707	**0.718**
	(0.002)	(0.007)	(0.010)	(0.011)	(0.005)	(0.012)	(0.010)	(0.012)	(0.005)	(0.004)	(0)
F_1 score	0.648	0.645	0.559	0.619	0.619	0.563	0.580	0.559	0.610	0.637	**0.665**
	(0.004)	(0.018)	(0.027)	(0.030)	(0.012)	(0.028)	(0.023)	(0.028)	(0.011)	(0.010)	(0)
Strong regularity	0.664	0.985	0.985	0.985	0.659	0.979	0.983	0.979	0.461	**0.999**	0.996
	(0.173)	(0.009)	(0.011)	(0.011)	(0.129)	(0.015)	(0.013)	(0.013)	(0.195)	(0.001)	(0)
Weak regularity	0.728	0.999	0.997	0.999	0.685	0.997	0.995	0.992	**1.000**	**1.000**	**1.000**
	(0.162)	(0.001)	(0.005)	(0.002)	(0.128)	(0.004)	(0.006)	(0.009)	(0.000)	(0.000)	(0)
Panel 2: LTDS data, sum-XGR											
Log-likelihood	−322.6	−328.8	**−317.8**	−335.2	−345.1	−339.7	−343.5	−340.1	−341.0	−335.5	−331.2
	(3.455)	(8.146)	(3.440)	(11.30)	(2.337)	(2.023)	(3.035)	(1.974)	(2.572)	(2.951)	(0)
Accuracy	0.737	0.720	0.740	0.717	0.725	0.734	0.724	0.733	0.729	0.733	**0.746**
	(0.007)	(0.013)	(0.007)	(0.021)	(0.004)	(0.006)	(0.009)	(0.006)	(0.006)	(0.008)	(0)
F_1 score	0.732	0.703	0.734	0.694	0.711	0.720	0.705	0.718	0.715	0.725	**0.740**
	(0.007)	(0.020)	(0.008)	(0.038)	(0.005)	(0.006)	(0.013)	(0.005)	(0.007)	(0.009)	(0)
Strong regularity	0.904	**0.999**	0.992	0.997	0.939	0.964	**0.999**	0.965	0.924	**0.999**	0.998
	(0.069)	(0.002)	(0.014)	(0.008)	(0.023)	(0.013)	(0.002)	(0.01)	(0.046)	(0.002)	(0)
Weak regularity	0.913	**1.000**	0.993	0.998	0.943	0.984	0.999	0.981	**1.000**	**1.000**	**1.000**
	(0.067)	(0.001)	(0.014)	(0.006)	(0.024)	(0.012)	(0.002)	(0.013)	(0.000)	(0.000)	(0)



(a) DNN



(b) DNN, sum-PGR

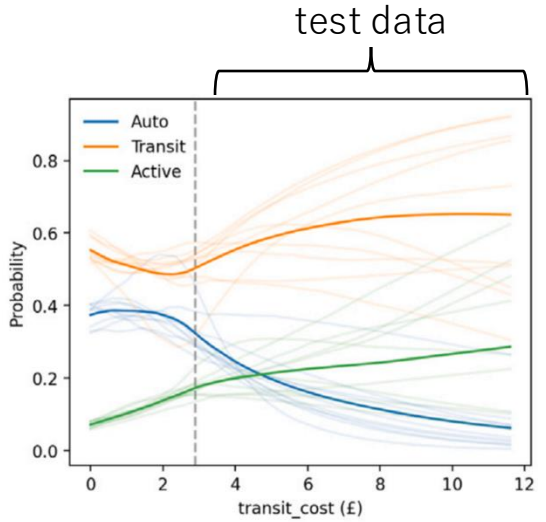
- 特にCMAPデータに対するregularityがDNN (no GR)はかなり小さい
- 向上幅を見ると大サンプルのときよりもGRが効果的と言える
- sum-XGRは予測力と規則性、両方の向上に寄与している

sum-XGR DNNは小サンプルではMNLと同等の総合的な性能を持つ

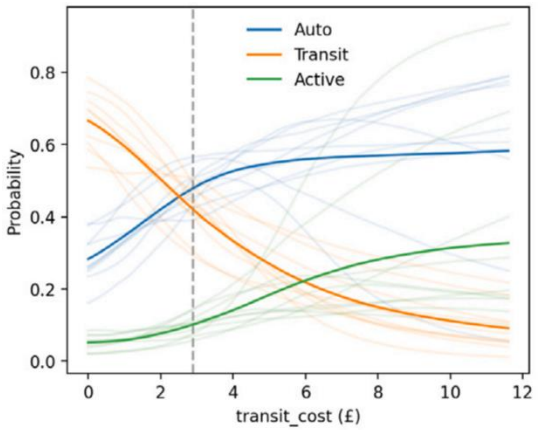
5-1. Enhancing model performance with gradient regularization

3. Out-of-domain generalization (10K-Sorted)

Metric:	DNN				TasteNet						RUM
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR	ReLU	Exp	MNL
Panel 1: CMAP data, sum-XGR											
Log-likelihood	−1356.1	−1243.5	**−1167.9**	−1232.4	−1485.9	−1873.7	−1901.4	−2003.9	−2315.4	−1995.0	−2025.7
	(134.1)	(55.15)	(27.64)	(54.87)	(47.92)	(45.52)	(36.62)	(50.23)	(395.4)	(51.83)	(0)
Accuracy	0.783	0.788	**0.789**	0.788	0.750	0.726	0.725	0.705	0.578	0.723	0.722
	(0.011)	(0.002)	(0.003)	(0.002)	(0.014)	(0.008)	(0.006)	(0.008)	(0.134)	(0.007)	(0)
F_1 score	0.722	0.726	**0.727**	0.724	0.707	0.721	0.720	0.708	0.603	0.719	0.721
	(0.010)	(0.006)	(0.003)	(0.005)	(0.006)	(0.005)	(0.004)	(0.006)	(0.112)	(0.005)	(0)
Strong regularity	0.317	0.857	0.923	0.865	**1.000**	0.980	0.978	0.981	0.933	0.969	0.984
	(0.240)	(0.099)	(0.087)	(0.071)	(0.000)	(0.004)	(0.004)	(0.004)	(0.169)	(0.005)	(0)
Weak regularity	0.487	0.974	0.983	0.977	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.230)	(0.037)	(0.040)	(0.023)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0)
Panel 2: LTDS data	sum-XGR (public transit cost)										
Log-likelihood	−1137.9	−1110.4	**−1108.4**	−1112.1	−1221.1	−1170.2	−1171.1	−1175.0	−1182.3	−1150.1	−1301.6
	(34.33)	(23.86)	(21.15)	(26.67)	(33.10)	(27.01)	(28.54)	(30.25)	(54.11)	(29.62)	(0)
Accuracy	0.777	0.784	**0.794**	0.782	0.776	0.786	0.785	0.783	0.782	0.781	0.780
	(0.010)	(0.007)	(0.007)	(0.006)	(0.009)	(0.008)	(0.008)	(0.007)	(0.017)	(0.008)	(0)
F_1 score	0.768	0.776	**0.778**	0.776	0.765	0.772	0.772	0.772	0.769	0.768	0.770
	(0.011)	(0.006)	(0.008)	(0.006)	(0.008)	(0.008)	(0.008)	(0.007)	(0.018)	(0.008)	(0)
Strong regularity	0.185	0.968	0.979	0.907	0.313	0.878	0.872	0.720	0.248	**0.982**	0.980
	(0.075)	(0.032)	(0.029)	(0.062)	(0.007)	(0.095)	(0.095)	(0.083)	(0.084)	(0.005)	(0)
Weak regularity	0.207	0.977	0.984	0.925	0.343	0.993	0.988	0.900	**1.000**	**1.000**	**1.000**
	(0.080)	(0.026)	(0.025)	(0.055)	(0.008)	(0.010)	(0.012)	(0.045)	(0.000)	(0.000)	(0)



(a) DNN



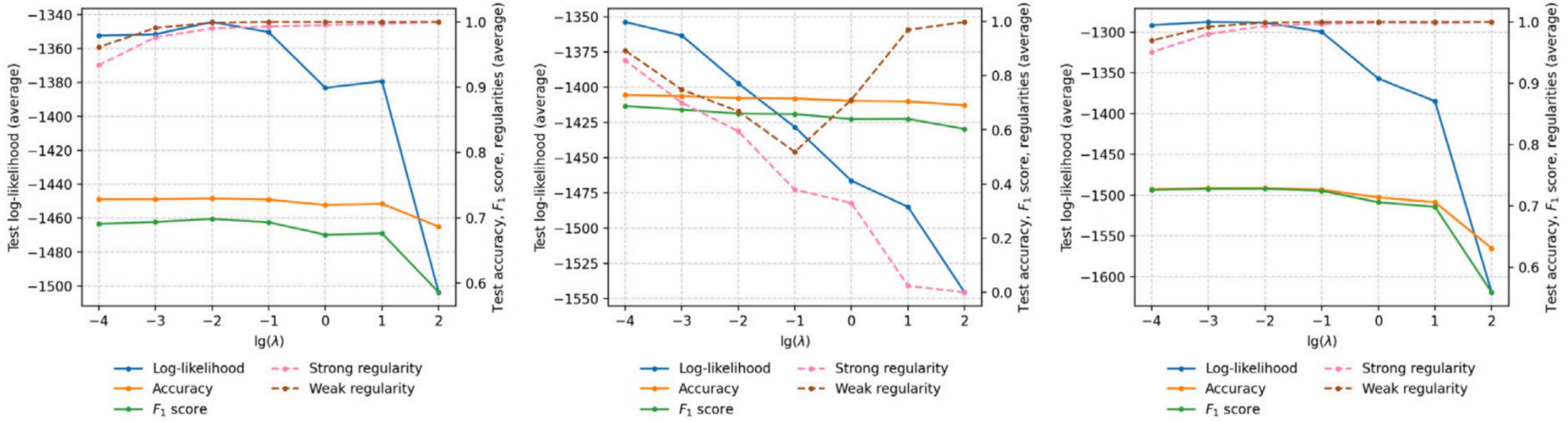
(b) DNN, sum-UGR

- 尤度でDNN (no GR)はMNLを大きく上回っているが、regularityは劣る
- sum-XGRは予測力と規則性の向上に寄与している
- TasteNet (hard)はLTDSでは規則性と予測力に向上が見られるが、CMAPでは効果なし

領域外汎化性でもsum-XGRは大きな効果を持つ / ハード制約の効果はケース依存

5-2. Trade-off between predictive power and behavioral regularity

1. Large sample scenario (10K-Random)



(a) DNN, sum-PGR (CMAP)　(b) DNN, norm-PGR (CMAP)　(c) DNN, sum-PGR (LTDS)

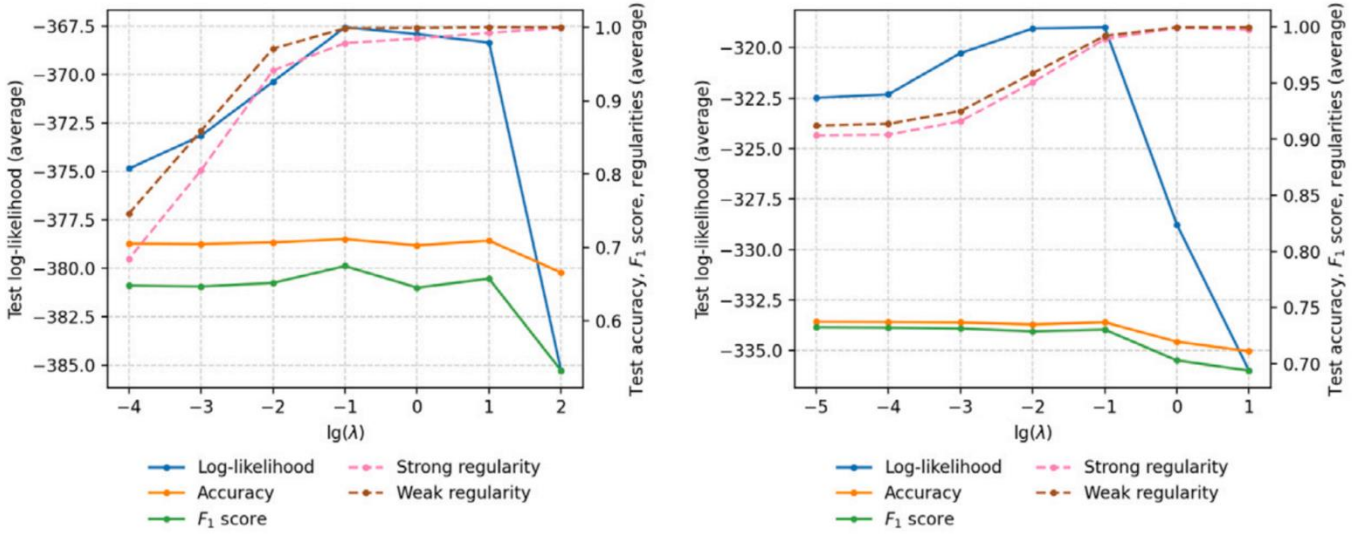
- λが大きくなるにつれて尤度は大きく低下しており、accuracyやF1 scoreよりも敏感
- 特に$\lambda = 10$の点でその変化が顕著である→ $\lambda < 0.1$では予測力の大きな低下を招かずに規則性向上を実現
- norm-PGRにおいて、λを大きくしたときweak regularityは1に近づき、strong regularityは0に近づく

→ノルム正則化により、勾配が平坦化されていることがわかる

大サンプルでは規則性の向上に予測力低下を伴う代替効果（substitution effects）が見られる
λを大きくしすぎないという工夫が必要

5-2. Trade-off between predictive power and behavioral regularity

2. Small sample scenario (1K-Random)



(a) DNN, sum-PGR (CMAP)

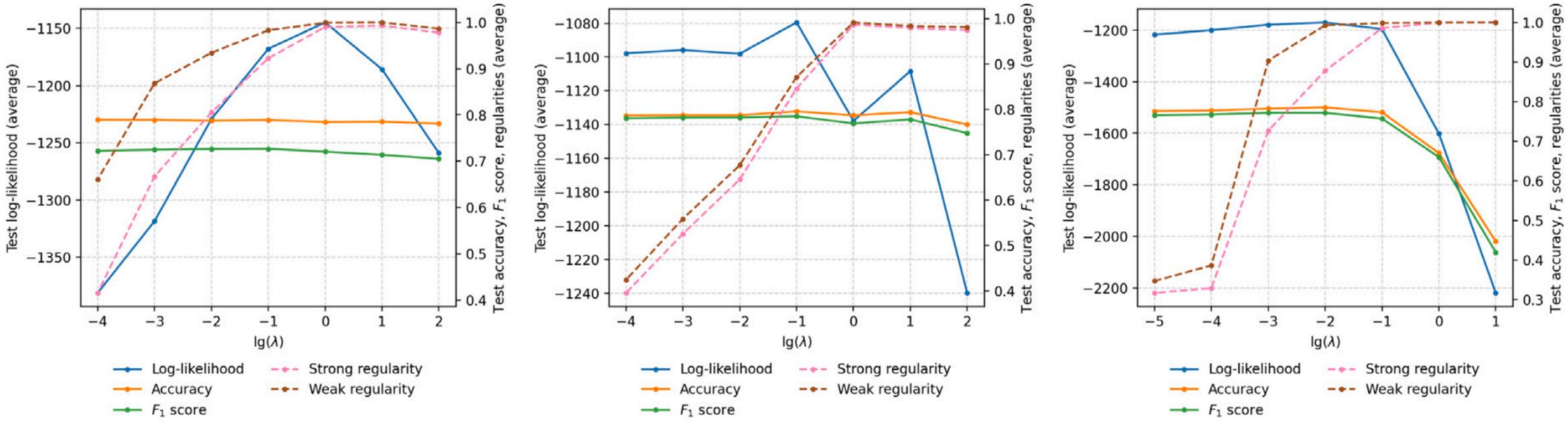
(b) DNN, sum-PGR (LTDS)

- どちらのデータセットでもDNN (sum-PGR)の予測力と規則性の両方を向上させているλの範囲が存在
  （例）CMAPにおいてλを10^{-4}から0.1に増やしたところ尤度は1.9%、strong regularityは29.4%改善
- DNNにおけるパレート効率の存在を示唆している
- 一方でλを大きくし過ぎると代替効果が表れる

小サンプルでは予測力と規則性が同時に向上する相補効果（complementary effects）が見られる
ここでもλを大きくしすぎないという工夫が必要

5-2. Trade-off between predictive power and behavioral regularity

3. Out-of-domain generalization (10K-Sorted)



(a) DNN, sum-UGR (CMAP)　(b) DNN, sum-UGR (LTDS)　(c) TasteNet, sum-PGR (LTDS)

- DNNではCMAPにおいて相補効果、LTDSにおいて代替効果が顕著

→NNのトレーニングのデータ依存性を示唆

- TasteNetではLTDSにおいて代替効果がより顕著（特にaccuracyとF1 scoreにおいて）

領域外汎化では大サンプルと小サンプルで見られた代替効果と相補効果の両方が見られる
領域外汎化においても勾配正則化が有用である

5-3. Summary of empirical findings

1. Large sample scenario

- DNN (no GR)でも高い行動的規則性を示す
- サンプルサイズが十分に大きい場合は提案指標や正則化は必要ないかもしれない

2. Small sample

- 正則化により行動的規則性だけでなく、予測力も向上させる

3. Out-of-domain generalization

- 勾配正則化がモデルの都市間転移にも有用であることが示された

総じて…

- sum-XGRが最も効果的
- 提案正則化は規則性と予測力の両方に寄与する可能性がある
- ベンチマークの時点で規則性が高い場合には提案勾配正則化が有益とは言えない

1. Introduction

2. Literature review

3. Methodology

4. Setup of experiments

5. Results

6.Conclusions

6. Conclusions

目的

- DNNにおいて予測力を保持したまま行動的規則性を改善する
- 行動的規則性の指標についてのコンセンサスを得る

手法

- 行動的規則性の指標
  - strong and weak behavioral regularities
- 制約付き最適化における正則化のフレームワーク
  - 6つの勾配正則化
  - 2つのデータセットにおいて、大サンプル・小サンプル・領域外汎化性の3つのシナリオで検証

結果

- sum-XGRが全てのサンプルにおいて予測力を犠牲にすることなく規則性を改善
- 大サンプル：代替効果、小サンプル：相補効果
- 領域外汎化においても勾配正則化が有用

貢献

- 単純な正則化により深層学習の計算能力と一般的な行動的規則性の過程を統合可能に
- データ分布が異なる都市間モデル転移にも信用性を担保

6. Conclusions

Future work

- 個人の行動的規則性を考慮し、ハード制約によりそれを強化
  - 本研究では集約的な指標を規則性として導入し、それをソフト制約により強化していた
  - 「個人vs集合」「ソフト制約vsハード制約」というのは経済学や最適化の分野では長い間議論されている
- cross partial derivatives(＝他の選択肢に関する説明変数で偏微分)に正則化
  - 本研究はdirect partial derivativesのみ
  - cross partial derivativesはより強い事前過程を課し、代替効果や相補効果を制御可能
- ハイパーパラメータチューニングを組み合わせることでλを自動的に学習
- DNNにおけるパラメータ同定（parameter identification）の議論
  - モデルの不確実性の定量化や統計的検定の設計が困難
  - 特にパラメータ数が過剰に多いover-parameterized modelにおいての検証が必要

所感

- 小サンプルにおいてMNLと結局同等の性能なのは少し残念
- 結果パートのまとめ方が若干分かりずらい...
- 提案指標を改善するために正則化されている感は否めない

- 論文中にもあったが単純な正則化法であるためにあらゆる枠組みに簡単に組み込むことができる
  - より複雑なモデルにおいて良さを発揮しそう
  - 機械学習が流行りの研究室
  - 卒修論ゼミとかでこの論文を思い出してもらえると嬉しい
- サンプルの少なさという普遍的な行動モデルの問題に答えているのが良い
- 領域外汎化の可能性についてはもっといろんな論文で議論されるべき
  - 全都市が同じような分布のデータを持つとは限らない
  - いろんな設定に使える手法というのも良いモデルの1つの評価基準
  - 近藤さん修論の外部検証

Appendix

1. Large sample scenario (10K-Random)

Metric:	DNN				TasteNet						RUM
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR	ReLU	Exp	MNL
Panel 1: CMAP data, sum-XGR											
Log-likelihood	−1351.9	**−1344.3**	−1350.2	−1347.4	−1438.3	−1438.4	−1439.0	−1438.4	−1463.0	−1439.7	−1426.3
	(4.697)	(5.521)	(5.803)	(5.110)	(5.834)	(6.027)	(5.992)	(6.099)	(20.05)	(6.073)	(0)
Accuracy	0.729	**0.730**	0.729	0.729	0.713	0.713	0.713	0.713	0.712	0.717	0.718
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0)
F_1 score	0.691	**0.698**	0.694	0.696	0.654	0.654	0.654	0.654	0.650	0.664	0.669
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.002)	(0.004)	(0)
Strong regularity	0.888	0.990	0.982	0.991	0.998	**0.999**	**0.999**	**0.999**	0.426	**0.999**	0.998
	(0.066)	(0.003)	(0.012)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.320)	(0.001)	(0)
Weak regularity	0.922	0.999	0.996	0.999	0.999	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.061)	(0.001)	(0.006)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0)
Panel 2: CMAP data, norm-XGR											
Log-likelihood	**−1351.9**	−1353.9	−1362.0	−1354.0	−1438.3	−1439.1	−1469.5	−1440.8	−1463.0	−1439.7	−1426.3
	(4.697)	(5.018)	(3.110)	(4.451)	(5.834)	(5.836)	(6.287)	(5.804)	(20.05)	(6.073)	(0)
Accuracy	**0.729**	**0.729**	0.725	0.727	0.713	0.713	0.710	0.712	0.712	0.717	0.718
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0)
F_1 score	**0.691**	0.688	0.677	0.683	0.654	0.653	0.643	0.652	0.650	0.664	0.669
	(0.005)	(0.007)	(0.009)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.002)	(0.004)	(0)
Strong regularity	0.888	0.857	0.706	0.815	0.998	0.998	0.929	0.997	0.426	**0.999**	0.998
	(0.066)	(0.069)	(0.051)	(0.068)	(0.002)	(0.002)	(0.030)	(0.004)	(0.320)	(0.001)	(0)
Weak regularity	0.922	0.893	0.756	0.851	0.999	0.999	0.941	0.998	**1.000**	**1.000**	**1.000**
	(0.061)	(0.067)	(0.051)	(0.069)	(0.002)	(0.002)	(0.026)	(0.003)	(0.000)	(0.000)	(0)
Panel 3: LTDS data, sum-XGR											
Log-likelihood	−1292.0	**−1288.6**	**−1288.6**	−1305.0	−1305.8	−1308.6	−1317.6	−1308.3	−1335.5	−1312.9	−1366.9
	(7.894)	(9.668)	(3.765)	(7.316)	(2.226)	(2.795)	(4.280)	(2.670)	(20.45)	(2.706)	(0)
Accuracy	0.729	0.729	0.727	0.724	0.732	0.730	0.728	0.730	0.725	**0.736**	0.730
	(0.005)	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)	(0.006)	(0.002)	(0)
F_1 score	0.727	0.728	0.725	0.721	0.728	0.726	0.723	0.726	0.721	**0.733**	0.726
	(0.004)	(0.004)	(0.002)	(0.006)	(0.003)	(0.002)	(0.004)	(0.002)	(0.006)	(0.002)	(0)
Strong regularity	0.950	0.994	0.994	**0.997**	0.942	0.964	0.987	0.972	0.933	0.990	0.993
	(0.034)	(0.004)	(0.009)	(0.003)	(0.021)	(0.013)	(0.008)	(0.012)	(0.043)	(0.002)	(0)
Weak regularity	0.969	0.999	0.998	**1.000**	0.966	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.027)	(0.001)	(0.004)	(0.000)	(0.019)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0)

Appendix

2. Small sample scenario (1K-Random)

Metric:	DNN				TasteNet						RUM
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR	ReLU	Exp	MNL
Panel 1: CMAP data, sum-XGR											
Log-likelihood	−375.1	**−367.9**	−374.2	−369.7	−389.3	−385.2	−383.1	−386.6	−386.3	−378.0	−380.2
	(3.096)	(4.181)	(3.298)	(3.304)	(2.838)	(3.351)	(3.139)	(3.447)	(2.506)	(4.123)	(0)
Accuracy	0.705	0.703	0.676	0.696	0.700	0.679	0.686	0.677	0.696	0.707	**0.718**
	(0.002)	(0.007)	(0.010)	(0.011)	(0.005)	(0.012)	(0.010)	(0.012)	(0.005)	(0.004)	(0)
F_1 score	0.648	0.645	0.559	0.619	0.619	0.563	0.580	0.559	0.610	0.637	**0.665**
	(0.004)	(0.018)	(0.027)	(0.030)	(0.012)	(0.028)	(0.023)	(0.028)	(0.011)	(0.010)	(0)
Strong regularity	0.664	0.985	0.985	0.985	0.659	0.979	0.983	0.979	0.461	**0.999**	0.996
	(0.173)	(0.009)	(0.011)	(0.011)	(0.129)	(0.015)	(0.013)	(0.013)	(0.195)	(0.001)	(0)
Weak regularity	0.728	0.999	0.997	0.999	0.685	0.997	0.995	0.992	**1.000**	**1.000**	**1.000**
	(0.162)	(0.001)	(0.005)	(0.002)	(0.128)	(0.004)	(0.006)	(0.009)	(0.000)	(0.000)	(0)
Panel 2: LTDS data, sum-XGR											
Log-likelihood	−322.6	−328.8	**−317.8**	−335.2	−345.1	−339.7	−343.5	−340.1	−341.0	−335.5	−331.2
	(3.455)	(8.146)	(3.440)	(11.30)	(2.337)	(2.023)	(3.035)	(1.974)	(2.572)	(2.951)	(0)
Accuracy	0.737	0.720	0.740	0.717	0.725	0.734	0.724	0.733	0.729	0.733	**0.746**
	(0.007)	(0.013)	(0.007)	(0.021)	(0.004)	(0.006)	(0.009)	(0.006)	(0.006)	(0.008)	(0)
F_1 score	0.732	0.703	0.734	0.694	0.711	0.720	0.705	0.718	0.715	0.725	**0.740**
	(0.007)	(0.020)	(0.008)	(0.038)	(0.005)	(0.006)	(0.013)	(0.005)	(0.007)	(0.009)	(0)
Strong regularity	0.904	**0.999**	0.992	0.997	0.939	0.964	**0.999**	0.965	0.924	**0.999**	0.998
	(0.069)	(0.002)	(0.014)	(0.008)	(0.023)	(0.013)	(0.002)	(0.01)	(0.046)	(0.002)	(0)
Weak regularity	0.913	**1.000**	0.993	0.998	0.943	0.984	0.999	0.981	**1.000**	**1.000**	**1.000**
	(0.067)	(0.001)	(0.014)	(0.006)	(0.024)	(0.012)	(0.002)	(0.013)	(0.000)	(0.000)	(0)

Appendix

3. Out-of-domain generalization (10K-Sorted)

Metric:	DNN				TasteNet						RUM
Mean (SD)	No GR	PGR	UGR	LGR	No GR	PGR	UGR	LGR	ReLU	Exp	MNL
Panel 1: CMAP data, sum-XGR											
Log-likelihood	−1356.1	−1243.5	**−1167.9**	−1232.4	−1485.9	−1873.7	−1901.4	−2003.9	−2315.4	−1995.0	−2025.7
	(134.1)	(55.15)	(27.64)	(54.87)	(47.92)	(45.52)	(36.62)	(50.23)	(395.4)	(51.83)	(0)
Accuracy	0.783	0.788	**0.789**	0.788	0.750	0.726	0.725	0.705	0.578	0.723	0.722
	(0.011)	(0.002)	(0.003)	(0.002)	(0.014)	(0.008)	(0.006)	(0.008)	(0.134)	(0.007)	(0)
F_1 score	0.722	0.726	**0.727**	0.724	0.707	0.721	0.720	0.708	0.603	0.719	0.721
	(0.010)	(0.006)	(0.003)	(0.005)	(0.006)	(0.005)	(0.004)	(0.006)	(0.112)	(0.005)	(0)
Strong regularity	0.317	0.857	0.923	0.865	**1.000**	0.980	0.978	0.981	0.933	0.969	0.984
	(0.240)	(0.099)	(0.087)	(0.071)	(0.000)	(0.004)	(0.004)	(0.004)	(0.169)	(0.005)	(0)
Weak regularity	0.487	0.974	0.983	0.977	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.230)	(0.037)	(0.040)	(0.023)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0)
Panel 2: LTDS data, sum-XGR (public transit cost)											
Log-likelihood	−1137.9	−1110.4	**−1108.4**	−1112.1	−1221.1	−1170.2	−1171.1	−1175.0	−1182.3	−1150.1	−1301.6
	(34.33)	(23.86)	(21.15)	(26.67)	(33.10)	(27.01)	(28.54)	(30.25)	(54.11)	(29.62)	(0)
Accuracy	0.777	0.784	**0.794**	0.782	0.776	0.786	0.785	0.783	0.782	0.781	0.780
	(0.010)	(0.007)	(0.007)	(0.006)	(0.009)	(0.008)	(0.008)	(0.007)	(0.017)	(0.008)	(0)
F_1 score	0.768	0.776	**0.778**	0.776	0.765	0.772	0.772	0.772	0.769	0.768	0.770
	(0.011)	(0.006)	(0.008)	(0.006)	(0.008)	(0.008)	(0.008)	(0.007)	(0.018)	(0.008)	(0)
Strong regularity	0.185	0.968	0.979	0.907	0.313	0.878	0.872	0.720	0.248	**0.982**	0.980
	(0.075)	(0.032)	(0.029)	(0.062)	(0.007)	(0.095)	(0.095)	(0.083)	(0.084)	(0.005)	(0)
Weak regularity	0.207	0.977	0.984	0.925	0.343	0.993	0.988	0.900	**1.000**	**1.000**	**1.000**
	(0.080)	(0.026)	(0.025)	(0.055)	(0.008)	(0.010)	(0.012)	(0.045)	(0.000)	(0.000)	(0)