Gaussian Process Latent Class Choice Models

Georges Sfeir, Machine Learning for Smart Mobility (MLSM), Technical University of Denmark (DTU)
Filipe Rodrigues, Machine Learning for Smart Mobility (MLSM), Technical University of Denmark (DTU)
Maya Abou-Zeid, American University of Beirut (AUB)

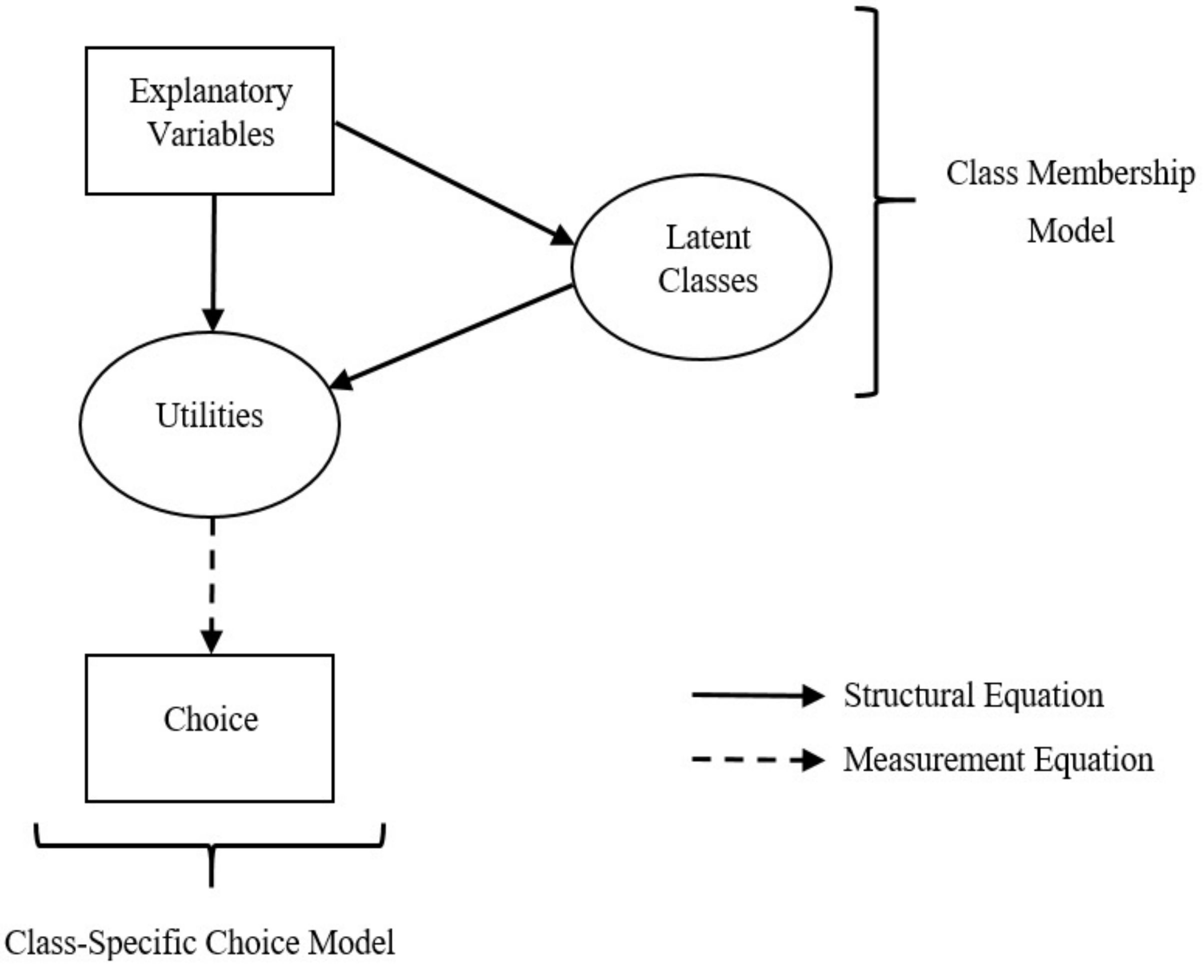
The 7th International Choice Modelling Conference 2022

Outline

- Introduction
- Modelling Framework
- Applications
- Conclusion

Introduction

Latent Class Choice Model (LCCM)



Adapted from Walker and Li (2007)

- ✓ Latent classes are easily interpreted

$$U_{nk} = S'_n\gamma_k + \nu_{nk}$$

Utility of latent class k for decision-maker n

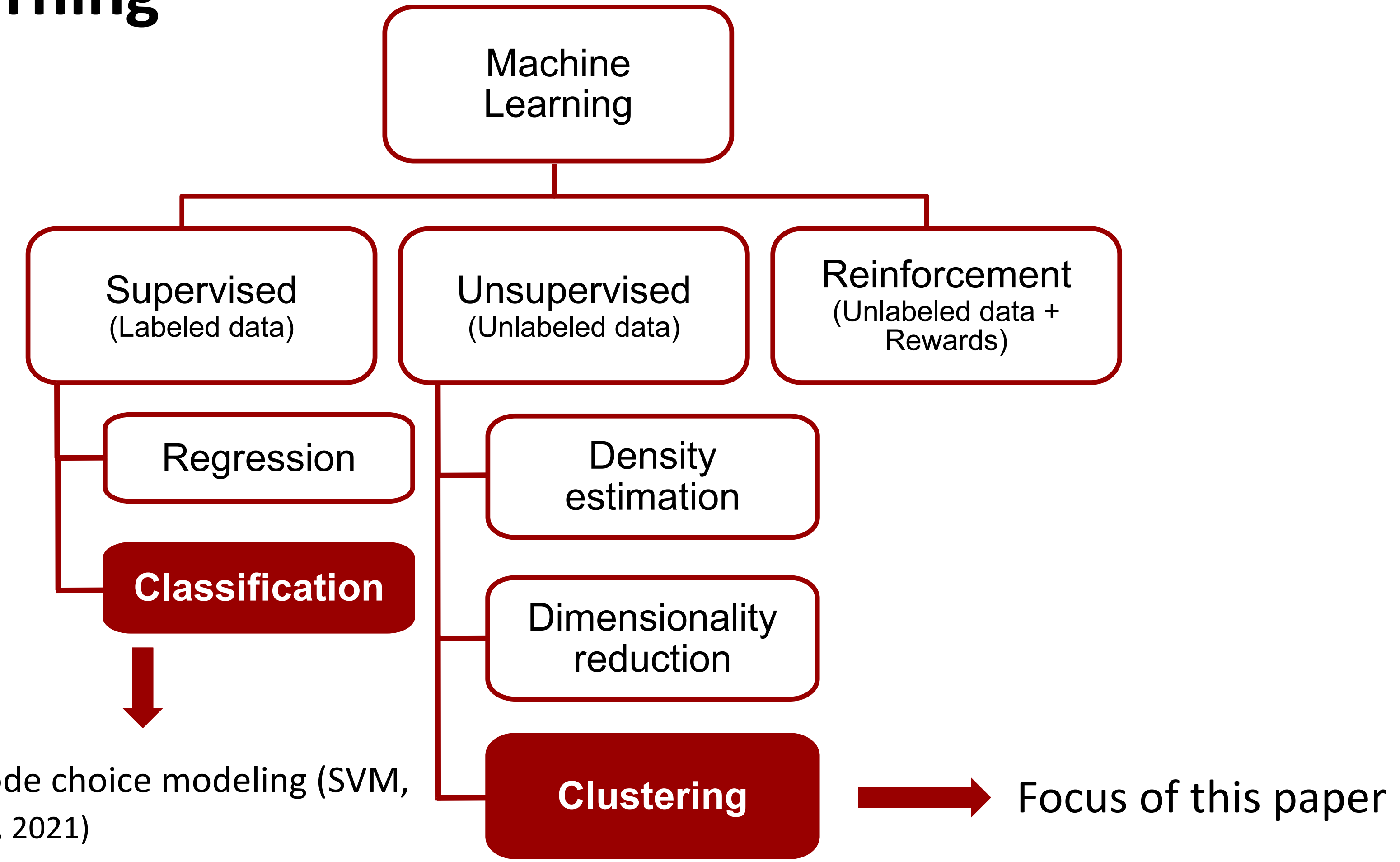
Characteristics of decision-maker n

Vector of unknown parameters

Disturbance term

- ✓ Fewer assumptions regarding the parameters' distributions forms
- ✗ Computational complexity grows as the number of classes increases
- ✗ Linear-in-parameters specification of the latent classes
- ✗ Discrete representation may oversimplify the heterogeneity

Machine Learning



- Most ML applications in mode choice modeling (SVM, ANN, Trees, etc.) (Hillel et al., 2021)
- ML usually provided better prediction accuracy than DCMs
- Not connected to economic theories → Lack of straightforward interpretability

Combining the Two Fields

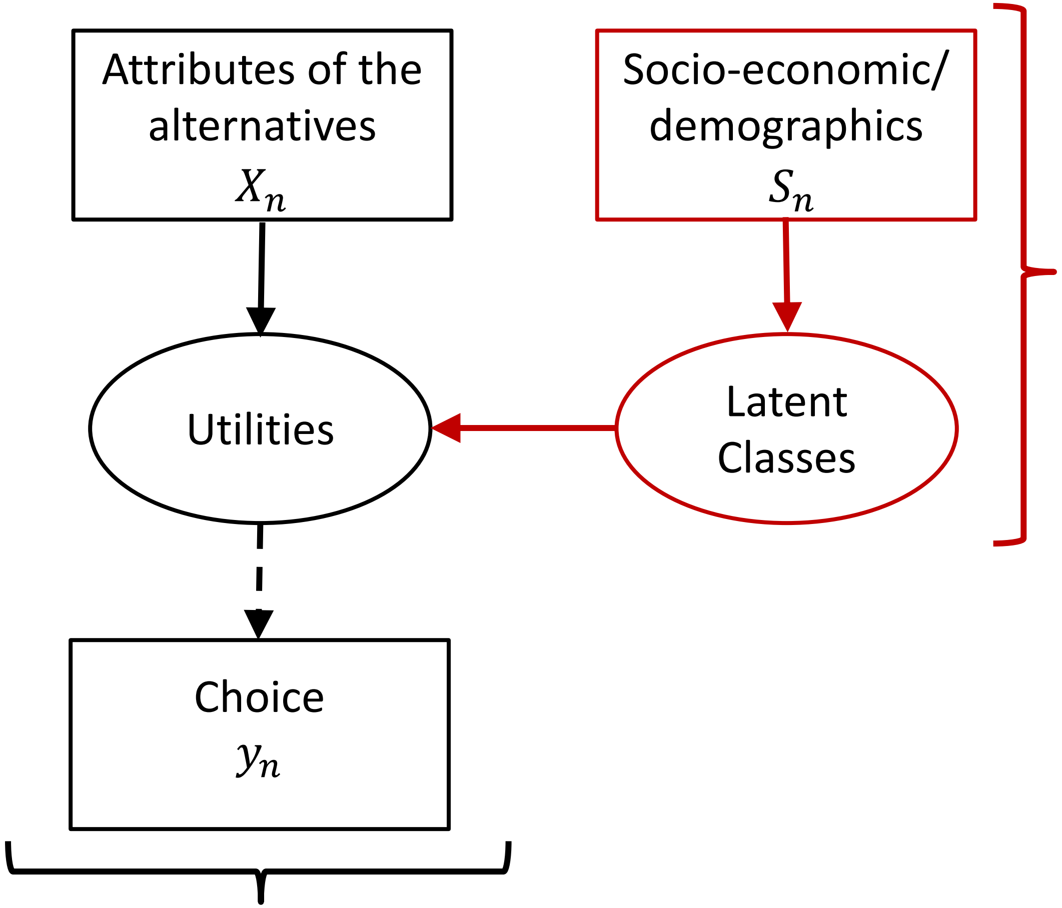
- Most previous applications:
  - Supervised classification tasks (mostly Neural Networks)
  - Better prediction accuracy but at the expense of economic interpretability
- Few studies have used ML clustering algorithms in choice modeling
  - Nonlinear-LCCM to specify the latent classes using Neural Networks (Han, 2019)

- This paper:
  - Integrates clustering ML with LCCM
  - Improves prediction accuracy and representation of heterogeneity
  - Abides by McFadden's four properties

➔ Maintain economic and behavioral interpretability

Modelling Framework

Modelling Framework



Class-Specific Choice Model:

- Discrete Choice Models

Class Membership/Clustering Model:

- **Gaussian Process (GP)**
  - Non-parametric probabilistic ML technique
  - Bayesian Framework
  - Kernel-based

Simultaneous estimation

More flexibility than utility specification

Interpretability

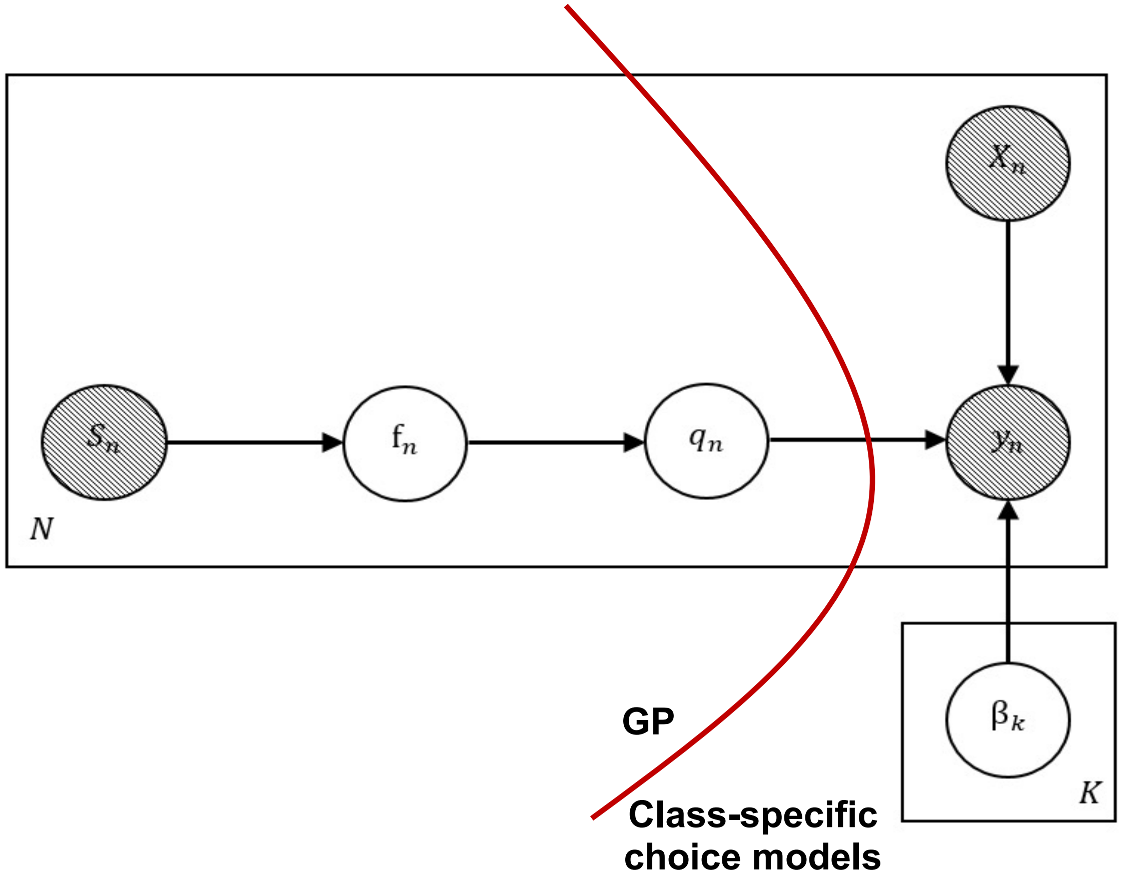
- **Gaussian-Bernoulli Mixture Model (GBM) (Sfeir et al., 2021)**

Parametric model-based clustering technique

GP-LCCM – Graphical Model

1) Class-specific choice models:
   - X_n: exogenous attributes of the alternatives
   - y_n : choices of individual n
   - β_k : vector of vector of unknown parameters

2) GP – Class membership model:
   - S_n: vector of socio-economic/demographic variables of individual n
   - f_n: latent variable
   - q_n: vector of q_{nk} - equal to 1 if individual n belongs to latent class k and 0 otherwise



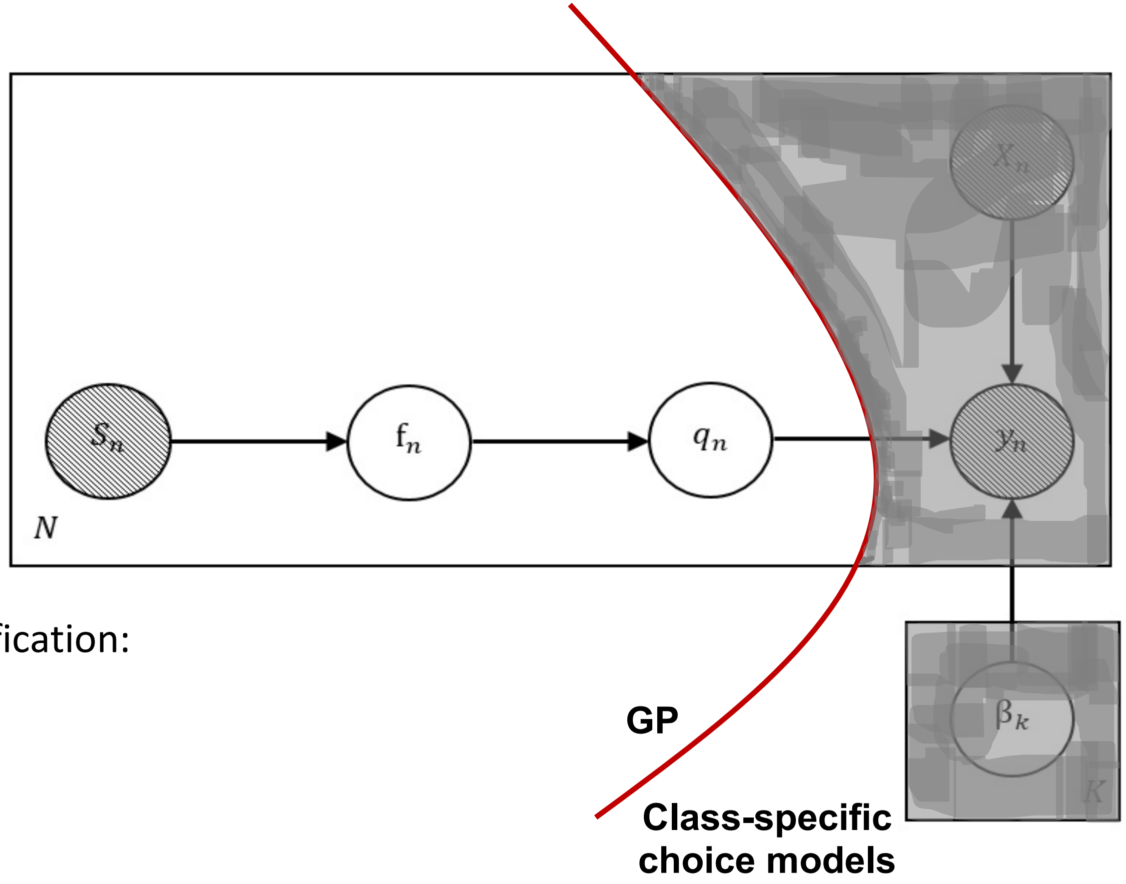
Gaussian Process (GP)

1) Instead of parametrizing the target variables or placing priors over the unknown parameters of a predefined distribution, a GP defines a prior distribution over a latent function:

$$\boldsymbol{f \sim GP(m(S_n) = 0, C(S_n, S_m))}$$

Mean function that represents the expected value of each latent variable $f(S_n)$

Covariance function (kernel) that represents the variance between every pair of latent variables $f(S_n)$ and $f(S_m)$



2) Specify a link function for the classes to obtain a probabilistic classification:

$$\boldsymbol{P(q_{nk}|f_n) = \frac{1}{1 + \exp(-f_n)}}$$

3) The posterior over $\boldsymbol{f}$ can be determined using Bayes' theorem:

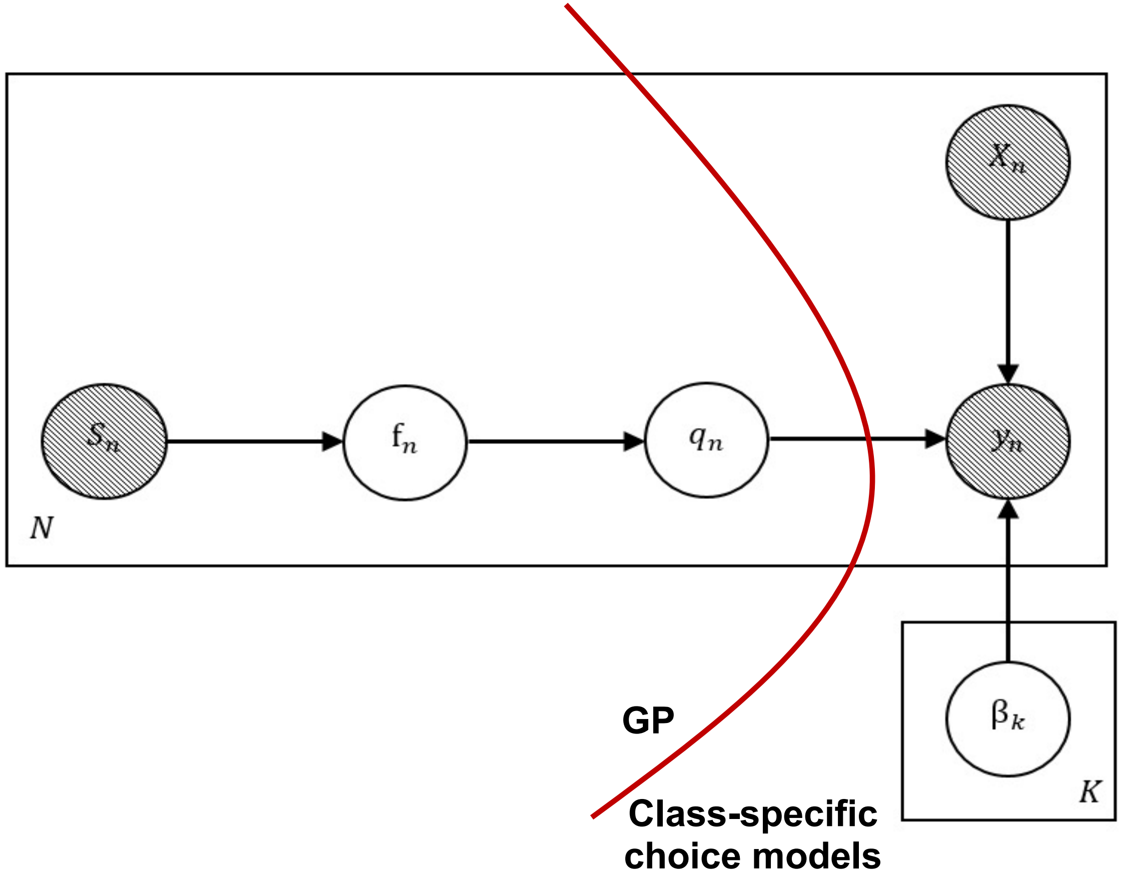
$$\boldsymbol{P(f_n|q_{nk}, S_n) = \frac{P(q_{nk}|f_n)P(f_n|S_n)}{P(q_{nk}|S_n)}}$$

4) The Laplace approximation approximates the posterior with a Gaussian by taking a second-order Taylor expansion of the log of the posterior around its maximum (Rasmussen and Williams, 2006)

GP-LCCM – Graphical Model

The joint probability of GP-LCCM:

$$P(f,y) = \prod_{n=1}^{N}\sum_{k=1}^{K} \underbrace{P(f_n|S_n)}_{\text{GP prior}}\underbrace{P(q_{nk}=1|f_n)}_{\text{Link function or likelihood of } q_{nk}}\underbrace{P(y_n|X_n,q_{nk}=1,\beta_k)}_{\text{Conditional choice probability}}$$



Estimation

- Maximum likelihood estimation techniques
- Maximizing the log-likelihood of both LCCM and GP-LCCM is a complex task:
  - Summation over k
  - No closed-form solutions
- Expectation-Maximization (EM) algorithm:
  - Repeatedly maximizes an expectation or a lower bound function of the likelihood
  - Two main steps:
    - **E-step:** estimates the expectations of the latent variables
    - **M-step:** estimates the unknown parameters

EM Algorithm

1. Select a kernel function GP (RBF, Matérn, periodic, etc.)
2. Write the joint likelihood assuming the clusters are observed

$$P(f,y) = \prod_{n=1}^{N}\prod_{k=1}^{K}[P(f_n|S_n)P(q_{nk}=1|f_n)]^{q_{nk}} \times \prod_{n=1}^{N}\prod_{k=1}^{K}\prod_{j=1}^{J}P\left(y_{nj}=1|X_{nj},q_{nk}=1,\beta_k\right)^{y_{nj}q_{nk}}$$

$$LL = \underbrace{\sum_{n=1}^{N}\sum_{k=1}^{K}q_{nk}log[P(f_n|S_n)P(q_{nk}=1|f_n)]}_{\text{Class Membership Model GP}} + \underbrace{\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{j=1}^{J}y_{njt}q_{nk}log\left[P\left(y_{nj}=1|X_{nj},q_{nk}=1,\beta_k\right)\right]}_{\text{Choice Model}}$$

EM Algorithm

3. Initialize the unknown parameters

4. E-Step: Evaluate the expectations of q_{nk} (class assignments)

5. M-Step:

   i. Estimate the parameters β_k by setting the derivatives of the expected LL to zero

   ii. Assign each individual to one class:

   if $E[q_{n0}] > E[q_{n1}]$ ➔ individual n belongs to class 0

   iii. Estimate the hyper-parameters of the kernel function:

   - Laplace approximation method (Bishop, 2006)

6. Evaluate the LL and check if the convergence criterion is met. If not, return to the **E-Step**

N.B. For multi-class problems, several binary one-versus rest classifiers are fitted

EM Algorithm

7. Finally, estimate the marginal choice probability
   - Compare with the traditional LCCM
   - Estimate out-of-sample prediction accuracies

$$P(y) = \prod_{n=1}^{N} \sum_{k=1}^{K} P(q_{nk} = 1 | f_n, S_n)\, P(y_n | X_n, q_{nk} = 1, \beta_k)$$

Applications

First Case Study: Swissmetro

- 1998 Stated Preference (SP) survey (Bierlaire et al., 2001)
- Sample:
  - 1,188 respondents and 10,692 observations
  - Choice set: Train, Car, and Swissmetro (SM)
- Cross validation – 80/20
- Choice models developed:
  - LCCM – K = 2 to 5
  - GP-LCCM – K = 2 to 7

Specification

- The utilities of the three alternatives are specified using:
  - Generic travel time and travel cost coefficients
  - Alternative-specific constants for the Train and Car alternatives
- The latent classes are characterized by the categorical variables:
  - AGE, MALE, INCOME, FIRST, LUGGAGE and PURPOSE

Variable	Description	Levels
AGE	The age class of respondents	Age ≤ 24*; 24 < Age ≤ 39; 39 < Age ≤ 54; 54 < Age ≤ 65; Age > 65
MALE	The respondent's gender	1: Male; 0: Female
INCOME	The respondent's income per thousand CHF per year	INCOME < 50*; 50 ≤ INCOME ≤ 100; INCOME > 100; M_INCOME: unknown income
FIRST	First class traveler	0: no; 1: yes
LUGGAGE	Number of luggage the respondent carries during a trip	0: none; 1: one piece; 2: more than one piece*
PURPOSE	Purpose of the trip	1: Commuter; 2: Shopping; 3: Business; 4: Leisure*

*: level kept as a base

Results

LCCM

K	Nb of Parameters		LL	AIC	BIC	Pred. LL
2	23		-5,930.76	11,907.52	12,069.75	-1,490.62
3	42		-5,202.71	10,489.41	10,785.67	-1,329.94
4	61		-4,870.51	9,863.02	10,293.29	-1,245.18
5	80		-4,687.99	9,535.99	10,100.28	-1,233.69

GP-LCCM – Matérn Kernel

K	Nb of Parameters	Joint LL	LL	AIC	BIC	Pred. LL
2	10	-5,930.02	-5,916.43	11,852.86	11,923.39	-1,493.52
3	18	-4,879.78	-5,176.06	10,388.11	10,515.08	-1,354.44
4	24	-4,260.21	-4,878.84	9,805.68	9,974.97	-1,263.21
5	30	-3,872.87	-4,825.55	9,711.11	9,922.72	-1,256.62
6	36	-3,564.44	-4,742.13	9,556.26	9,810.19	-1,237.09
7	**42**	**-3,346.24**	**-4,649.20**	**9,382.39**	**9,678.65**	**-1,213.44**

Results (Cont.)

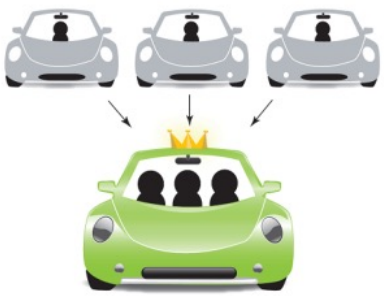
Class-Specific Choice Models of GP-LCCM (K = 7)

Parameter	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
ASC (Train)	1.65 (11.07)	-0.427 (-2.45)	0.0571 (0.33)	0.860 (3.71)	-2.12 (-5.58)	-2.89 (-16.66)	1.09 (6.35)
ASC (Car)	-3.47 (-4.37)	-3.48 (-12.63)	2.19 (19.69)	5.03 (17.65)	-0.532 (-3.57)	-1.81 (-12.39)	0.912 (5.03)
Travel Time	-1.24 (-13.67)	-5.56 (-41.79)	-4.91 (-53.65)	-2.93 (-21.20)	-6.44 (-47.29)	-0.331 (-4.62)	-0.0719 (-1.38)
Travel Cost	-0.441 (-6.57)	-0.570 (-10.30)	-3.21 (-43.56)	-1.30 (-9.16)	-3.91 (-39.87)	-6.63 (-57.86)	-0.719 (-5.73)
Class Share	11.69%	14.54%	29.61%	7.15%	26.17%	6.98%	3.87%
VOT (CHF/min)	2.82	9.75	1.53	2.26	1.65	0.05	0.10

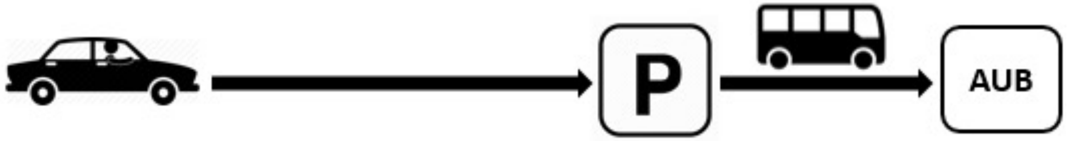
Around 75% of individuals have reasonable VOTs
(Bierlaire et al., 2001)

Second Case Study: AUB Dataset

- 2017 Stated Preference (SP) survey (Sfeir et al., 2020)
- Data about intended weekly frequency of commuting by three modes
  - Shared-Taxi (ST)
  - Shuttle (SH)
  - Current mode of commute to AUB: Car or PT



Shared-Taxi (ST)



Shuttle (SH)

Second Case Study: AUB Dataset (Cont.)

- Sample used:
  - Car users who come 5 days per week to AUB
  - 650 respondents and 2,600 choice scenarios
- 5-fold cross validation
- Choice models developed:
  - LCCM – K = 2
  - GBM-LCCM – K = 2 and 3
  - GP-LCCM – K = 2 and 3

Specification

- The class-specific utility functions of each alternative are characterized by the corresponding:
  - Travel time and travel cost
  - Constants related to the frequency of using the available modes
- The latent classes are characterized by the categorical variables:
  - Age, Grade, C/D, Nb

Variable	Description
Age	Age of the respondent (in years/10)
$Grade$	A number between 1 and 16 used to specify the salary of a staff member (Grade/10)
C/D	Ratio of number of cars available over number of licensed drivers per household
Nb	Number of people who are usually present in the car during the trip from home to AUB

Results

K	Model	Nb of parameters	Joint LL	LL	AIC	BIC	Average Pred. LL
2	LCCM	47		-4,910.92	9,915.84	10,191.41	-1,024.93
	GBM-LCCM	61	-8,533.22	-4,911.08	9,944.16	10,301.82	-1,012.62
	GP-LCCM	44	-4,905.31	**-4,877.73**	**9,843.46**	**10,101.44**	**-995.76**
3	GBM-LCCM	80	-7,042.21	-4,893.29	9,946.58	10,415.64	-998.41
	GP-LCCM	72	**-4,480.70**	**-4,691.25**	**9,526.50**	**9,948.66**	**-935.23**

Latent Classes – GP-LCCM

- GPs make the latent classes less transparent

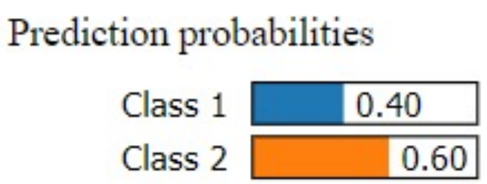
- Model-agnostic techniques
  - Infer explanations from a trained “black-box” model

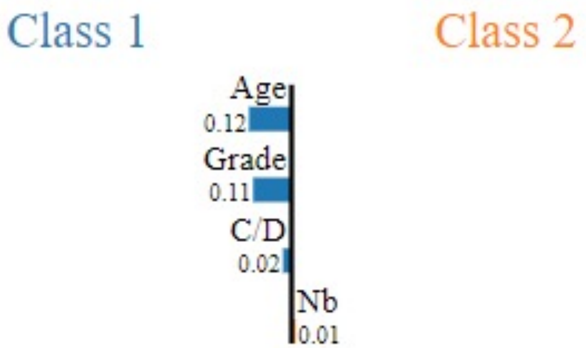
- Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016)
  - Individual (local) interpretation

Latent Classes – GP-LCCM (Cont.)

<u>Individual 1:</u>

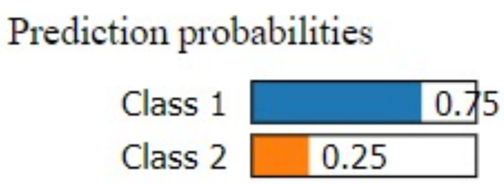
- Age = 24
- Grade = 0
- C/D = 0.5
- Nb = 0

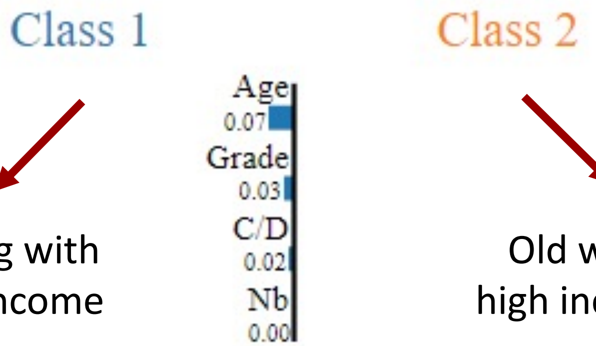




<u>Individual 2:</u>

- Age = 45
- Grade = 16
- C/D = 1
- Nb = 4





Young with
low income

Old with
high income

Conclusion

Findings

	GP-LCCM
Class membership model (latent classes)	Non-parametric
Class interpretability	Less transparent Model-agnostic techniques are required
Class-specific choice models	DCM
Choice models interpretability (marginal effects, economic indicators, etc.)	Maintained
Heterogeneity representation	More flexible than LCCM and GBM-LCCM
Predictive power	Better than LCCM and GBM-LCCM
Overfitting	More likely
Computational time	Minutes to hours

Future Directions

- Investigate whether the findings generalize to different applications, specifications, and attribute transformations
- Extend the models to account for within-class heterogeneity
- Automate the task of finding an appropriate kernel(s) function
- Explore Bayesian Variational Inference (VI) estimation techniques

References

- Walker, J. L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. Journal of Geographical Systems, 9(1), 77–101. https://doi.org/10.1007/s10109-006-0030-0
- Hillel, T., Bierlaire, M., Elshafie, M.Z.E.B., Jin, Y., 2021. A systematic review of machine learning classification methodologies for modelling passenger mode choice. J. Choice Model. 38. https://doi.org/10.1016/j.jocm.2020.100221
- Han, Y., 2019. Neural-Embedded Discrete Choice Models. Massachusetts Institute of Technology (PhD Thesis).
- Sfeir, G., Abou-Zeid, M., Rodrigues, F., Pereira, F.C., Kaysi, I., 2021. Latent class choice model with a flexible class membership component: A mixture model approach. J. Choice Model. 41. https://doi.org/10.1016/j.jocm.2021.100320
- Rasmussen, C.E., Williams, C., 2006. Gaussian Processes for Machine Learning. the MIT Press.

References

- Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer, New York.
- Bierlaire, M., Axhausen, K.W., Abay, G., 2001. The acceptance of modal innovation: The case of Swissmetro, in: 1st Swiss Transport Research Conference. Ascona, Switzerland.
- Sfeir, G., Abou-Zeid, M., Kaysi, I., 2020. Multivariate count data models for adoption of new transport modes in an organization-based context. Transp. Policy 91, 59–75. https://doi.org/10.1016/j.tranpol.2020.03.014
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Model-Agnostic Interpretability of Machine Learning. arXiv:1606.05386.

DTU MLSM
Machine Learning for Smart Mobility

Thank you

Reach us:

- MLSM website: https://mlsm.man.dtu.dk/
- Twitter: @mlsm_dtu
- Email: mlsm@man.dtu.dk – geosaf@dtu.dk

The 7th International Choice Modelling Conference 2022