

Validation of probabilistic classifiers

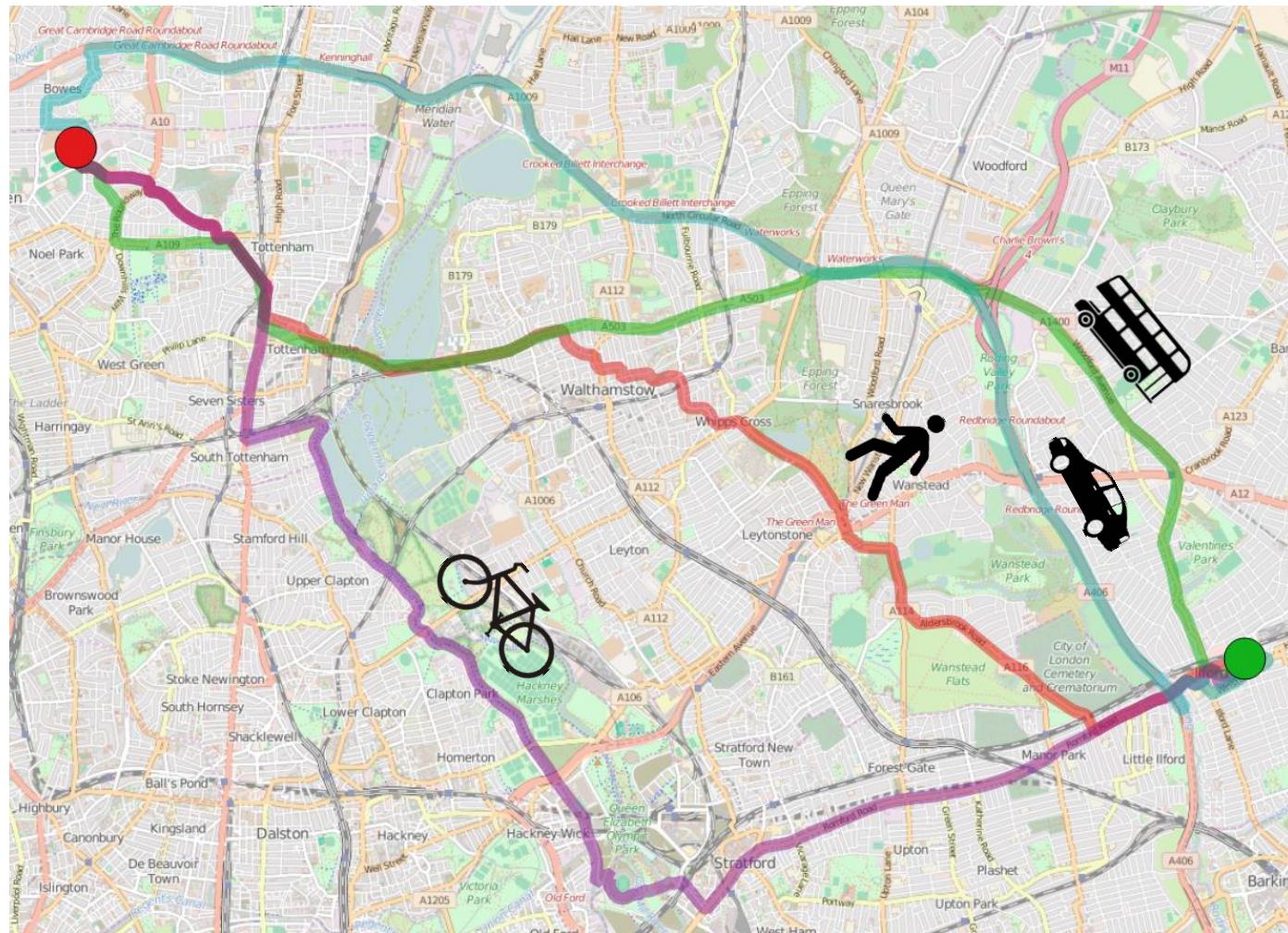
STRC 2018

Tim Hillel, Michel Bierlaire, Mohammed Elshafie, Ying Jin

École Polytechnique Fédérale de Lausanne EPFL

University of Cambridge

Motivation



UNIVERSITY OF
CAMBRIDGE

Motivation

P^m

P^r

?

$P \dots$

Motivation

$$U_i = \beta_t * T_i + \dots$$

$$U_i = \beta_t * T_i + \dots$$

Motivation

$$U_i = \beta_t * T_i + \dots \quad U_i = \beta_t * T_i + \dots$$

Likelihood ratio test

Motivation

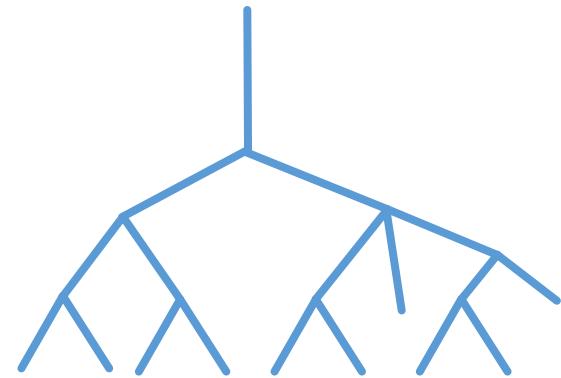
$$U_i = \beta_t * T_i + \dots \quad U_i = \beta_t * T_i + \dots$$

Likelihood ratio test

AIC/BIC

Motivation

$$U_i = \beta_t * T_i + \dots$$



Overview

1. Statistical tests
 - a. Single true-model
 - b. Kullback-Leibler divergence
 - c. Convex combination of classifiers
2. Initial experiments
3. Conclusions and further work

Single true-model test

$$P^m$$

Single true-model test

$$\mathcal{L}^m = \sum_{n=1}^N \sum_{i=1}^J \ln P^m(i|x_n) y_{in}$$

Single true-model test

$$\mathcal{L}^m = \sum_{n=1}^N \sum_{i=1}^J \ln P^m(i|x_n) y_{in}$$

$$E[\mathcal{L}^m|x_V] = \sum_{n=1}^N \sum_{i=1}^J \ln P^m(i|x_n) P^*(i|x_n)$$

Single true-model test

$$\text{E}[\mathcal{L}^m|x_V] = \sum_{n=1}^N \sum_{i=1}^J \ln P^m(i|x_n)P^*(i|x_n)$$

$$H_0 : P^m = P^*$$

Single true-model test

$$\text{E}[\mathcal{L}^m|x_V] = \sum_{n=1}^N \sum_{i=1}^J \ln P^m(i|x_n)P^m(i|x_n)$$

$$H_0 : P^m = P^*$$

Kullback-Leibler divergence test

$$P^r \ P^m$$

Kullback-Leibler divergence test

$$\mathcal{D}(P^r \| P^m) = \sum_{n=1}^N \sum_{i=1}^J \ln \left(\frac{P^r(i|x_n)}{P^m(i|x_n)} \right) P^r(i|x_n)$$

Kullback-Leibler divergence test

$$\text{E}[\mathcal{D}(P^r \| P^m)] = \sum_{n=1}^N \sum_{i=1}^J \ln \left(\frac{P^r(i|x_n)}{P^m(i|x_n)} \right) P^r(i|x_n)$$

Kullback-Leibler divergence test

$$\text{E}[\mathcal{D}(P^r \| P^m)] = \sum_{n=1}^N \sum_{i=1}^J \ln \left(\frac{P^r(i|x_n)}{P^m(i|x_n)} \right) P^r(i|x_n)$$

$$H_0 : P^m = P^r$$

Kullback-Leibler divergence test

$$\mathbb{E}[\mathcal{D}(P^r \| P^m)] = \sum_{n=1}^N \sum_{i=1}^J \ln \left(\frac{P^r(i|x_n)}{P^m(i|x_n)} \right) P^r(i|x_n)$$

$$H_0 : P^m = P^r$$

$$H_0 : \mathcal{D}(P^r \| P^m) = 0$$

Convex combination of classifiers test

$$P(i|x_n; \lambda) = \lambda P^r(i|x_n) + (1 - \lambda) P^m(i|x_n)$$

Convex combination of classifiers test

$$P(i|x_n; \lambda) = \lambda P^r(i|x_n) + (1 - \lambda) P^m(i|x_n)$$

$$H_0 : \lambda = 0$$

$$H_1 : \lambda \neq 0$$

Methodology

- SwissMetro dataset (*Bierlaire, Axhausen, Abay, 2001*)

Methodology

- SwissMetro dataset (*Bierlaire, Axhausen, Abay, 2001*)



Methodology

Variable		Alternative		
		SM	Car	SBB
ASC	Constant	SM	Car	-
TT	Travel time	B-Time	B-Time	B-Time
Cost	Travel cost	B-Cost	B-Cost	B-Cost
Freq	Frequency	B-Freq	-	B-Freq
GA	Annual season	B-GA	-	B-GA
Age	Age in classes	-	-	B-Age
Luggage	Pieces of luggage	-	B-Luggage	-
Seats	Airline seating	B-Seats	-	-

Models

1. Base model
2. True model
3. No-luggage
4. No-age
5. Time-cost

Base model

Variable		Alternative		
		SM	Car	SBB
ASC	Constant	SM	Car	-
TT	Travel time	B-Time	B-Time	B-Time
Cost	Travel cost	B-Cost	B-Cost	B-Cost
Freq	Frequency	B-Freq	-	B-Freq
GA	Annual season	B-GA	-	B-GA
Age	Age in classes	-	-	B-Age
Luggage	Pieces of luggage	-	B-Luggage	-
Seats	Airline seating	B-Seats	-	-

True model

Variable		Alternative		
		SM	Car	SBB
ASC	Constant	SM	Car	-
TT	Travel time	B-Time	B-Time	B-Time
Cost	Travel cost	B-Cost	B-Cost	B-Cost
Freq	Frequency	B-Freq	-	B-Freq
GA	Annual season	B-GA	-	B-GA
Age	Age in classes	-	-	B-Age
Luggage	Pieces of luggage	-	B-Luggage	-
Seats	Airline seating	B-Seats	-	-

No-luggage model

Variable		Alternative		
		SM	Car	SBB
ASC	Constant	SM	Car	-
TT	Travel time	B-Time	B-Time	B-Time
Cost	Travel cost	B-Cost	B-Cost	B-Cost
Freq	Frequency	B-Freq	-	B-Freq
GA	Annual season	B-GA	-	B-GA
Age	Age in classes	-	-	B-Age
Luggage	Pieces of luggage	-	B Luggage	-
Seats	Airline seating	B-Seats	-	-

No-age model

Variable		Alternative		
		SM	Car	SBB
ASC	Constant	SM	Car	-
TT	Travel time	B-Time	B-Time	B-Time
Cost	Travel cost	B-Cost	B-Cost	B-Cost
Freq	Frequency	B-Freq	-	B-Freq
GA	Annual season	B-GA	-	B-GA
Age	Age in classes	-	-	B-Age
Luggage	Pieces of luggage	-	B-Luggage	-
Seats	Airline seating	B-Seats	-	-

Time-cost model

Variable		Alternative		
		SM	Car	SBB
ASC	Constant	SM	Car	-
TT	Travel time	B-Time	B-Time	B-Time
Cost	Travel cost	B-Cost	B-Cost	B-Cost
Freq	Frequency	B Freq	-	B Freq
GA	Annual season	B GA	-	B GA
Age	Age in classes	-	-	B Age
Luggage	Pieces of luggage	-	B Luggage	-
Seats	Airline seating	B Seats	-	-

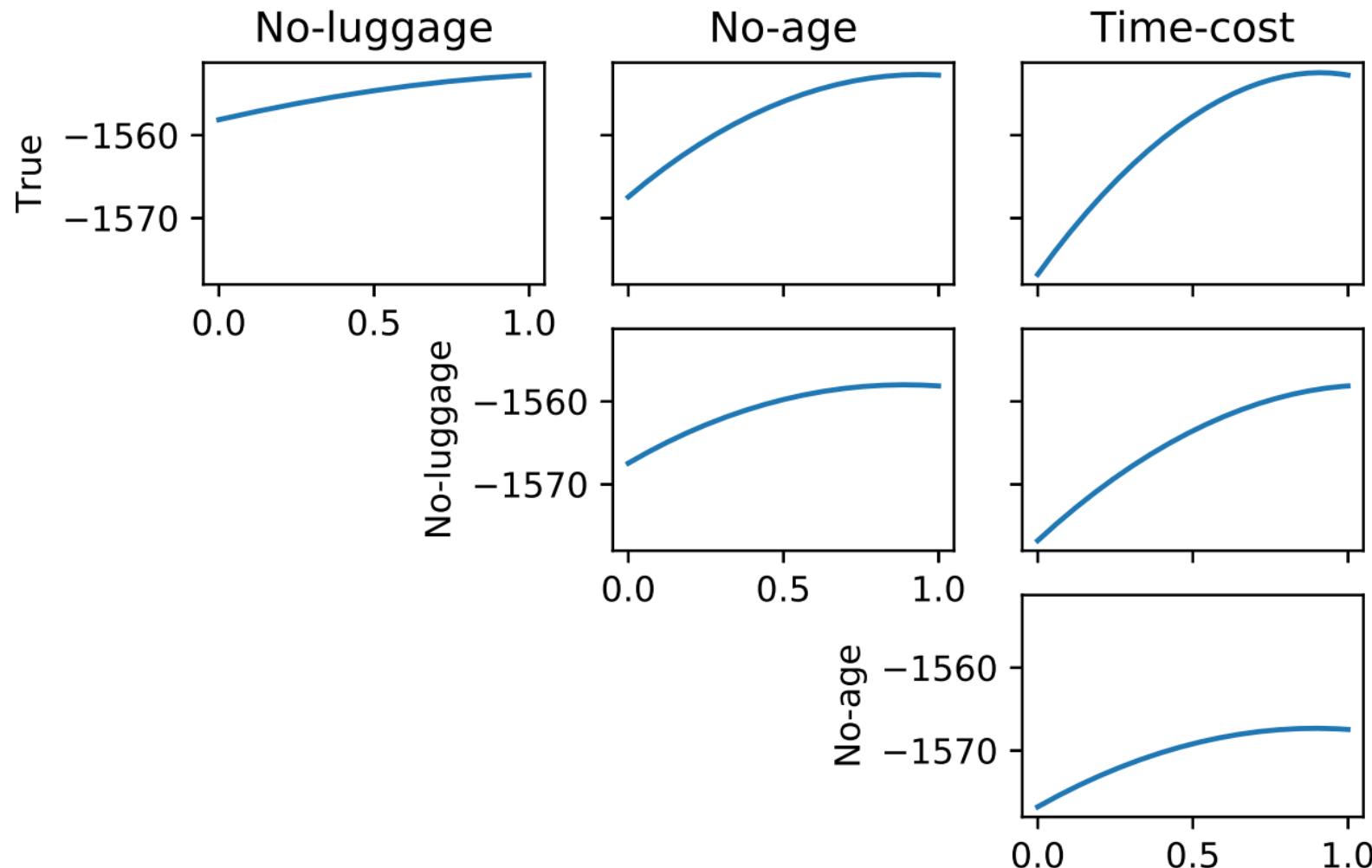
Results: Single true-model test

Model	Base	True	No-luggage	No-age	Time-cost
Train LL (N=4374)	-3712.07	-3772.96	-3772.69	-3781.53	-3811.95
Test LL (N=2034)	-1545.31	-1552.75	-1558.13	-1567.46	-1576.81
μ_m	-1593.11	-1593.11	-1619.01	-1625.42	-1636.49
σ_m	26.39	26.39	26.20	26.16	26.23
z	1.811	1.529	2.324	2.216	2.275
p	0.0701	0.1262	0.0201	0.0267	0.0229

Results: Kullback-Leibler divergence test

	True	No-luggage	No-age	Time-cost
True	-	0.1809	0.0046	0.0001
No-luggage	0.1900	-	0.0167	0.0038
No-age	0.0048	0.0162	-	0.0103
Time-cost	0.0001	0.0042	0.0117	-

Results: Convex combination of classifiers test



Conclusions/further work

- 3 proposed tests for parametric and non-parametric models
- Initial results indicate potential consistency with parametric tests
- Need formal testing and simulation on multiple scenarios and datasets