Are you the Beatles or the Rolling Stones? Learning the choice preferences of air travelers

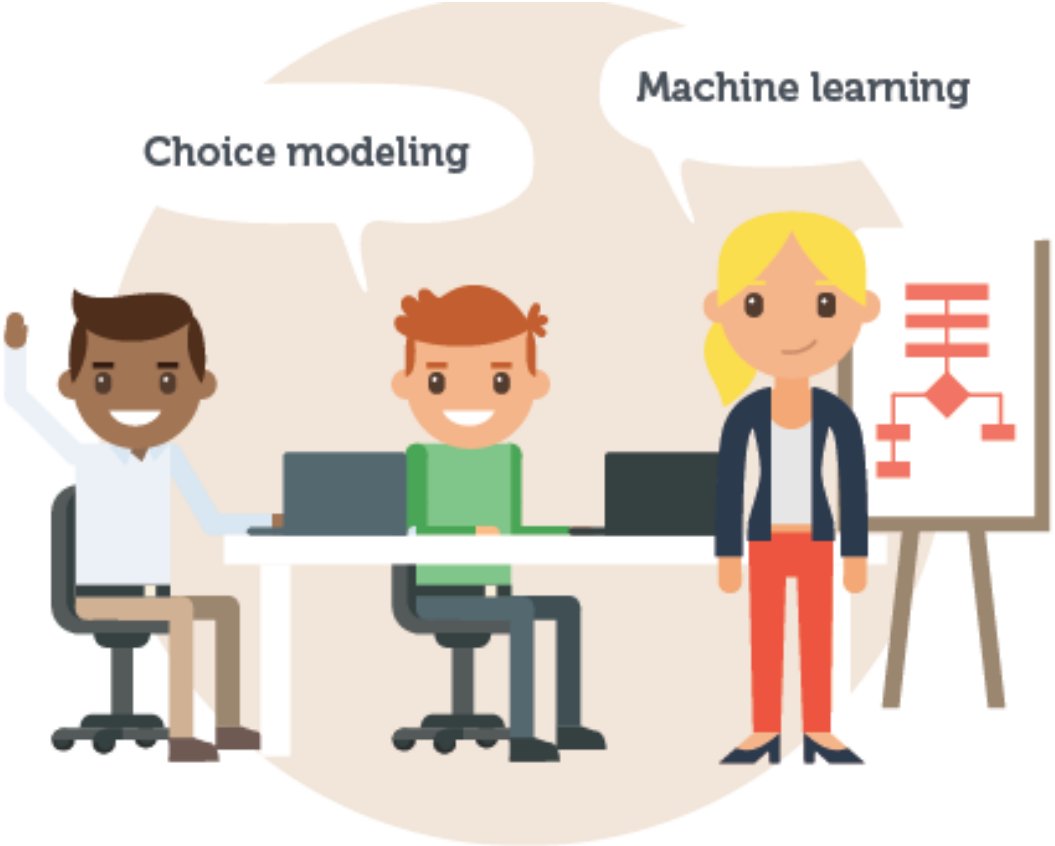
Virginie Lurkin

Eindhoven University of Technology

2019 AGIFORS RM Study Group
Panama City

Motivation

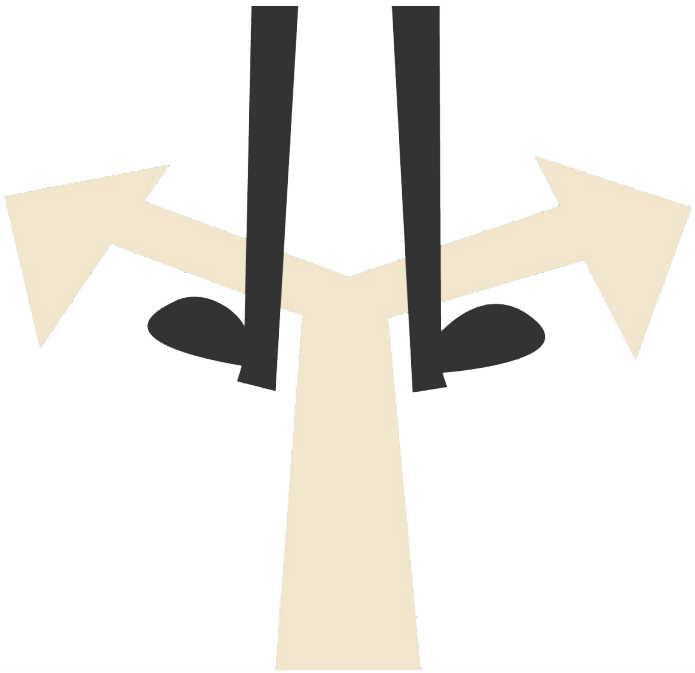
ARE YOU
THE BEATLES
OR
the Rolling Stones



"The new and good ideas come from having a very broad and multidisciplinary range of interests."

Robin Chase

Discrete choice models



Predict individual choice behavior
using mathematical models

Discrete choice models

Understand individual choice behavior
using mathematical models

Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

$$V_{in} = ... + \beta_c Cost_{in} + \beta_t Time_{in} + ...$$

Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

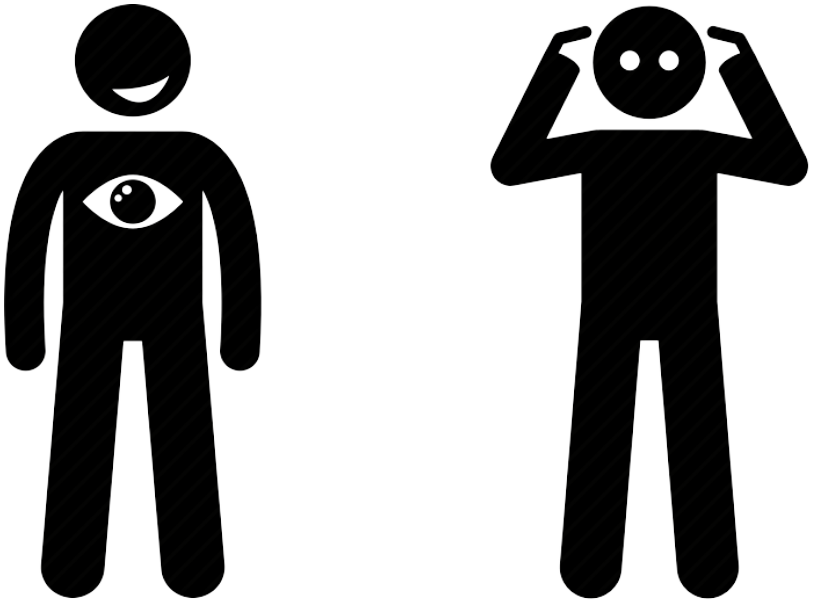
$$V_{in} = \underbrace{... + \beta_c Cost_{in} + \beta_t Time_{in} + ...}_{\text{Utility specification}}$$

Utility specification

Determining the *best* specification for a particular application is a difficult task

Utility specification

Determining the *best* specification for a particular application is a difficult task



Expertise + Intuition

Utility specification

Determining the *best* specification for a particular application is a difficult task

Inspiration + Experience

Utility specification

Determining the *best* specification for a particular application is a difficult task

Trial-and-error

Utility specification

Determining the *best* specification for a particular application is a difficult task

Time consuming

Utility specification

Determining the *best* specification for a particular application is a difficult task

↓

Utility specification

Utility specification

Determining the *best* specification for a particular application is a difficult task

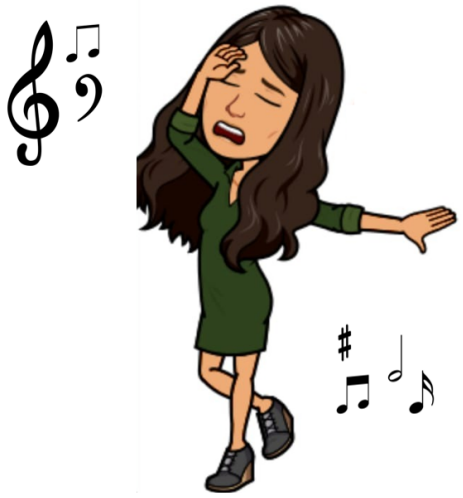
↓

Utility specification

Prediction accuracy + Interpretability

Utility specification

Determining the *best* specification for a particular application is a difficult task



Help me if you can, I'm feeling down

Machine learning

Neural networks

Machine learning

Neural networks

Very good prediction accuracy

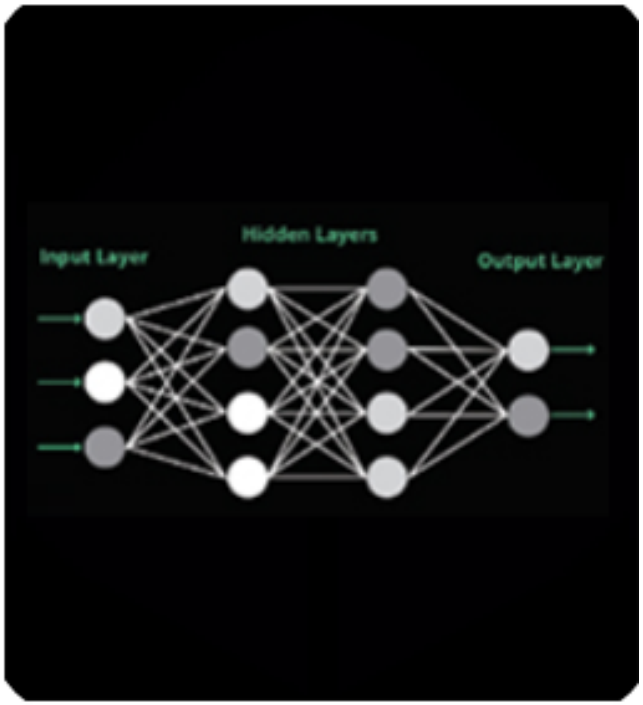
Machine learning

Neural networks

Very good prediction accuracy

Machine learning

Machine learning



Our idea

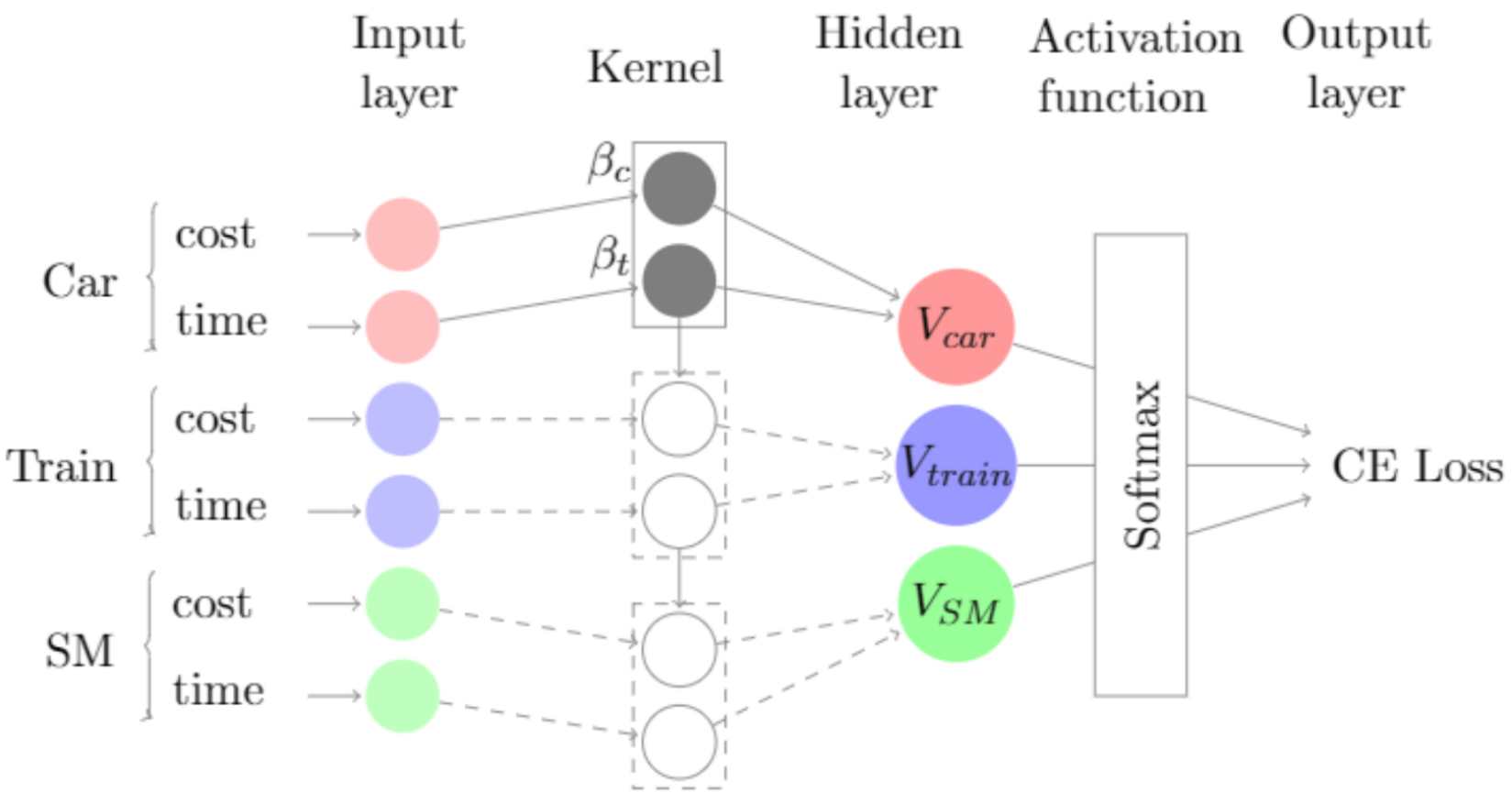
- Use the **predictive strength of Neural Networks** to automatically discover from available data good representation needed for utility specification, **while keeping interpretability**.

- Interpretability is critical for researchers and practitioners to get insights into the complex human decision-making process.

- For instance linear specifications allow for straightforward derivation of the VOT that constitutes a highly relevant measure in a wide range of critical decisions including aircraft purchasing, codeshare agreements, scheduling decisions..

Learning Choice Models

MNL as CNN


Input layer
Kernel
Hidden layer
Activation function
Output layer
Car
cost
time
Train
cost
time
SM
cost
time
β_c
β_t
V_{car}
V_{train}
V_{SM}
Softmax
CE Loss

MNL as CNN

Softmax activation function:

$$(\sigma(\mathbf{V}_n))_i = \frac{e^{V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}}}$$

Categorical cross-entropy function:

$$H_n(\sigma, \mathbf{y}_n) = -\sum_{i \in \mathcal{C}_n} y_{in} \log \left[\sigma_i(\mathbf{V}_n)\right]$$

Learning MNL

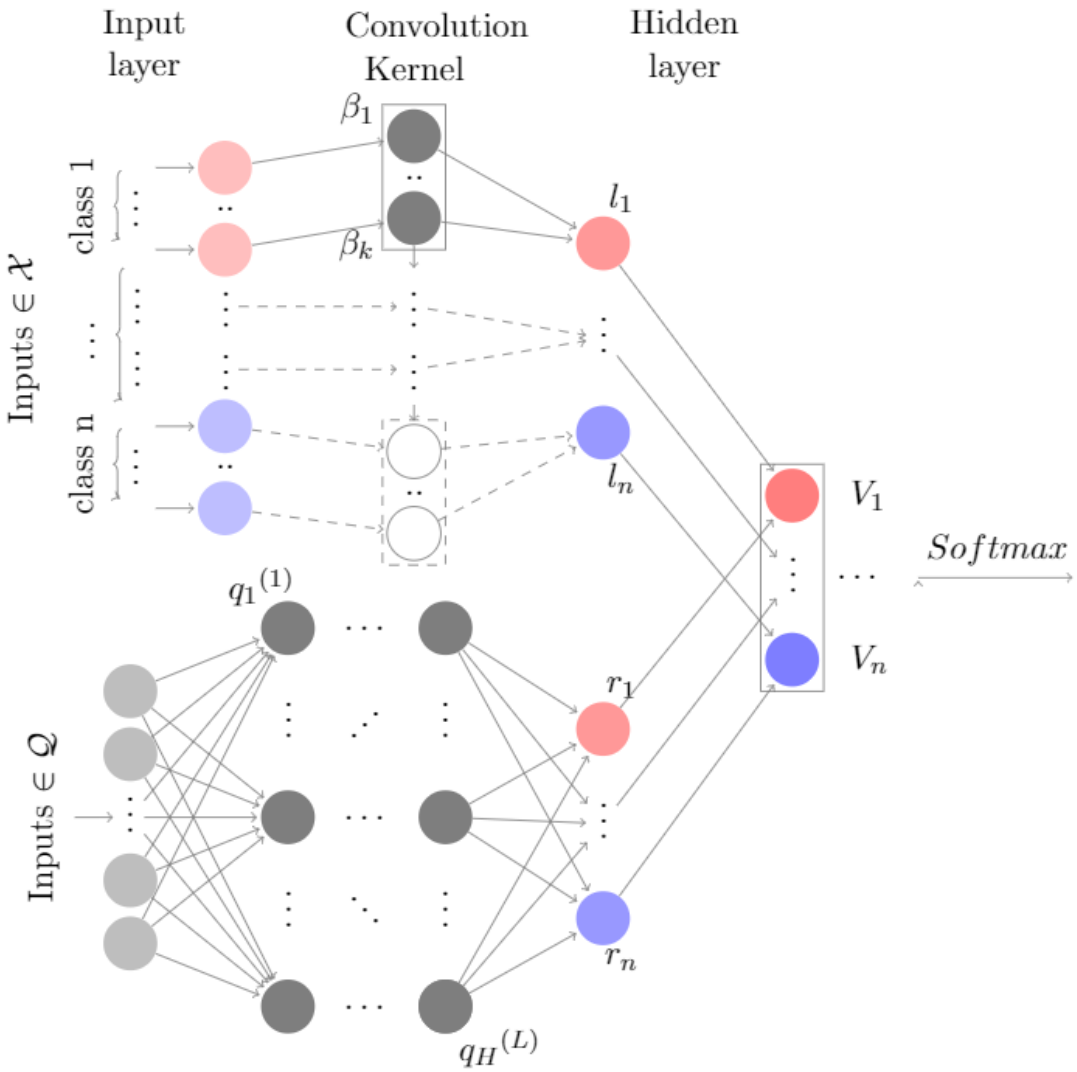
$$V_{in} = \ell_{in} + r_{in},$$

where

- $\ell_n = f(\mathcal{X}_n; \beta)$ is the assumed **interpretable part of the utility specification**, modeled using a function f whose unknown model parameters β are a linear combination of the explanatory variables (or input features) $\mathcal{X}_n$.

- $r_n = \psi(\mathcal{Q}_n, w)$ is a **representation that is learned** from a set of explanatory variables (or input features) $\mathcal{Q}_n$ for which no *a priori* relationship is assumed. This term is best recovered with an ANN, but is generalizable to any other representation learning method.

L-MNL architecture



Experiments - Synthetic data

Data generation

We consider a simple binary choice model:

$$U_n = V_n + \varepsilon_n,$$

where

$$V_n = \underbrace{\beta_1 \cdot x_1 + \beta_2 \cdot x_2}_{linear} + \underbrace{\beta_3 \cdot x_3 x_4 + \beta_4 \cdot x_3 x_5}_{non-linear}.$$

For the explanatory variables, we assume that

$$x_k \sim \mathcal{N}(0,1), \qquad \forall k = 1, ...5.$$

Data generation

The synthetic choice y_n can be seen as a Bernoulli random variable:

$$y_n \sim \text{Bern}\left(P_n(x)\right), \qquad \forall n = 1, ...N.$$

where, under logit assumption:

$$P_n(x) = \frac{e^{V_n}}{1 + e^{V_n}},$$

We randomly sample 100 times from

$$\beta_k \sim \mathcal{U}([-5, 5] \setminus [-0.5, 0.5]), \qquad \forall k = 1, ...5,$$

and for each specification, we generate $1,000$ individual observations.

Benchmarking models

$$V_n = \underbrace{\beta_1 \cdot x_1 + \beta_2 \cdot x_2}_{linear} + \underbrace{\beta_3 \cdot x_3 x_4 + \beta_4 \cdot x_3 x_5}_{non-linear}.$$

- L-MNL$(100, \mathcal{X}, \mathcal{Q})$, with $\mathcal{X} = \{x_1, x_2\}$, and $\mathcal{Q} = \{x_3, x_4, x_5\}$,
- Logit$(\mathcal{X}_1)$, with $\mathcal{X}_1 = \{x_1, x_2, x_3, x_4, x_5\}$,
- Logit$(\mathcal{X}_2)$, with $\mathcal{X}_2 = \{x_1, x_2\}$,
- Hruschka2004$(100, \mathcal{Q})$, with $\mathcal{Q} = \{x_1, x_2, x_3, x_4, x_5\}$,
- Hruschka2007$(3, \mathcal{X})$, with $\mathcal{X} = \mathcal{Q} = \{x_1, x_2, x_3, x_4, x_5\}$,
- Hruschka2007$(100, \mathcal{X})$, with $\mathcal{X} = \mathcal{Q} = \{x_1, x_2, x_3, x_4, x_5\}$.

Log likelihood results

Model	Train set		Test set	
	$\overline{LL}$	$s.e.(LL)$	$\overline{LL}$	$s.e.(LL)$
L-MNL$(100, \mathcal{X}, \mathcal{Q})$	-207	47	-51	13
Logit$(\mathcal{X}_1)$	-464	90	-94	19
Logit$(\mathcal{X}_2)$	-466	90	-94	19
Hruschka2004$(100, \mathcal{Q})$	-187	46	-55	14
Hruschka2007$(3, \mathcal{X})$	-214	47	-52	12
Hruschka2007$(100, \mathcal{X})$	-214	46	-52	12

Relative errors in parameters estimates

Model	$\overline{e}_{\beta_1}$	$s.e.(e_{\beta_1})$	$\overline{e}_{\beta_2}$	$s.e.(e_{\beta_2})$	$\overline{e}_{\beta_2/\beta_1}$	$s.e.(e_{\beta_2/\beta_1})$
L-MNL$(100, \mathcal{X}, \mathcal{Q})$	9.2	8.2	8.5	8.5	6.7	8.0
Logit$(\mathcal{X}_1)$	50.4	11.4	51.6	11.6	9.9	8.8
Logit$(\mathcal{X}_2)$	50.1	11.1	52.0	11.3	10.0	8.7
Hruschka2007$(3, \mathcal{X})$	37.4	13.7	39.1	13.0	13.2	11.5
Hruschka2007$(100, \mathcal{X})$	36.7	13.9	38.6	15.7	16.6	16.8

$$e_\beta = \left| \frac{\beta - \widehat{\beta}}{\beta} \right|, e_{\beta_i/\beta_j} = \left| \frac{e_{\beta_i} - e_{\beta_j}}{1 - e_{\beta_j}} \right|.$$

Relative errors in parameters estimates

Model	**% of experiments not rejecting H_0**	
	β_1 and β_2	β_2/β_1
L-MNL$(100, \mathcal{X}, \mathcal{Q})$	78.5	98
Logit$(\mathcal{X}_1)$	1.5	97
Logit$(\mathcal{X}_2)$	1.5	97
Hruschka2007$(3, \mathcal{X})$	10	91
Hruschka2007$(100, \mathcal{X})$	9.5	83

$$H_0 : \widehat{\beta} = \beta,\ H_1 : \widehat{\beta} \neq \beta.$$

Conclusions

Conclusion

- New hybrid model
- NN predictive power: better fit in terms of likelihood
- Ability learn complex relationships from the data
- Don't compromise the interpretability of the linear parameters

classic mixes
1
VOLUME
I LOVE
THE BEATLES &
THE ROLLING STONES
Volume 1
DMC
DMC 35
DMC CLASSIC PRODUCTIONS • STRICTLY DJ USE ONLY

Thank you!