

# **TasteNet-Multinomial Logit Model**

January 20, 2025

# Contents

1. TasteNet .....	3
1.1. Ideas generales .....	3
1.2. Modelación de los gustos .....	3
1.3. Arquitectura de TasteNet .....	4
1.4. Consideraciones de la modelación .....	4
Bibliography .....	6
Index of Figures .....	7

# 1. TasteNet

El presente archivo busca ser una explicación general de la arquitectura de TasteNet-MNL, para ello comienza dando ideas generales en la modelación del problema y como los autores del paper lo abordan, el valor de este archivo es que busca ser clarificador respecto a como se construye la arquitectura y agrega observaciones y comentarios respecto a su funcionamiento, los cuales fueron recogidos a lo largo del estudio del modelo. Cabe mencionar que este archivo no busca ser una revisión exhaustiva del paper, si no que una revisión enfocada en el modelo propuesto por lo que diversos puntos comentados en el paper no han sido tratados. Entre ellos, se puede mencionar discusiones respecto a las redes neuronales y su papel como aproximadores universales, redes neuronales embebidas en modelos de elección discreta, una muy interesante sección dedicada a la interpretabilidad en modelos de elección discreta, entre otros.

## 1.1. Ideas generales

Al comenzar a modelar problemas de elección discreta es común modelar a los individuos como consumidores racionales, es decir, que buscan escoger la opción que maximice su utilidad. En base a esto, surge una gran familia de modelos de elección discreta que se diferencian entre sí por la modelación que hace cada uno de su función de utilidad. En general, se asume que la función de utilidad tiene como variables los atributos de cada opción, las características del individuo, errores de observación asociados a ellos (tanto atributos como características individuales) y el conjunto de alternativas que posee para escoger.

Con estas consideraciones, los autores mencionan que las redes neuronales se han implementado en modelos de elección discreta principalmente con un enfoque, que es aproximar la función de utilidad. Esta tarea puede ser sumamente difícil dada la naturaleza de la misma, por lo que proponen fijar una meta más acotada e intentar aproximar parte de ella, que corresponde a los “tastes” de los individuos.

Los “tastes” se pueden entender como gustos, valoraciones o predisposiciones de cada persona en base a sus características, por ejemplo, si un individuo tiene entre sus características que es una persona de altos ingresos y trabaja con jornada completa, es posible que esta persona tenga una mayor valoración por el tiempo pese a elevar los costos de la alternativa que escoga.

## 1.2. Modelación de los gustos

No es sencillo modelar como los gustos de las personas actúan al momento de asignar la utilidad a una de las alternativas que se le presenta, esta puede ser una función de cualquier tipo en los parámetros, sin embargo, para mantener simplicidad en la modelación y fortalecer la interpretabilidad del modelo los autores representan que los gustos son una función lineal con respecto a sus características. Con ello, su función de utilidad resultante se puede escribir de la siguiente manera:

$$V_{in} = \beta_{in}^{TN} \cdot x_{in}^{TN} + f(x_i^{MNL}, z_n^{MNL}; \beta_i^{MNL}) \quad (1)$$

Acá,  $\beta_{in}^{TN} = TasteNet(z_n^{TN}; \omega)_i'$  corresponde una red neuronal que busca modelar los gustos del n-ésimo individuo como una función flexible a partir de las características de él, su output es un vector de gustos  $\beta^{TN}$  mientras  $\omega$  corresponde a los pesos asociados a la red neuronal. Por otro lado,  $f$  corresponde a una función de utilidad determinística o paramétrica cualquiera,

que usualmente se toma lineal. Por último,  $x$  corresponde a un vector de atributos y  $z$  a un vector de características, así,  $x_{in}^{TN}$  representará a los atributos de la  $i$ -ésima alternativa que posee el  $n$ -ésimo consumidor, el supra índice representa que esos atributos serán ponderados por los estimadores obtenidos a partir de TasteNet, mientras que  $z_n^{TN}$  representa un vector de características del  $n$ -ésimo individuo que serán entregadas a TasteNet para que esta red neuronal retorne los parámetros de gustos que estima para él. Análogamente para los super índices MNL, siendo estos acrónimos del modelo Logit Multinomial, que es el escogido por los autores en este caso para obtener las probabilidades de elección de cada alternativa.

Los autores mencionan restricciones sobre los vectores  $z_n^{TN}$ ,  $x_{in}^{MNL}$  y  $z_n^{MNL}$  que fueron obtenidas en el paper de L-MNL con el objetivo de reducir sesgos en el modelo MNL, las cuales son las siguientes:

$$z_n^{TN} \cap x_{in}^{MNL} = \emptyset, \quad (2)$$

y

$$z_n^{TN} \cap z_n^{MNL} = \emptyset \quad (3)$$

Que son razonables en la modelación general, ya que la primera se tiene siempre y la segunda dice que si tenemos claro como una característica de un individuo afecta en sus gustos, entonces se la entregamos como input al MNL, en caso contrario dejamos que la red neuronal lo aprenda. Pese a ello, es una imposición que puede no ser cumplida y en ese caso, la red neuronal aprenderá los residuos de la interacción que no es capturada por la modelación hecha por el experto, lo que eventualmente generaría sesgos u overfitting.

De lo anterior, es importante notar la flexibilidad de la modelación hecha por los autores, ya que si bien ellos escogieron que los gustos sean función lineal de las características es posible escribirlo de una manera diferente acorde al problema que se esté afrontando, de la misma manera con la elección del modelo Logit Multinomial, ya que este se puede escoger libremente cualquier modelo que esté basado en el principio RUM y con ello afrontar de manera más específica el problema de elección discreta que se esté abordando.

### 1.3. Arquitectura de TasteNet

La modelación es lo suficientemente general como para que la arquitectura de la red neuronal no sea estricta, es decir, se puede escoger libremente cualquier tipo de red neuronal que se estime conveniente. En el paper los autores se apegan a lo simple y escogen la red neuronal más sencilla, que es el perceptrón multicapa (MLP, por sus siglas en inglés) o una red neuronal feed-forward.

### 1.4. Consideraciones de la modelación

En la figura (1) se muestra la arquitectura general de la modelación, notando que los autores agregan en la función de pérdida un término de regularización  $\lambda_p \|\omega\|^p$ , este cambio en la función de pérdida se conoce en la literatura como regularización, cuyo objetivo es mitigar el *overfitting* al no permitir que exista pesos muy grandes en la estructura de la red neuronal, evitando así que alguna neurona sea demasiado importante.

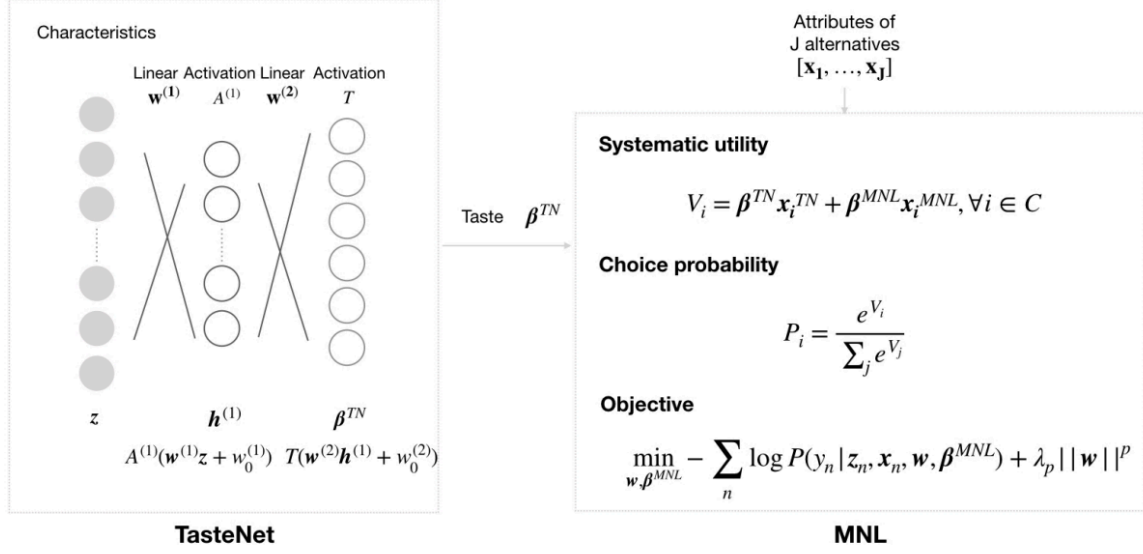


Figure 1: Diagrama de TasteNet-MNL (TasteNet como un MLP de una capa oculta)

Otra consideración importante que explican los autores, es poner restricciones en la capa de salida de la red neuronal para evitar incongruencias con la teoría y como método de regularización, proponen que los estimadores retornados por la NN tengan restricciones fijadas por quien modele el problema. Por ejemplo, digamos que  $\beta_{cost}$  es el coeficiente estimado para el costo por TasteNet, entonces es esperable que este tenga signo negativo, ya que el aumento en el costo uno esperaría que tuviera un impacto negativo en la utilidad de la alternativa, es decir, si es más caro, menos lo quiero. Por lo que para restringir que  $\beta_{cost}$  sea no positivo, se puede usar en la capa de salida de la red neuronal una función como  $-ReLU(-\beta)$  o  $-exp(-\beta)$ .

Mencionan que una ventaja de hacer esto de esta manera es que las restricciones se mantendrán estrictamente, tanto con data de entrenamiento como con data no vista, a diferencia de otros métodos que añaden una penalización en la función de pérdida cuando no se cumplen las restricciones esperadas. Sin embargo, imponer restricciones a los modelos de DCM es una decisión polémica y debe ser considerada con suma atención por la persona que modele el problema a tratar.

## **Bibliography**

[1] Yafei Han, Francisco Camara Pereira, Moshe Ben-Akiva, Christopher Zengras. (2022). A Neural-embedded Discrete Choice Model: Learning Taste Representation with Strengthened Interpretability

## **Index of Figures**

Figure 1: Diagrama de TasteNet-MNL (TasteNet como un MLP de una capa oculta) .....	5
--	---