

D. Juan V. Juan

# Modelos de regresión y AutoML

Equipo Innovación Copec



Roberto Muñoz

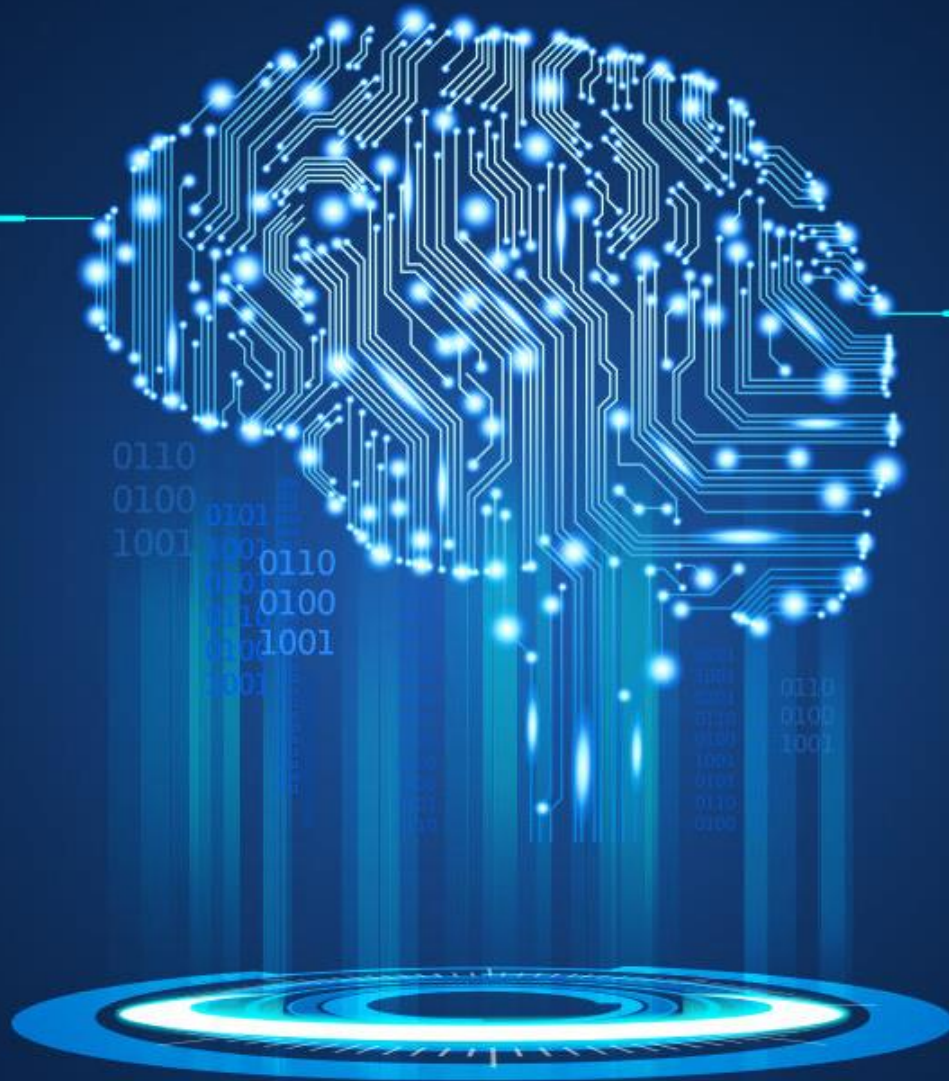
Senior Data Scientist  
Digital Data Analytics  
EY



■ The better the question. ■ The better the answer. ■ The better the world works.



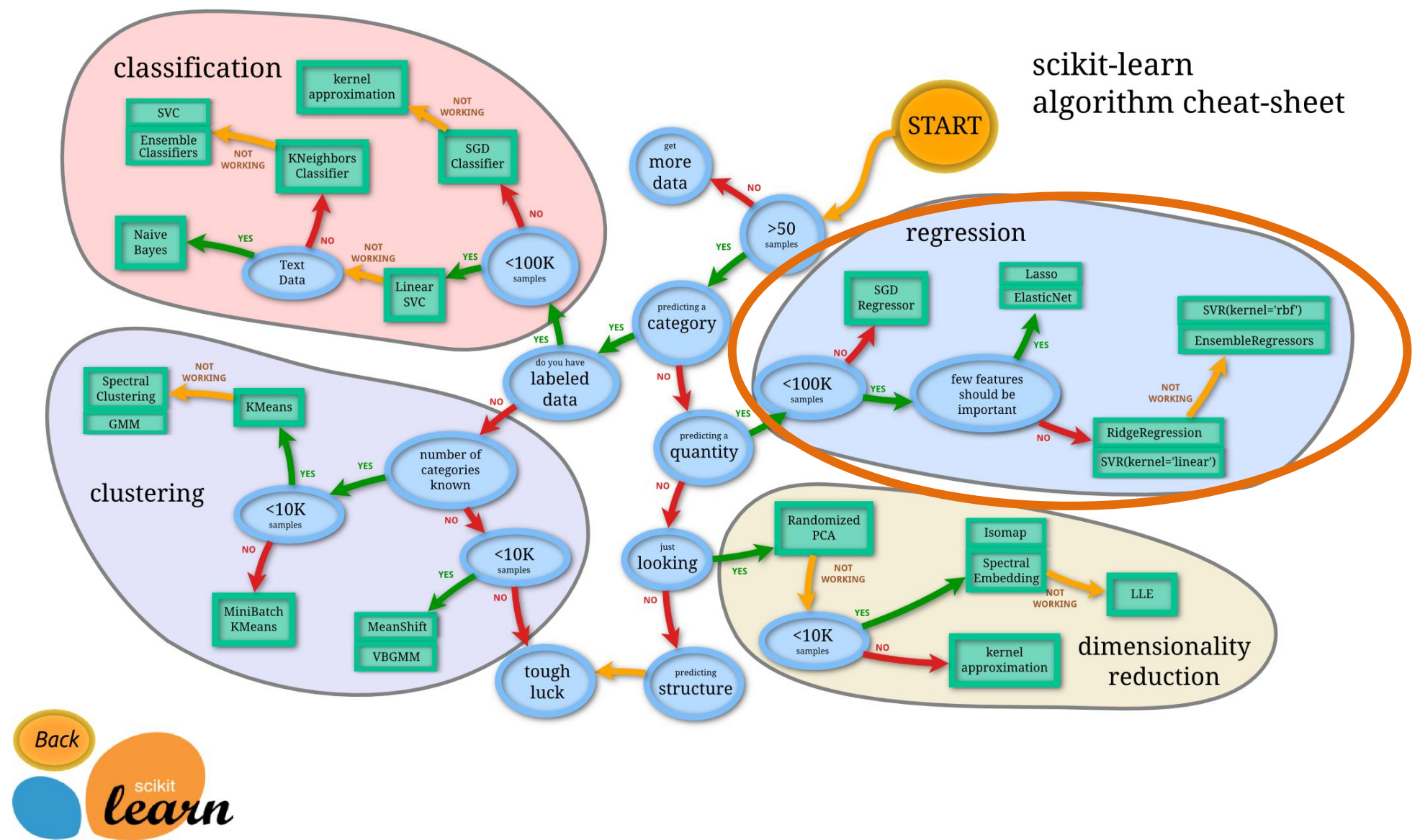
Building a better  
working world



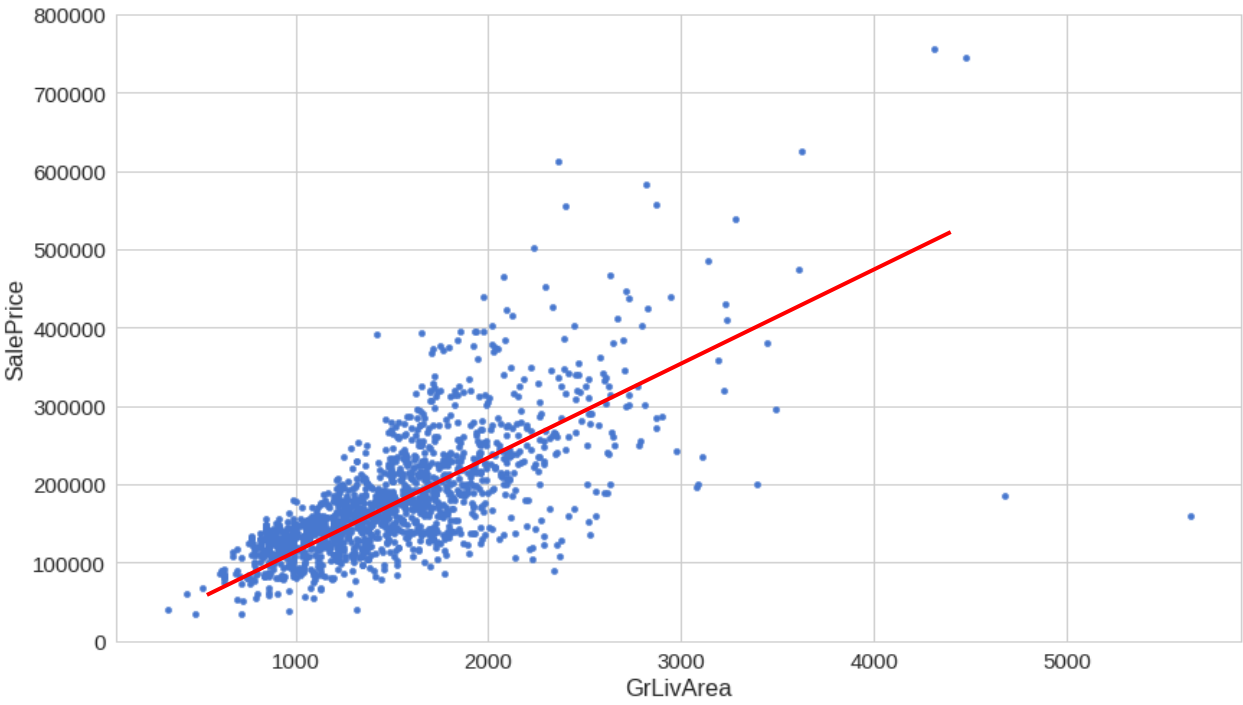
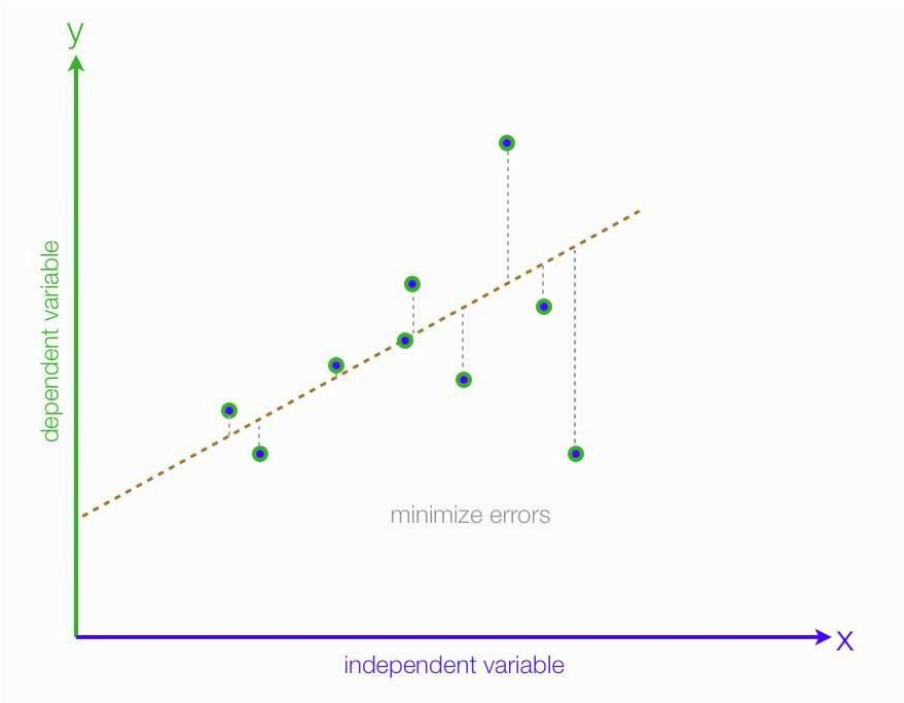
# Modelos de regresión



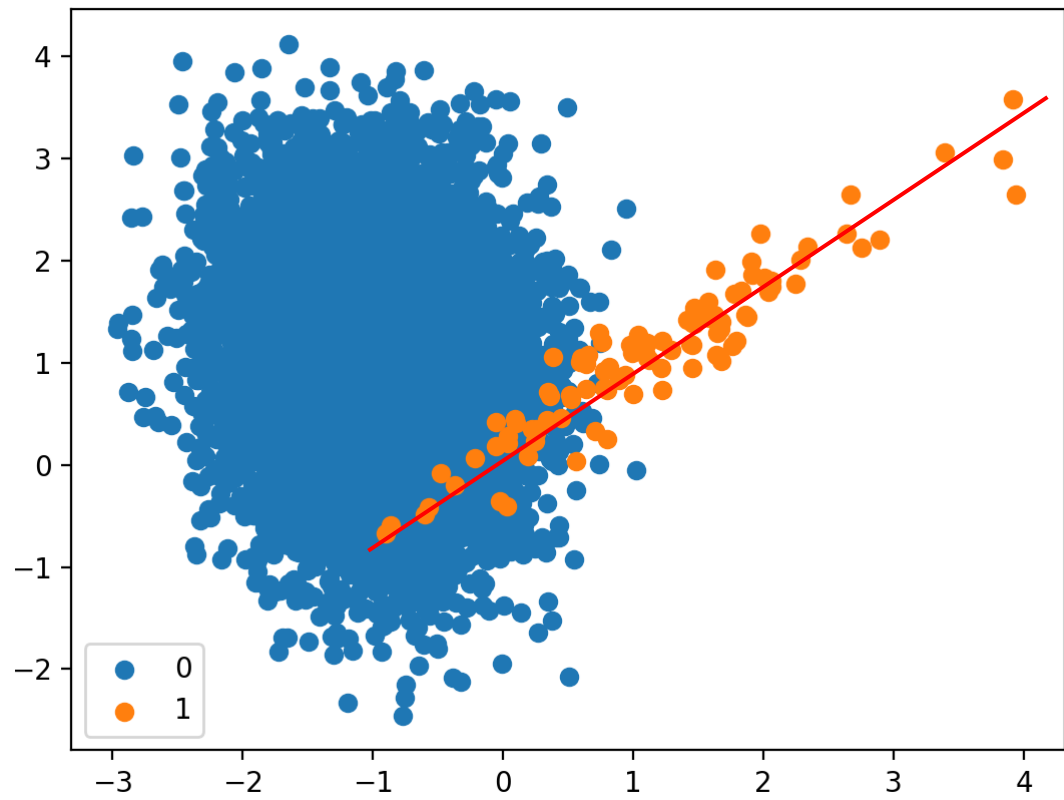
# Scikit-learn methods



# Regresión lineal



# Clustering y Regresión



1. Preparar y normalizar datos
2. Hacer clustering (K-means)
3. Identificar clusters de interés y ajustar modelo regresión (LinearRegression)

# Métricas de evaluación – Modelos de regresión

- Mean Square Error (MSE) L2

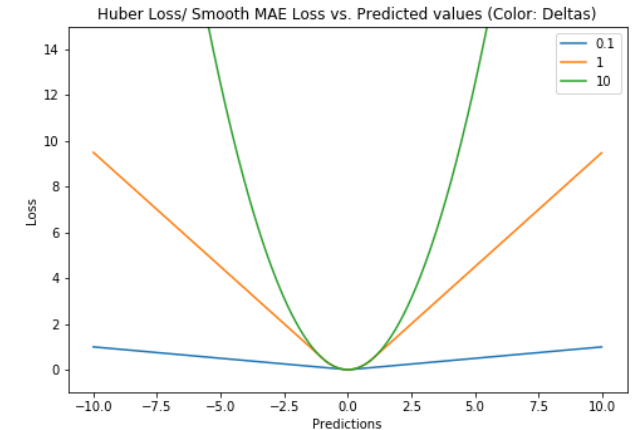
$$MSE = \frac{\sum_{i=1}^n (y_i - y_i^p)^2}{n}$$

- Mean Absolute Error (MAE) L1

$$MAE = \frac{\sum_{i=1}^n |y_i - y_i^p|}{n}$$

- Huber loss, Smooth Mean Absolute Error

$$L_{\delta}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta, \\ \delta |y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases}$$

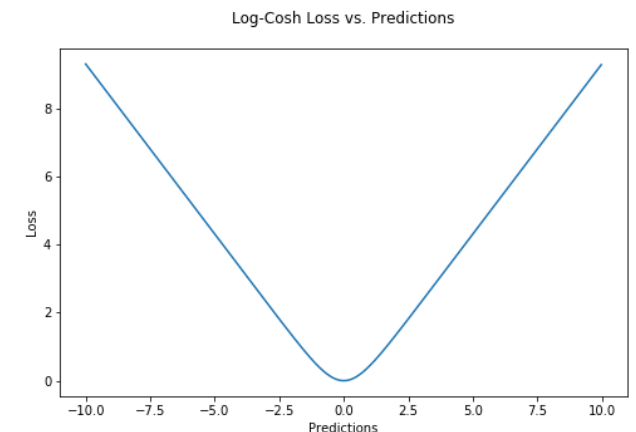


- Log-Cosh loss

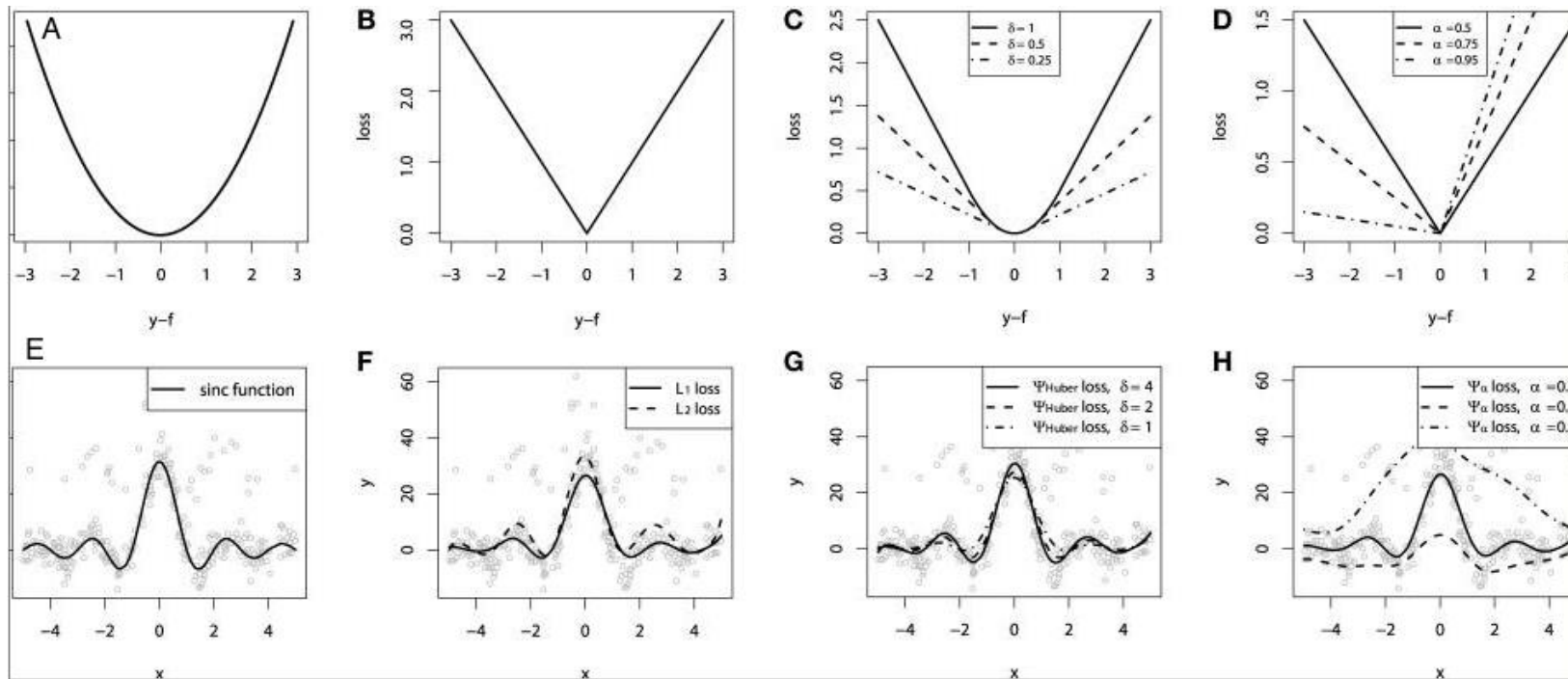
$$L(y, y^p) = \sum_{i=1}^n \log(\cosh(y_i^p - y_i))$$

- Quantile loss

$$L_{\gamma}(y, y^p) = \sum_{i=y_i < y_i^p} (\gamma - 1) \cdot |y_i - y_i^p| + \sum_{i=y_i \geq y_i^p} (\gamma) \cdot |y_i - y_i^p|$$



# Métricas de evaluación – Modelos de regresión



(A) MSE loss function, (B) MAE loss function, (C) Huber loss function, (D) Quantile loss function

Demonstration of fitting a smooth GBM to a noisy sinc(x) data:

(E) original sinc(x) function; (F) smooth GBM fitted with MSE and MAE loss; (G) smooth GBM fitted with Huber loss with  $\delta = \{4, 2, 1\}$ ; (H) smooth GBM fitted with Quantile loss with  $\alpha = \{0.5, 0.1, 0.9\}$ .



The background of the image is a dark blue field filled with a complex network of glowing nodes and connecting lines. The nodes are small spheres in white, yellow, and cyan, while the lines are thin, reddish-orange. A bright, multi-colored starburst of light is positioned at the center of the image. Faint, vertical columns of binary code (0s and 1s) are visible in the background, adding to the digital aesthetic.

# Taller