LocalGLMnet: interpretable deep learning

Mario V. Wüthrich
RiskLab, ETH Zurich

Joint work with Ronald Richman (Old Mutual Insure)

October 13, 2021
Online Joint Section Colloquium of the IAA
Regression modeling: a car insurance example

'\text{data.frame}': 678007 obs. of 13 variables:

$\text{IDpol}$ : num 1 3 5 10 11 13 15 17 18 21 . . .

$\text{Exposure}$ : num 0.1 0.77 0.75 0.09 0.84 0.52 0.45 0.27 0.71 0.15 . . .

$\text{Area}$ : Factor w/ 6 levels "A","B","C","D","E", . . .: 4 4 2 2 2 5 5 3 3 2 . . .

$\text{VehPower}$ : int 5 5 6 7 7 6 6 7 7 7 . . .

$\text{VehAge}$ : int 0 0 2 0 0 2 2 0 0 0 . . .

$\text{DrivAge}$ : int 55 55 52 46 46 38 38 33 33 41 . . .

$\text{BonusMalus}$ : int 50 50 50 50 50 50 50 68 68 50 . . .


$\text{VehGas}$ : Factor w/ 2 levels "Diesel","Regular": 2 2 1 1 1 2 2 1 1 1 . . .

$\text{Density}$ : int 1217 1217 54 76 76 3003 3003 137 137 60 . . .

$\text{Region}$ : Factor w/ 22 levels "R11","R21","R22", . . .: 18 18 3 15 15 8 8 20 20 12 . . .

$\text{ClaimTotal}$ : num 0 0 0 0 0 0 0 0 0 0 . . .

$\text{ClaimNb}$ : num 0 0 0 0 0 0 0 0 0 0 . . .

\textbf{Goal.} Find suitable regression function $\mu(\cdot)$ such that

$$x \mapsto \mu(x) = \mathbb{E}_x[Y],$$

where $x \in \mathbb{R}^q$ are the covariates (explanatory variables) describing claim $Y \sim F_x$. 
GLM and neural networks

- **GLM**: Choose strictly monotone link function \( g \) and assume

\[
x \mapsto g \left( \mu_{\text{GLM}}(x) \right) = \langle \beta, x \rangle = \beta_0 + \sum_{j=1}^{q} \beta_j x_j.
\]

That is, we have a linear function in covariate \( x \) after applying link \( g \).

- **Neural network**: Set for regression function

\[
x \mapsto g \left( \mu_{\text{NN}}(x) \right) = \left\langle \beta, z^{(d:1)}(x) \right\rangle,
\]

where \( x \mapsto z^{(d:1)}(x) \) is a neural network of depth \( d \).

- The neural network learns a new representation \( z^{(d:1)}(x) \) of the covariate \( x \). Typically, a well trained neural network outperforms a GLM.

- **Drawback.** The network solution is often not interpretable and explainable.
LocalGLMnet: definition

- **GLM:**

  \[
  \mathbf{x} \mapsto g(\mu_{GLM}(\mathbf{x})) = \langle \mathbf{\beta}, \mathbf{x} \rangle = \beta_0 + \sum_{j=1}^{q} \beta_j x_j.
  \]

- **Idea.** Let a neural network learn regression parameter \( \mathbf{\beta} = \mathbf{\beta}(\mathbf{x}) \).

- Choose a neural network of depth \( d \)

  \( z^{(d:1)} : \mathbb{R}^q \to \mathbb{R}^q, \quad \mathbf{x} \mapsto \mathbf{\beta}(\mathbf{x}) = z^{(d:1)}(\mathbf{x}). \)

- **LocalGLMnet:** Set for regression function

  \[
  \mathbf{x} \mapsto g(\mu(\mathbf{x})) = \langle \mathbf{\beta}(\mathbf{x}), \mathbf{x} \rangle = \beta_0 + \sum_{j=1}^{q} \beta_j(\mathbf{x}) x_j.
  \]
LocalGLMnet: interpretations

- LocalGLMnet:

\[ x \mapsto g(\mu(x)) = \langle \beta(x), x \rangle = \beta_0 + \sum_{j=1}^{q} \beta_j(x)x_j. \]

* If \( \beta_j(x) \equiv 0 \): drop term \( x_j \).
* If \( \beta_j(x) \equiv \beta_j (\neq 0) \): we have a GLM term in \( x_j \).
* If \( \beta_j(x) = \beta_j(x_j) \): no interactions of term \( x_j \) with \( x_{j'}, j' \neq j \).
* Interactions: study gradient

\[ \nabla \beta_j(x) = \left( \partial_{x_1} \beta_j(x), \ldots, \partial_{x_q} \beta_j(x) \right)^\top \in \mathbb{R}^q. \]

* We do not have identifiability as we may still receive

\[ \beta_j(x) x_j = x_{j'}, \]

by learning a regression attention \( \beta_j(x) = x_{j'}/x_j \) We did not encounter this.
Synthetic example

- Choose regression function

\[ \mathbf{x} = (x_1, \ldots, x_8)^\top \in \mathbb{R}^8 \mapsto \mu(\mathbf{x}) = \frac{1}{2} x_1 - \frac{1}{4} x_2^2 + \frac{1}{2} |x_3| \sin(2x_3) + \frac{1}{2} x_4 x_5 + \frac{1}{8} x_5^2 x_6. \]

- Note that \( x_7 \) and \( x_8 \) do not enter the regression function.

- We simulate Gaussian observations \( Y \) with means \( \mu(\mathbf{x}) \) and unit variance.

- We fit a LocalGLMnet for \( \beta(\mathbf{x}) = z^{(d:1)}(\mathbf{x}) \) of depth \( d = 4 \) with \((20, 15, 10, 8)\) hidden neurons.

- Fitting is done with stochastic gradient descent exploring early stopping.
Estimated \( \hat{\beta}(x) \) and variable selection

\[
\mu(x) = \frac{1}{2} x_1 - \frac{1}{4} x_2^2 + \frac{1}{2} |x_3| \sin(2x_3) + \frac{1}{2} x_4 x_5 + \frac{1}{8} x_5^2 x_6.
\]
Estimated terms $\hat{\beta}_j(x)x_j$

$$
\mu(x) = \frac{1}{2}x_1 - \frac{1}{4}x_2^2 + \frac{1}{2}|x_3|\sin(2x_3) + \frac{1}{2}x_4x_5 + \frac{1}{8}x_5^2x_6.
$$
Gradients for interactions $\nabla \hat{\beta}_j(x)$

\[
\mu(x) = \frac{1}{2} x_1 - \frac{1}{4} x_2^2 + \frac{1}{2} |x_3| \sin(2x_3) + \frac{1}{2} x_4 x_5 + \frac{1}{8} x_5^2 x_6.
\]
Variable importance: car example

\[ \text{VI}_j = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{\beta}_j(x_i) \right|. \]
Take Aways

- A neural network as an extension of a GLM (representation learning).
- Neural networks suffer the deficiency of not being interpretable.
- We have introduced the LocalGLMnet that estimates the GLM parameters through a neural network.
- This leads to interpretable results that allow for the study of variable importance.
- Variable selection can be done in LocalGLMnets.
- Interactions can be studied in LocalGLMnets.

- LocalGLMnet: interpretable deep learning for tabular data.  
  *SSRN Manuscript 3892015* (2021).