EMBEDDING LAYER AND LOCAL GLMNET FEATURE SELECTION IN ACTUARIAL SETTING

Author: Warnauts Aymeric Professor: Hainaut Donatien

aymeric.warnauts@student.uclouvain.be

Abstract

The variable selection and the training of regressors are of increasing concern in the actuarial field as regulators now impose insurance company which offer contracts to hold enough capital to fulfill the promises they sell, companies need to allocate premiums carefully. Inspired by the structure of generalized linear models, the Local GLMnet propose a new network architecture that shares similar features as generalized linear models, but provides superior predictive power benefiting from the representation learning. This architecture allows for variable selection of tabular data and for interpretation of the calibrated deep learning model.

Assumptions

The generalization of this network structure is made for GLM's wich share the properties of the exponential distribution:

$$f(y_i) = exp\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} - c(y_i)\}$$

and thus can be applied to **Poisson** distributed claim count, **Binomial** default probability or normally distributed targets with respectively $log(.)$, $log(\frac{\pi}{1-\pi})$ (logit) or identity canonical link functions. Moreover, an offset may be added to take care of exposure as in case of Poisson distributed target:

$$\theta_i = log(\mu) = log(\mathbb{E}[\mathbb{Y}]) = log(b'(\theta_i)) = log(\frac{N_i}{v_i}) = log(N_i) - log(v_i) = \sum_{j=1}^{J} \beta_{ij}X_{ij}$$

Methodology

It differs with what has been done with methods like **PDP**, **LIME**, **ALE** or **SHAP** recently as here the internal structure allows for interpreting, explaining, general interactions and benefiting from the transparent β interpretation of GLM's.

FFN extension of GLM

For the $m-th$ layer of the fully-connected feed forward neural network, raw features are first non-linearly transformed before performing the scalar product of the GLM such that $z_j^{(m)}(x) = \phi_m(w_{0,j}^{(m)} + \sum_{l=1}^{q_{m-1}} w_{l,j}^{(m)} x_l)$.

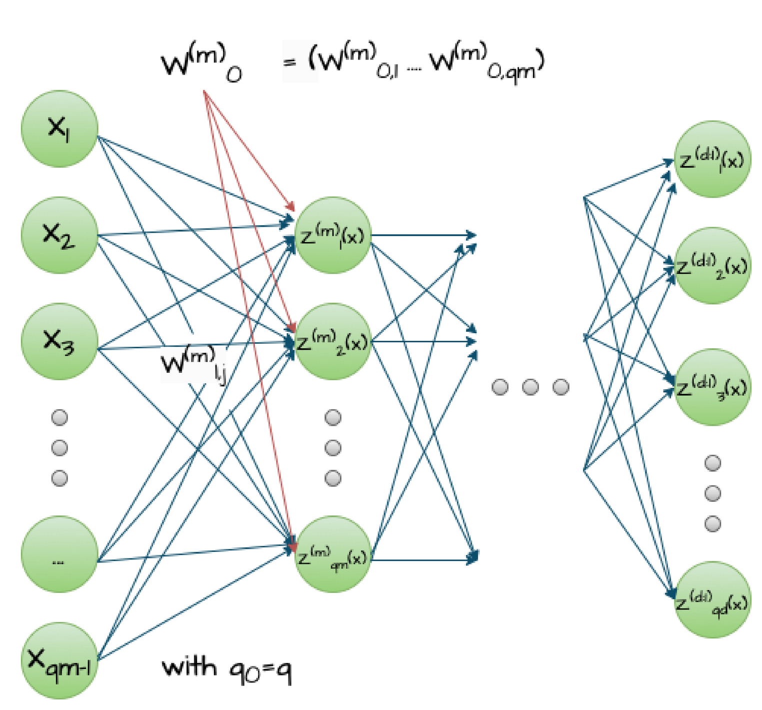

Figure 1: Fully-connected feed forward neural network architecture

For a FFN with d layers, the last layer output is given by $x \mapsto z^{(d:1)}(x) = (Z^{(d)} \circ ... \circ Z^{(1)})(x)$. And it enters the GLM providing the regression function:

$$x \mapsto g(\mu(x)) = \beta_0 + \langle \beta, z^{(d:1)}(x) \rangle$$

Local GLM network

As we face the black-box interpretation issue of deep networks architectures, the x_j influence on the response $\mu(x)$ becomes not clear. The idea will be to transform *regression parameter* β_j into network learned *regression attention* $\beta_j(x)$ such that:

$$x \mapsto g(\mu(x)) = \beta_0 + \langle \beta(x), x \rangle$$

where $\beta(x) = z^{(d:1)}(x)$ that in a small environment $\mathbb{B}(x)$ allows us to approximate $\beta(x')$, $x' \in \mathbb{B}(x)$ by a constant regression parameter leading to local GLM interpretation. Moreover, $\beta_j(x)$ may be interpreted as attention given to x_j.

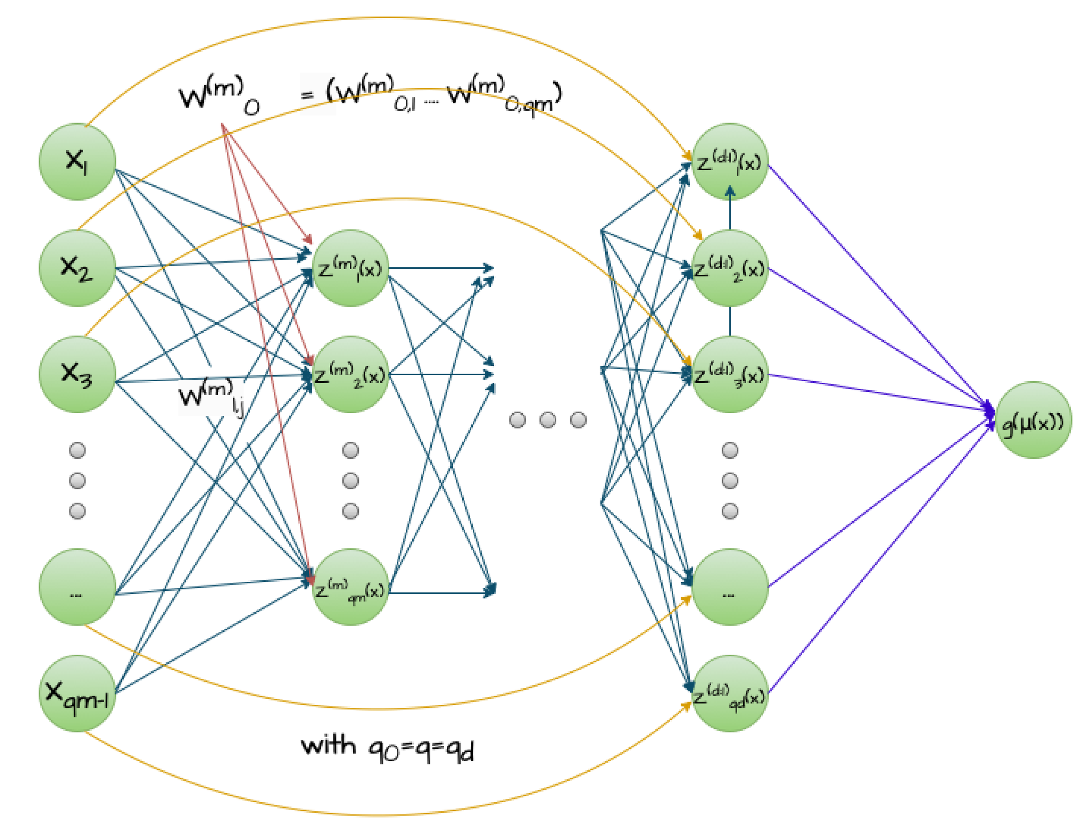

Figure 2: Deep LocalGLMnet architecture

A *LocalGLMnet layer* can be defined by replacing the scalar product by a component-wise product such that:

$$x \mapsto \beta^{(1)}(x) \otimes x = (\beta_1^{(1)}(x)x_1, ..., \beta_q^{(1)}(x)x_q)^\top$$

and finally the definition of the deep LocalGLMnet is obtained by combining such layers:

$$x \mapsto g(\mu) = g(\mu(x)) = \beta_0^{(2)} + \langle \beta^{(2)}(\beta^{(1)}(x) \otimes x), x \rangle$$

for an example with two layers.

In addition, the *LocalGLMnet* solves the explanation problem of the Shapley values extension for deep learning models as it directly postulates an additive decomposition for $\mu(.)$ after applying the link function.

Table 1: Interpretation of $\beta_j(x)x_j$

$\beta_j(x) \equiv 0$	$\beta_j(x) \equiv \beta_j$	$\beta_j(x) \equiv \beta_j(x_j)$	$\beta_j(x) \equiv x_{j'}/x_j$
Drop x_j	GLM term in x_j, not feature dependent	No interactions of x_j with $x_{j'}$ $j \neq j'$	Not full interpretability in model identifiability

Thus, the gradients $\nabla\widehat{\beta}_j(x)$ for $1 < j < q$ allows to study the derivative of regression attention with respect to x_k for fixed j with $\nabla\hat{\beta}_j(x_i)$, $1 < i < n$ with Spline fits to the sensitivities by regressing :

$$\partial x_k \widehat{\beta}_j(x_i) \sim x_{i,j}$$

[1].

Thesis proposal

This thesis will focus on the application of the *LocalGLMnet* architecture and the well known *Embedding layers* in an actuarial context. Consideration of extensions of the model to procedures such as Shapley values or Partial Dependence plots. Gradient Boodting Machine and Regression Trees based on the same computation techniques.

References

[1] R. Richman and M. V. Wuthrich. Localglmnet: interpretable deep learning for tabular data. University of the Witwatersrand, & RiskLab Department of Mathematics, ETH Zurich, 2021. arXiv:2107.11059v1.