ECE 6254 Statistical Machine Learning
Team Project Report

Chin Wei Herng, Keunhyoung Park, Ritarka Samanta, Chang Guoyun

May 2, 2024

1 Project Summary

This project aims to study the effect of noise in regression. The motivation originated from the assumption that the noise does not follow Gaussian distribution, which is generally accepted. The theoretical scenarios of noise are developed and tested through ideally generated datasets.

Methodology

A set of test is designed and performed to see whether having MSE as the objective function for regression has a limit when a PDF of noise is different from the Gaussian.

A simple function is assumed as the reference model as shown in the following. Although the study can be generalized into more complex problem, limiting the function to the first order term helps to visualize the result.

$$y = f(x) = \beta x + x_0 \tag{1}$$

The following chart represents the main process of testing the effect of an assumed noise scenario.

Objective: Study the effect of the noise in a system having pdf in regression of true model.
Step 1: Generate x
Step 2: Generate $y_o = f_o(x)$ from a true function $f_o(x) = bx + b_o$
Step 3: Generate $y = f(x)$ from a test function $f(x)$, such as $y = bx + b_o + \varepsilon$
Step 3-1: ε is generated from a selected probability model
Step 4: $g(x)$ is based on (x, y) to predict $f_o(x)$
Step 5: Evaluate the regression effectiveness of $g(x)$ by $e = g(x) - f_o(x)$

The residuals of the function, which is the noise, can have probability model such as Gaussian, Uniform, and Beta distribution. For each case, a proper objective function is considered. Thanks to the simple form of the function, the noise can be presented as the following simple equation.

$$\begin{aligned} g(\boldsymbol{Y}) = \boldsymbol{X}\beta + \varepsilon \rightarrow \boldsymbol{Y} &= g^{-1}(\boldsymbol{X}\beta + \varepsilon) \\ \rightarrow \varepsilon &= y_i - g^{-1}(x_i^T \beta) \end{aligned} \tag{2}$$

2 Cases of Probability Function

2.1 Uniform

Uniform distribution is commonly considered for the background noise (white noise) in the electric engineering field. Data measurement through the sensors contains the white noise, which affects on the precision and resolution capability of the sensor system.

$$f(\varepsilon \mid a, b) = \left[\begin{cases} \frac{1}{b-a} & \varepsilon \in [a, b] \\ 0 & otherwise \end{cases}\right] \tag{3}$$

The loss function for Uniform distribution noise is derived.

$$L(\theta \mid a, b) = \prod f(y_i - x_i\theta \mid a, b) = \prod \frac{1\{(y_i - x_i\theta) \in [a, b]\}}{b - a} \tag{4}$$

2.2 Beta Distribution Noise

Noise with Beta distribution is considered to include the skewed distribution case. Although Beta distribution noise is uncommon, it is significant to generalize the study of various noise scenarios. Furthermore, it helps to realize distinguishing noise model stemmed from a specific physic system.

$$\begin{aligned} f(y; \alpha, \beta) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)} \quad & \text{Where, Mean} : \mu = \frac{\alpha}{\alpha+\beta} \\ & \text{Variance} : \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ & B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\ & \Gamma(z) = \int_0^\infty t^{z-1}e^{-z}dt \end{aligned} \tag{5}$$

Where B and Γ are the Beta and Gamma function. Here, g^{-1} can be assumed as a logistic function for ensuring that Y will be in the range of 0 and 1. Log-Likelihood is the reasonable to be used because of the natural exponential terms in the noise distribution.

$$\begin{aligned} L(\varepsilon_i; \alpha, \beta) &= -\sum_{i=1}^{n} \log\left[\frac{(y_i - g^{-1}(x_i^T\beta))^{\alpha-1}(1 - (y_i - g^{-1}(x_i^T\beta)))^{\beta-1}}{B(\alpha, \beta)}\right] \\ &= -\sum_{i=1}^{n}\Big[(\alpha-1)\log(y_i - g^{-1}(x_i^T\beta)) + (\beta-1)\log(1 - (y_i - g^{-1}(x_i^T\beta))) - n\log B(\alpha, \beta)\Big] \end{aligned} \tag{6}$$

The loss function for Beta distribution noise is derived as the following form.

$$L(y|\alpha, \beta) = -\sum_{i=1}^{n} \log(y_i^{\alpha-1}(1-y_i)^{\beta-1}) + n\log B(\alpha, \beta) \tag{7}$$

2.3 Laplace Distribution

Laplace distribution, or Double Exponential distribution, is uded to describe a system having deviated mean in either direction with exponential decay of probabilities, which is useful for modeling changes or differences such as financial returns or signal processing.

$$
\begin{aligned}
f(x|\mu,b) = \frac{1}{2b}exp\left(-\frac{|x-\mu|}{b}\right) \rightarrow & P(n_i) = \frac{1}{2b}exp\left(-\frac{|n_i-\mu|}{b}\right) \\
& \rightarrow P(y_i|x_i;\theta) = \frac{1}{2b}exp\left(-\frac{|y_i-x_i\theta-\mu|}{b}\right)
\end{aligned}
\tag{8}
$$

The exponential term in the PDF leaded to consider the log-likelihood as the loss function.

$$
\begin{aligned}
L(\theta) = \prod_{i=0}^{n} P(y_i|x_i;\theta) \rightarrow Log(L(\theta)) &= \sum_{i=0}^{n} Log(P(y_i|x_i;\theta)) \\
&= \sum_{i=0}^{n}\left[-\frac{|y_i-x_i\theta-\mu|}{b} + log\left(\frac{1}{b}\right)\right]
\end{aligned}
\tag{9}
$$

2.4 Heteroscedastic

Heteroscedastic case is relevant for the case that varying variability at different levels or that measurement precision might depend on the size or concentration of the item being measured.

$$
Y = X\theta + N \quad \text{where} \quad N\ (0, |X|\sigma^2) \tag{10}
$$

It is similar to Gaussian distribution except the deviation of distribution function has X term in it.

$$
P(n_i) = \frac{1}{|x|\sigma\sqrt{2\pi}}exp(-\frac{1}{2}\frac{n_i^2}{|x|^2\sigma^2}) \rightarrow P(y_i|x_i;\theta) \quad = \frac{1}{|x|\sigma\sqrt{2\pi}}exp(-\frac{1}{2}\frac{(y_i-x_i\theta)^2}{|x|^2\sigma^2}) \tag{11}
$$

Log-likelihood is proper to consider the loss function due to the exponent terms.

$$
\begin{aligned}
L(\theta) = \prod_{i=0}^{n} P(y_i|x_i;\theta) \rightarrow Log(L(\theta)) &= \sum_{i=0}^{n} Log(P(y_i|x_i;\theta)) \\
&= \sum_{i=0}^{n} -\frac{1}{2}\frac{(y_i-x_i\theta)^2}{|x|^2\sigma^2} log(|x_i|\sigma\sqrt{2\pi})
\end{aligned}
\tag{12}
$$

2.5 Gaussian Mixture Models

Mixing multiple Gaussian models is also considered to cover a combined complex system.

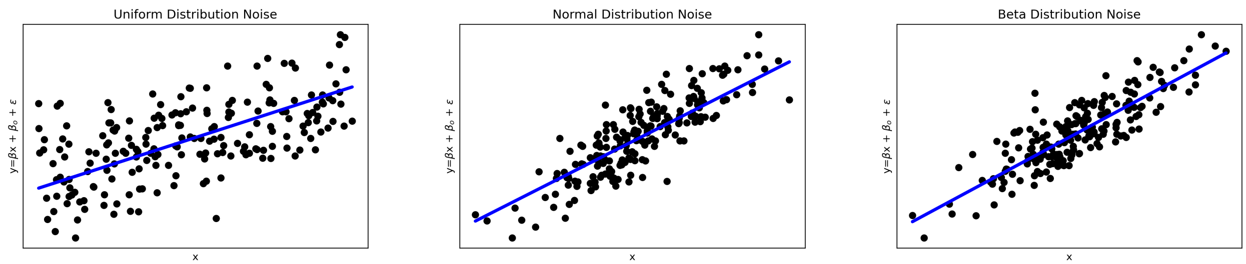
$$
P(y_i|x_i;\theta) = \frac{p}{\sigma_1\sqrt{2\pi}}exp\left(-\frac{(y_i-x_i\theta-mu_1)^2}{2\sigma_1^2}\right) \quad + \frac{q}{\sigma_2\sqrt{2\pi}}exp\left(-\frac{(y_i-x_i\theta-mu_1)^2}{2\sigma_2^2}\right) \tag{13}
$$

3 Results

Our investigations into the impact of various noise distributions on regression models yielded enlightening findings. The primary objective was to assess whether mean squared error (MSE) is effective as an objective function under non-Gaussian noise conditions.

3.1 Effectiveness of MSE Under Different Noises

Uniform distribution in noises results in noise that does not favour any particular segment of the data range over another. MSE helps in capturing the average performance of a model across the entire range of data. Meanwhile, MSE is highly sensitive to outliers in the context of beta-distributed noise. The test model has 0.3 and 2 for coefficient and intersect.



(a) Uniform Distribution Noise **(b)** Normal Distribution Noise **(c)** Beta Distribution Noise

Figure 1 Regression of the function with noises of varying PDFs

Unfortunately, the challenging regression process prevented the study from further processing. Specifically, theoretically deriving the uniform distribution noise was not resolved due to the limited form of the loss function's derivation. The limited the error domain of Beta distribution noise within the range from 0 to 1 hurt or disabled the training process. Gaussian distribution noise case suffered divergence in training for regression. Conclusively, this study was finalized by the theoretical study.

Noise PDF	MSE Loss Function		Modified Loss Function	
(Result)	(β,β_o)	(MSE)	(β,β_o)	(MSE)
Uniform	0.301833	2.009793	-	-
Normal	0.284884	2.018960	-	-
Beta	0.308963	2.859334	-	-

4 Conclusion

This project underscored the critical impact of noise distribution on the performance of regression models. The assumption of Gaussian noise, while prevalent and often effective, does not universally hold, particularly in real-world scenarios where noise characteristics may vary significantly. In conclusion, our study highlights the necessity for a critical examination of the underlying assumptions in regression analysis, especially regarding noise distribution. By embracing more versatile and adaptive approaches, the field can better accommodate the complexities of real-world data, leading to more accurate and reliable predictive models.

Chin Wei Herng and Chang Guoyun collected various material for study and studied mixed Gaussian and uniform model in addition to preparation of presentation materials. Keunhyoung Park worked on study of Beta, testing on the modified loss functionsand final report. Ritarka Samanta studied Laplace, and Heteroscedastic models, in addition to coordinating the team project activities from schedule to distributing the tasks.

5 References

Baek, Insung, Sung Jin Hwang, and Seoung Bum Kim. "CowSSL: contrastive open-world semi-supervised learning for wafer bin map." Journal of Intelligent Manufacturing (2024): 1-13.

Cherkassky, Vladimir, and Yunqian Ma. "Comparison of loss functions for linear regression." 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). Vol. 1. IEEE, 2004.

Davies, Harry J., et al. "Segmented Error Minimisation (Semi) for Robust Training of Deep Learning Models with Non-Linear Shifts in Reference Data." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

Smithson, Michael, and Jay Verkuilen. "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables." Psychological methods 11.1 (2006): 54.

Wan, Hua-Ping, et al. "Unsupervised deep learning approach for structural anomaly detection using probabilistic features." Structural Health Monitoring (2024): 14759217241226804.

Yang, Nan, Zeyu Zheng, and Tianran Wang. "Model loss and distribution analysis of regression problems in machine learning." Proceedings of the 2019 11th International Conference on Machine Learning and Computing. 2019.

https://www.stat.cmu.edu/ cshalizi/uADA/12/lectures/ch06.pdf

https://periodicos.fgv.br/bre/article/download/1515/938/2821

https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf