STA413
Assignment 2

Ngcamphalala Vuyo S.
202101291

University of Eswatini

Faculty of Science & Engineering

Statistics and Demography Department

December, 2024

Question

Use and link as many Statistical Learning Theory concepts as you know (Diffusion of Innovation of models, Information Theory, Systems' capacities, loss functions, optimization, regularization, etc.) with reality concepts (philosophical, spiritual, etc.) to decipher the events in this video. Originality, imagination, and creativity are key!

Abstract

This assignment/research seeks to explore the concepts of *inter-being* and *inter-are*; the idea that everything in the universe is interconnected. By establishing relationships between Statistical Learning Theory concepts and Reality concepts, we decode the layers of information, loss, optimization, and regularization inherent in the conflict. This approach illustrates how these principles converge to offer actionable insights into human behavior and conflict resolution. The following sources are in the form of images and dialogue to vividly showcase the events of this TikTok video.

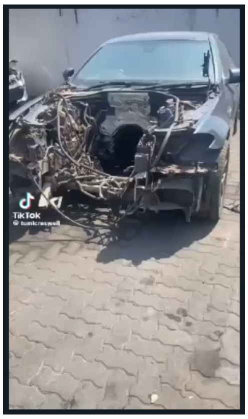


Figure 1: front of the stripped off car

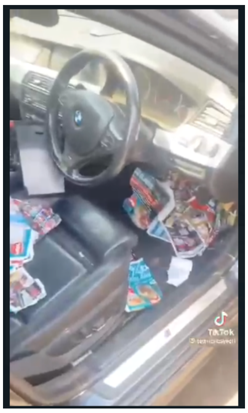


Figure 2: inside of the stripped off car

Dialogue

This is the dialogue of a dispute between a lady and a mechanic about a stripped-off car shown in the above figures that belonged to the lady.

> **Lady:** What's happening to my car, you are going to fix this car now; I want it the way it was when it came here. I want my car.
>
> **Mechanic:** Ma'am, I can't do that until we've put the fuel pump control module in.
>
> **Lady:** No, no, look, don't put anything back, just put back everything the way it was, my car wasn't like this, now it looks like a scrapyard.
>
> **Mechanic:** Ma'am, there is a procedure.
>
> **Lady:** You took advantage of us, right?
>
> **Mechanic:** No.
>
> **Lady:** You took advantage because we called...
>
> **Mechanic:** You can phone any auto electrician and they will tell you the procedure, ma'am.
>
> **Lady:** I don't want to call anybody, Philip. Listen, I don't want to call any auto, I don't want to call anything, I don't want to call no one. You are going to put back everything as it was, okay? Because several times that we called you and then you never responded. I even texted you with a different number and I said to you, now that because you know that it's me, probably you are not going to answer, and then you answered. From there, we keep on sending you messages and you ignored us. You know that you did this to our car, how can you?
>
> **Mechanic:** Because that's the procedure of finding the fault, ma'am.
>
> **Lady:** Okay, procedure or no procedure, just put back everything as it was. I'm not leaving with...

Introduction

The intricacies of human experiences and conflicts frequently reflect the frameworks and dynamics found in computational models. Statistical Learning Theory, a cornerstone of machine learning, provides a framework for understanding how systems learn from data to make predictions and decisions. Concepts such as information theory, loss functions, optimization, regularization, and the diffusion of innovation underpin computational algorithms while resonating with philosophical, spiritual, and societal dynamics (Vapnik, 1998; Shannon, 1948; Rogers, 2003).

Humans/systems need to connect all the dots, taking into consideration all intervening, collider, and antecedent variables for proper decision-making. Human interactions, at their core, are systems of communication, decision-making, and adaptation, where emotions, intentions, and external factors serve as the "data" driving outcomes. Similarly, Statistical Learning Theory addresses these dynamics by analysing how signals are transmitted in noisy environments (Shannon, 1948), how costs are minimized through loss functions, and how optimal solutions emerge within constrained systems (Hastie, Tibshirani, & Friedman, 2009). By drawing parallels between mathematical principles and the abstract realities of human existence, we uncover a unified lens to interpret conflicts, innovation, and growth.

Logistic Growth Model for Innovation Diffusion

The diffusion process is described by the following differential equation:

$$\frac{dN(t)}{dt} = rN(t)\left(1 - \frac{N(t)}{K}\right) \tag{1}$$

Where:

- $N(t)$: Number of adopters at time t
- r: Adoption rate (intrinsic growth rate)
- K: Saturation level or carrying capacity (maximum potential adopters)
- $\frac{dN(t)}{dt}$: Rate of change of adoption over time

This equation describes how the initial adoption grows exponentially but slows as the number of adopters approaches the saturation level K.

Bass Diffusion Model

Another widely used model for diffusion incorporates both innovation and imitation:

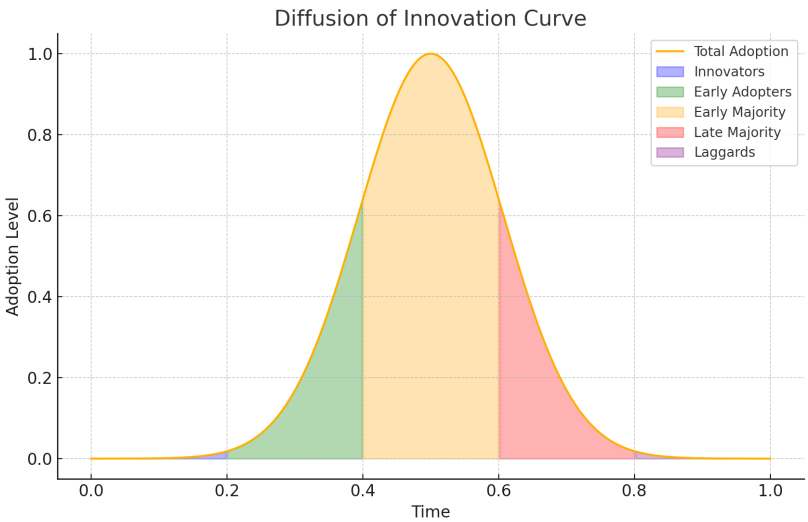
$$\frac{dN(t)}{dt} = p(K - N(t)) + q\frac{N(t)}{K}(K - N(t)) \tag{2}$$

Where:

- p: Coefficient of innovation (external influences)
- q: Coefficient of imitation (social influence)
- K: Saturation level or carrying capacity (maximum potential adopters)
- $N(t)$: Number of adopters at time t
- $\frac{dN(t)}{dt}$: Rate of change of adoption over time

This equation reflects the balance between external triggers (media, education) and internal dynamics (peer influence, community norms), emphasizing the role of human connections in spreading ideas.

This is illustrated in figure 3



If a car mechanic straps off a BMW and removes the fuel pump control module (FPCM), they are likely addressing an issue related to the fuel delivery system, which includes:

- Faulty fuel pump control module
- Fuel pump malfunction
- Electric issues
- Programming or software update
- Diagnosing a non-start condition
- Fuel system overhaul

The fuel pump control module is a critical component in modern BMWs, as it regulates the operation of the fuel pump, ensuring the engine receives the correct amount of fuel under various driving conditions. The mechanic might also strap down the vehicle to stabilize it during repair, especially if it's being worked on in a position that could make it shift or move unexpectedly.

Once the FPCM is removed and either repaired or replaced, it may need:

- **Programming:** The new or existing module might need to be coded to the vehicle's specific configuration.
- **Calibration:** Ensuring the fuel pump delivers fuel at the right pressure and flow rate.
- **System Testing:** Verifying the entire fuel system works correctly after repairs.

This process underscores the interconnected nature of modern vehicle components, particularly in high-performance cars like BMWs, Hece this is the procedure the mechanic was talking about.

The diagnosis and repair of issues related to the fuel pump control module (FPCM) in BMWs and other modern vehicles have significantly evolved over time. These improvements reflect broader advancements in automotive technology, diagnostic tools, and system designs.

Evolution of FPCM Diagnostics

Mechanical Diagnosis Era (Pre-1980s)

Diagnosis Approach:

- Early vehicles relied on mechanical fuel pumps that operated without electronic modules. Diagnosing fuel system issues primarily involved manual inspection of mechanical components like fuel lines, filters, and pumps.
- Mechanics used basic tools like pressure gauges to check fuel pressure and manually inspected fuel delivery systems for clogs or leaks.

Challenges:

- Diagnosis was time-consuming and imprecise.
- Limited ability to detect intermittent issues in the fuel delivery system.

This initial innovation can be interpreted using the General Linear Models (GLM) given by:

$$E(Y \mid X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_P X_P + e_i \tag{3}$$

The $\beta_i X_i$ are the non-somatic inputs (information accounted for the model), where specification errors happen due to the choice of distribution, measurement errors, and sampling errors.This approach assumes that the model is correctly specified and that the data follows a particular distribution, and were treated differently according to their families, these assumptions do not apply in reality. The assumptions are as follows:

Assumptions:

- **Linear assumption:** Assumes a linear relationship between predictors and the response variable, which may not capture complex relationships.
- **Normality of residuals:** Assumes errors in diagnosis are normally distributed, which is not always the case.
- **Outliers:** Sensitive to outliers, which could disproportionately influence results.
- **No Multicollinearity:** The issue of this car my be a combination of highly correlated predictor variables but this model assumes that it does bot exist.

Introduction of Electronic Fuel Injection (1980s–1990s)

Advancements:

- Electronic Fuel Injection (EFI) systems introduced electrically powered fuel pumps and control modules.
- Diagnostics became more complex, requiring a basic understanding of electronic systems.

Tools and Methods:

- Mechanics began using multimeters to test electrical continuity and voltage in fuel pump circuits.
- Fault tracing became more systematic but still relied on trial and error.

Improvements:

- Faster and more accurate fault detection compared to purely mechanical systems.
- Introduction of basic onboard diagnostics (OBD-I) allowed for electronic fault codes.

The Generalized Linear Models (GLM) approach was introduced:

$$g(E(Y \mid X)) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + e_i \tag{4}$$

where $g(.)$ is a link function (e.g., log link for specific distributions).

Poisson Distribution ($Y_i \sim$ Poisson(θ))

For a Poisson-distributed response variable with mean θ, the log-link function is given by:

$$\log(E[Y_i]) = \log(\theta).$$

Binomial Distribution ($Y_i \sim$ Binomial(n, p))

For a Binomial-distributed response variable with size n and probability p, the log-link function for the expected value is:

$$\log(E[Y_i]) = \log(n \cdot p),$$

where $E[Y_i] = n \cdot p$.

Normal Distribution ($Y_i \sim N(\mu, \sigma^2)$)

For a Normally distributed response variable with mean μ and variance σ^2, the log-link function for the expected value is:

$$\log(E[Y_i]) = \log(\mu).$$

It is worth noting that the generalized linear models are an extension of general linear models by allowing the response variables to have other distributions rather than the normal distribution thus not treating the distributions separately. Thereafter, the Generalized Additive Model (GAM) was introduced to cater for non-linearity and loss of information in data processing estimation

Integration of Onboard Diagnostics (OBD-II) (1996–2000s)

Advancements:

- OBD-II standard provided access to diagnostic trouble codes (DTCs).
- BMWs included more sophisticated FPCM systems.

Tools and Methods:

- OBD-II scanners could read error codes related to the fuel pump and its control module.
- Mechanics could identify problems like low fuel pressure, circuit faults, or communication errors with the ECU.

Improvements:

- Standardized diagnostics reduced guesswork.
- Improved ability to detect and diagnose intermittent electrical faults.

The Generalized Additive Model (GAM) is expressed as:

$$g(E[Y \mid X]) = \alpha + f_1(x_1) + \cdots + f_p(x_p) + \epsilon_i,$$

where:

- $g(\cdot)$ is the link function.
- $E[Y \mid X]$ is the expected value of the response variable Y given the predictors $X = (x_1, \ldots, x_p)$.
- α is the intercept.
- $f_i(x_i)$ are smooth, non-parametric functions of the predictors x_i, capturing the structure or trend in the data.
- ϵ_i represents the residual error.

Key Characteristics of GAM

- The theory of GAM emphasizes the distribution of the response variable rather than the individual data points.
- The functions $f_i(x_i)$ are non-parametric, meaning they are not restricted to a specific parametric form. Instead, they rely on the observed structure or trend in the distribution.
- This flexibility allows GAMs to model complex relationships between the predictors and the response variable without assuming a linear or predefined form.

Machine Learning and Deep Learning

ML and DL models are an extended mind hence they mimic the process of the human brain. We look at the concept of biomimicry to understand how the brain processes information from 1 neuron to another which corrects for errors, optimize and update which leads to proper decision making in the procedure of fixing the car.

Artificial Neural Network

A fit-forward neural network with at least one hidden layer containing a finite number of neurons that can approximate any continuous function on compact subsets in the space of any degree of accuracy thus it handles high dimensional data and learn complex relationships between the possible faults of the car. This is done through their layered architecture and it performs well to unseen data like in this case.

Modern Diagnostic Innovations (2020s–Present)

Advancements:

- Use of Artificial Intelligence (AI) and Machine Learning in predictive diagnostics.
- Implementation of cloud-based diagnostics for remote access.

Tools and Methods:

- Advanced scan tools integrating with BMW systems for live data streams.
- Automated diagnostic systems for comprehensive testing of FPCM.

Improvements:

- Real-time remote diagnostics minimize downtime.
- Predictive maintenance reduces the likelihood of catastrophic failures.

Future Innovations

- **IoT-Enabled Diagnostics:** Continuous monitoring of FPCM and other systems.
- **Self-Healing Systems:** Adaptive control systems to stabilize issues temporarily.
- **Augmented Reality (AR) Diagnostics:** Visualizing the internal state of components in real-time.

Philosophical Perspective

Equation 1 mirrors how awareness of an idea spreads in a society. The term $rN(t)$ reflects early adopters' enthusiasm, while $\frac{1-N(t)}{K}$ represents societal constraints, such as skepticism or resource limitations. As with the spread of new ideas in philosophy or spirituality, there is a tipping point where resistance diminishes, leading to widespread acceptance.

$$rN(t) \quad \text{reflects the early adopters' enthusiasm}$$
$$\frac{1-N(t)}{K} \quad \text{represents societal constraints, like skepticism or resource limitations}$$

Spiritual Perspective

The sigmoid adoption curve derived from the logistic model is:

$$N(t) = \frac{K}{1 + e^{-r(t-t_0)}}$$

Here, t_0 is the inflection point. This curve reflects spiritual journeys where individuals progress from initial ignorance to eventual enlightenment, resembling the phases of innovation adoption: innovators, early adopters, majority, and laggards.

Information Theory and Systems Capacity

Information about the truth is encoded in probability functions or wave functions, and once a system accounts for all information, there will be no variability. In this scenario, the mechanic and the lady represent two different systems with different information, leading to the conflict observed. The different energies propagate different wavelengths, and when the frequencies interact, they create different formations, either disruptive or constructive.

Using information theory, the spread of innovation can be quantified by Shannon entropy:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

or equivalently:

$$H(X) = -\int f(x_i) \log_2 f(x_i)\, dx_i$$

Where $P(x_i)$ and $f(x_i)$ represent the probability of individuals in the population adopting a specific innovation. A decrease in entropy over time corresponds to the growing consensus or widespread adoption of the innovation.

It follows from the Data Processing Inequality (DPI):

$$I(\theta_i; f(x)) \geq I(\theta_i; x_1, x_2, \ldots, x_n) \geq I(\theta_i; \overline{X})$$

This inequality states that conditioning on a random variable can only decrease entropy and increase probability.

In the video, the communication between the lady and the mechanic, denoted by $H(X)$, is disrupted because the mechanic is not given the chance to respond. This leads to high entropy (uncertainty). This can be interpreted as a loss of mutual information, represented by the Kullback-Leibler divergence (KL) between the distributions of the lady's expectation and the mechanic's actions:

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{i=1}^{n} P(x_i) \log \left(\frac{P(x_i)}{Q(x_i)} \right)$$

where:

- P is the distribution of the lady's expectations.
- Q is the distribution of the mechanic's actions.

One can also say that the mechanic and the lady have different priors. The mutual information between the lady's expectation X and the mechanic's actions Y can be defined as:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

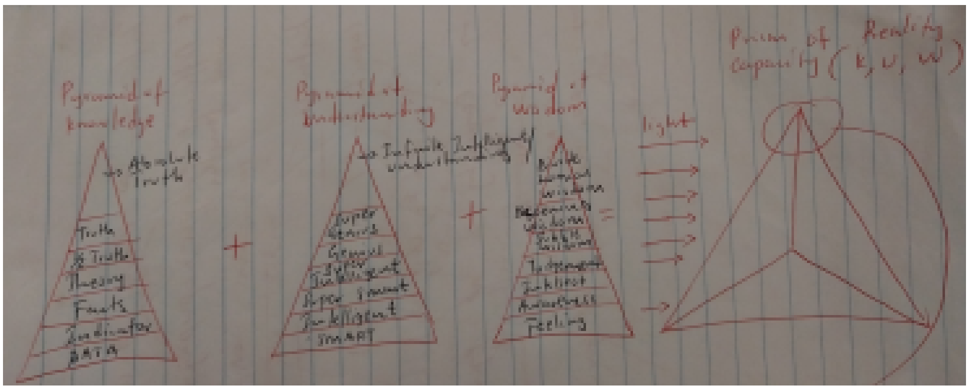
where:

- $H(X)$ is the entropy of the lady's expectations.
- $H(Y)$ is the entropy of the mechanic's actions.
- $H(X,Y)$ is the joint entropy of both.

If the mutual information $I(X;Y)$ is low, it means that the mechanic and the lady are not aligned in their understanding. As clearly shown in the video, the mutual information is very low. The high entropy in the mechanic's actions leads to a lower mutual information with the lady's expectations.

Different systems/humans have different capacities which assist them to connect all the dots in reality and it is related to quantum fields that help connect variables and the different capacities limit us from knowing the truth, hence only the supreme being knows the truth. Capacity is a function of Knowledge, Understanding and Wisdom:

$$\text{capacity} = f(K_i, U_i, W_i)$$



The somatic, extra-somatic, non-personal extra-somatic sensors help us in measurements as we try to account for more information about the truth. The lady relies only on somatic sensors to make the conclusion that the mechanic is not doing justice in fixing her car, while the mechanic uses all the somatic, extra-somatic and non-personal extra-somatic sensors as he accounts for more information since he is also a specialist in the field.

The capacity of the lady is at equilibrium, for instance:

$$f(K_2, U_2, W_2)$$

There is no curiosity to change, no imagination, no intuition, and no one can tell her otherwise (stubborn) since when the mechanic tries to advise her on calling other auto mechanics, she refuses. Therefore, she now believes that the mechanic is taking advantage of them and that sustains her belief, which is pretty much the same as religion, where the people cannot be convinced otherwise.

Understanding is genetic, since the lady keeps saying "us," it is most likely that she is referring to a family member that has the same understanding as her, and therefore they cannot be wise or knowledgeable beyond their understanding. To reach the peak in the prism of reality (Capacity), we need to do it collectively; hence, we all need to contribute to the image of God.

Loss Functions

A loss function quantifies the error between the predicted values $\hat{y}$ and the true or actual values y in a model. It is essential in determining how well a machine learning model performs. The goal of training a model is to minimize this loss function. In the case of the mechanic and the lady, the loss can represent emotional dissatisfaction or unmet expectations.

Mathematically, a general loss function can be written as:

$$L(y, \hat{y}) = \text{cost of discrepancy between true value } y \text{ and predicted value } \hat{y}$$

Where:

y represents the true value (e.g., the lady's ideal repair time or cost).

$\hat{y}$ represents the predicted or actual value (e.g., the actual time or cost the mechanic provides).

The loss function expresses the cost associated with incorrect predictions, and in the conflict scenario, this would be the emotional and practical "cost" of unmet expectations.

Common Types of Loss Functions

Squared Error Loss (L2 Loss)

$$L_{\text{sq}} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

This is the most common loss function in regression problems. The squared error penalizes large errors more significantly, which may be used to model situations where larger discrepancies (e.g., higher costs or longer repair times) are more painful or disruptive for the lady.

Absolute Error Loss (L1 Loss)

$$L_{\text{abs}} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

This loss function is less sensitive to outliers and can be used to model situations where small errors are not significant, but the emotional loss from a mismatch is still important.

Huber Loss (Combination of L1 and L2 Loss)

$$L_{\text{Huber}} = \frac{1}{n}\sum_{i=1}^{n} \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{for } |y_i - \hat{y}_i| \leq \delta \\ \delta\,|y_i - \hat{y}_i| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$

The Huber loss combines the properties of both L1 and L2 losses. It is more robust to outliers (when large discrepancies occur) while still penalizing smaller errors efficiently. This could represent a conflict where occasional large misunderstandings (e.g., pricing or timing issues) should not dominate the emotional loss but still contribute to the overall dissatisfaction.

Real-Life Example in Conflict

In the mechanic-lady conflict, the mechanic's failure to meet the lady's expectations results in a "loss" that can be quantified. For example, if the lady expected the car repair to be finished in 2 hours ($y_{\text{lady}} = 2$ hours), but it took 5 hours ($\hat{y}_{\text{lady}} = 5$ hours), her loss might be:

$$L_{\text{lady}} = |2 - 5| = 3 \text{ hours of delay}$$

Similarly, the mechanic may have expected full payment for a high-quality repair, but if the lady refuses or negotiates for a discount, the mechanic's loss could be:

$$L_{\text{mechanic}} = |y_{\text{mechanic}} - \hat{y}_{\text{mechanic}}|$$

where y_{mechanic} is the expected payment and $\hat{y}_{\text{mechanic}}$ is the actual payment received.

Optimization: Minimizing the Loss Function

Definition

Optimization refers to the process of adjusting the parameters of a model to minimize a given objective function—in this case, the loss function. Optimization is central to finding the best possible outcome (or "solution") that balances the interests of both parties in the conflict. In machine learning, optimization algorithms like gradient descent are used to minimize the loss function by iteratively adjusting the model parameters.

For the mechanic and the lady, the optimization process represents the dynamic adjustments both parties make in their expectations and behavior to minimize emotional dissatisfaction (loss). This could involve the mechanic adjusting repair strategies, timeframes, or costs, and the lady adjusting her expectations regarding these factors.

Gradient Descent

The most common method of optimization is gradient descent, which updates the parameters (or expectations in this case) in the direction of the negative gradient of the loss function. For a general loss function $L(\theta)$, the update rule for gradient descent is:

$$\theta(t+1) = \theta(t) - \eta \nabla_\theta L(\theta)$$

where:

$$\theta(t) \text{ is the parameter (expectation) at time step } t,$$

$$\eta \text{ is the learning rate (step size)},$$

$$\nabla_\theta L(\theta) \text{ is the gradient of the loss function with respect to } \theta.$$

In the context of the mechanic and the lady, both parties "learn" from each interaction, gradually adjusting their expectations and behaviors in an attempt to minimize the total emotional loss.

Stochastic Gradient Descent (SGD)

In a real-world scenario, both the mechanic and the lady are likely to interact in many small steps (conversations, decisions, etc.), each of which provides partial feedback. Stochastic Gradient Descent (SGD) applies optimization iteratively, using small batches of feedback at each step.

$$\theta(t+1) = \theta(t) - \eta \nabla_\theta L_i(\theta)$$

where $L_i(\theta)$ refers to the loss calculated from one small piece of feedback. Thus, each small interaction may slightly adjust their expectations, allowing them to find a middle ground that minimizes the emotional loss over time.

Regularization: Preventing Overfitting and Over-Attachment

Definition

Regularization refers to techniques used to prevent the model from fitting the data too closely, which can lead to overfitting. In statistical learning, overfitting occurs when the model learns noise or minor fluctuations in the training data, rather than the underlying patterns. This causes the model to perform poorly on unseen data.

In the context of the conflict, overfitting would correspond to over-attaching to rigid expectations. If either the mechanic or the lady is too inflexible, it would lead to a conflict that is hard to resolve. Regularization helps both parties remain flexible and adaptive in their expectations.

L2 Regularization (Ridge Regression)

One common form of regularization is L2 regularization, which penalizes large weights (extreme expectations) by adding a penalty term to the loss function. The regularized loss function becomes:

$$L_{\text{reg}} = L(y, \hat{y}) + \lambda \|\theta\|_2$$

where:

$\|\theta\|_2$ is the sum of squared values of the model parameters (or expectations),

λ is the regularization parameter controlling the strength of the penalty.

In the conflict scenario, this regularization term can be seen as a way to prevent either the mechanic or the lady from over-committing to rigid expectations.

L1 Regularization (Lasso)

Alternatively, L1 regularization (Lasso) can be used to encourage sparsity, i.e., reducing the complexity of the model (or expectations) by forcing some weights to zero. This might correspond to simplifying the expectations in the conflict:

$$L_{\text{reg}} = L(y, \hat{y}) + \lambda \|\theta\|_1$$

where $\|\theta\|_1$ is the sum of absolute values of the parameters (expectations).

This would lead to fewer or simpler expectations from both the mechanic and the lady, preventing over-complication in their communication and facilitating a resolution.

Elastic Net Regularization

A combination of L1 and L2 regularization, called Elastic Net, can be used to balance the benefits of both types of regularization:

$$L_{\text{elastic}} = L(y, \hat{y}) + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta\|_2$$

This regularization allows for both flexibility and simplicity in their expectations and interactions.

Conclusion

By intertwining Statistical Learning Theory with philosophical and spiritual ideas, we see this dispute not merely as a conflict but as a complex optimization problem. The key lies in balancing emotional and material loss, reducing noise, and finding equilibrium through regularization (forgiveness) and diffusion of innovative solutions (mediation).

The car strapped off or otherwise is both the problem and the metaphorical "dataset" that challenges the capacity of the human system to learn, adapt, and grow

References

- Vapnik, V. (1998). Statistical Learning Theory. Wiley-Interscience.
- Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423.
- Rogers, E. M. (2003). Diffusion of Innovations (5th ed.). Free Press.
- Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.
- Arora, S. (2019). Theory of deep learning: Highlights of progress and a roadmap. Proceedings of the National Academy of Sciences, 116(44), 22064–22072.