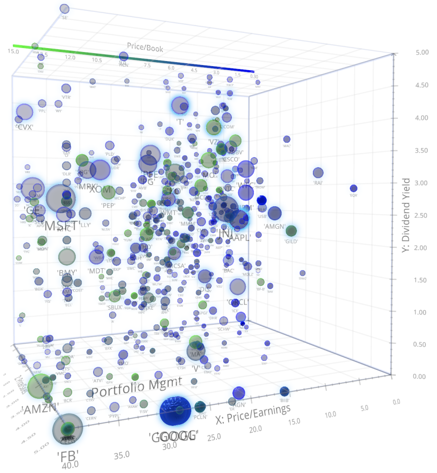
Immersion Analytics

Introduction

In an era where Artificial Intelligence (AI) increasingly influences most sectors of the economy, the need for effective AI governance is paramount. As models grow increasingly sophisticated, so too does the complexity of data they process and generate, typically multidimensional in nature. Despite this, understanding of such data and models today remains distorted by summary statistics, pair plots, and data tables. This is a critical root cause impeding development of safe and effective AI aligned with human needs. It amplifies the need for tools that enable diverse stakeholders to productively engage in development and governance of such models.

Our Innovative Solution

Our patented visualization technology simplifies otherwise complex, multidimensional data. Rendering up to 18 dimensions is accomplished by layering visual effects such as glow, pulsation, shimmer and translucency onto each data point; intensity of each effect visually encoding an added numeric dimension. We refer to this rendering technology as the Dimensional Engine™. It's made intelligible by incrementally layering dimensions as effects, one at a time, in a process we call Stepwise Storytelling™. This is supported by first principles:



1. Cognitive Load Theory - Introducing too much complexity at once overwhelms cognition. Stepwise introduction regulates cognitive load for more effective comprehension.
2. Spiral Learning - Building understanding progressively in cycles allows concepts to be scaffolded and integrated deeper over time.
3. Gestalt Psychology - Harnessing innate human perception tendencies, this approach emphasizes the use of naturally intuitive visual patterns. Gestalt psychology suggests that the human mind prefers to perceive a whole rather than disparate parts.

Other explainable AI (XAI) tools focus on interpreting and describing model logic after creation, treating symptoms reactively by debugging already-developed models, or using secondary AI for model interpretation.

In contrast, our solution enables more complete understanding before model development even begins, and throughout the process. This fosters development of inclusive, safe, and effective AI. It facilitates understanding beyond just AI experts to notice and address anomalies and bias.

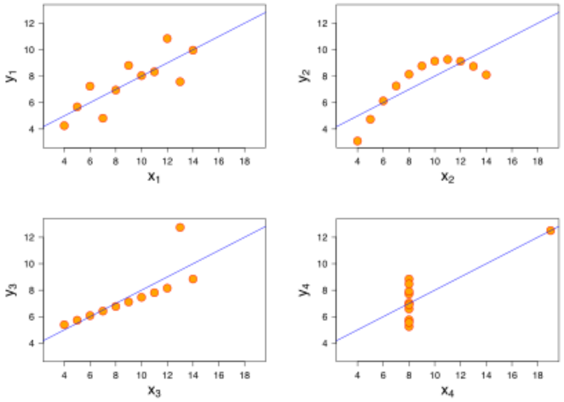
By visually highlighting when models perform poorly on specific outliers, you transcend average-case XAI to see, discuss and proactively mitigate corner cases that matter. Consider the loan applicant, denied credit due only to a faulty model, or an autonomous military drone mistaking a schoolhouse as a valid target. In developing AI, mastery of corner cases can mean the difference between life and death.

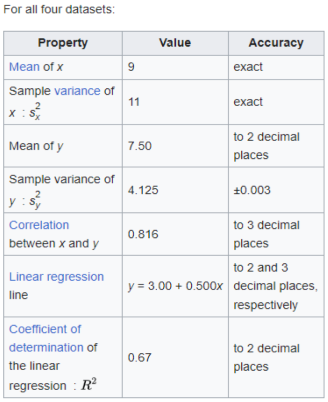
"Everything should be made as simple as possible, but no simpler." – Albert Einstein

Current State of the Field and Limitations

The current state of AI has seen substantial advancements, with a growing dependency on complex models which process multidimensional inputs. Traditionally, AI professionals rely on:

- **Statistical Analysis**: Statistical properties of model outputs may offer insight into the behavior of models. Alas, summary statistics yield an incomplete view. Figure - Anscombe's quartet:





For all four datasets:

Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x : s_x^2	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y : s_y^2	4.125	±0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression : R^2	0.67	to 2 decimal places

All four datasets are identical when examined using summary statistics, yet vary considerably when graphed.

- **Data Visualization**: Visualizing data and the results of models is a key way to understand what a machine learning algorithm is doing.
  - Pair plots to survey the unordered combinations of data dimensions, the number of such combinations expands geometrically as dimensions are added.
  - Dimensionality reduction techniques like PCA and t-SNE, though useful for clustering, can obscure interpretability by merging multiple variables, such as governance factors, into fewer axes. This risks conflating distinct aspects like hate speech and harassment into a single axis, thus hindering stakeholders from comprehending individual variables.

- **Cross-validation**: Cross-validation evaluates machine learning model performance and generalizability using tools like confusion matrices, ROC curves, AUC, precision-recall curves, learning/validation curves, feature importance, box plots, violin plots, and heatmaps.

- **Feature Importance Analysis**: By evaluating the importance of features in a model, we can gain understanding of which parts of the data are driving the model's decisions. An ability to see higher dimensional data may make the influence of individual features on AI model decisions more accessible and comprehensible to a diverse range of users.

- **Model Interpretation Tools**: Several tools and techniques can help in interpreting model predictions. Examples include SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), PDP (Partial Dependence Plots), and ALE (Accumulated Local Effects). The ability to visualize higher-dimensional data may e.g.
  - Complement SHAP by providing a clearer picture of how features interact and influence predictions, something that is not always apparent from SHAP values alone.
  - Impart a more comprehensive understanding of the model's behavior by visually representing both local explanations (from LIME) and broader data patterns.
  - Compare and contrast explanations from LIME across different models, helping to identify consistencies and discrepancies in a more intuitive manner.
  - Providing insights into complex relationships that are not easily captured by PDPs.
  - Improve accessibility and understanding from ALE plots, especially for non-experts.

- **Sensitivity Analysis**: Sensitivity analysis involves changing the input variables in various ways to see how the output of a model changes. Seeing higher dimensional data may complement sensitivity analysis by providing more intuitive and detailed visual representations of how changes in input variables affect model outputs.
- **Ensembling**: Ensembling methods, such as stacking or bagging, may provide insights by looking at it from multiple perspectives. Intuitive multidimensional visualization capabilities may
  - Enhance the comparative analysis of various models, facilitating the exploration of model diversity and correlation, essential for assembling robust ensembles.
  - Render insights into model behavior and effectiveness by seeing how various models process each data point, in the context of key features across all data points.
  - Improve predictive accuracy and also ensure creation of more balanced, effective and well-adapted ensemble models.
- **Disentangling Causality**: This is about understanding which variables have causal effects on the outcomes. Various techniques like Do-calculus, Causal Graphical Models, or Interventional Perturbation are used for this. Adaptations of the proposed visualizations may more clearly illustrate the relationships and potential causal connections between variables.
- **Counterfactual Explanations**: This involves understanding model decisions by asking "what if" questions. By slightly changing the input and observing the output, we can gain insights into the model's decision-making process. The proposed visualizations may more effectively illustrate the outcomes of "what if" scenarios, showcasing how slight changes in inputs impact model outputs.
- **Understanding Bias and Fairness**: It is important to understand and evaluate the bias in the data and the predictions made by the model. Tools like Fairlearn and AI Fairness 360 provide metrics and methods to audit and mitigate bias and discrimination in the models. The proposed visualizations enhance the ability to leverage domain knowledge to identify and comprehend complex patterns of bias, supporting more effective auditing and mitigation efforts.
- **Debugging ML models**: Debugging tools like TensorBoard, What-If Tool, or Explainable Boosting Machines (EBMs) provide interfaces to understand and debug ML models. Our visualizations render a unique visual lens for seeing how the model functions against each data point, at each epoch, without relying on summary statistics.

Visualization aspects of current techniques are limited by the conventional X vs Y (and perhaps Z) plot, hence presenting an incomplete and/or distorted view. As explained above, this poses significant challenges in understanding and analyzing higher-dimensional data underpinning models. Lack of adequate data visualization has become a major bottleneck and blindspot. Our solution fills this widening gap to mitigate potential threats as AI advances rapidly.

Solutions

Machine Learning Solution

These enable stakeholders to gain deeper insight into complex AI models, aiding in bias detection, model optimization, and inclusive governance by engaging diverse stakeholders.

Use Cases

1. **Enhanced Predictive Analytics**: The ability to improve accuracy of predictive models by enhanced Exploratory Data Analysis (EDA), Data Quality and Model Selection & Ensembling features underpins the ability to forecast and make informed decisions. Whether it's predicting maintenance needs, consumer behavior or health outcomes, this plays a key role in strategic planning and operational efficiency.

2. **Efficient Hyperparameter Tuning**: This use case is critical because it directly impacts the effectiveness and efficiency of AI models. By reducing time and computational resources needed for tuning, this capability not only accelerates the model development process but also enables creation of more accurate and reliable models.
3. **Real-Time Performance Monitoring**: The ability to continuously monitor and evaluate the performance of AI systems in real-time, as enabled by the Production Monitoring feature, is essential for maintaining the reliability and effectiveness of these systems. This use case is particularly important because it ensures that AI applications remain functional, accurate, and efficient over time, adapting to new data and conditions. This ongoing monitoring is crucial in scenarios where AI systems control critical processes or make important decisions, ensuring that any deviations or anomalies are quickly identified and addressed.

Capabilities

1. **Exploratory Data Analysis (EDA)**: This solution offers intuitive visualizations enabling you to explore data before building AI models, allowing you and your team to apply your domain knowledge to understanding data patterns and trends. By seeing the data first-hand, you design models informed by your expertise and context. This enhances model relevance and effectiveness by ensuring it's grounded in a deep, human-informed understanding of the underlying data.
2. **Data Quality**: Novel visualization enables your domain experts to leverage tacit knowledge at scale for identifying subtle, context-specific issues and anomalies that may be missed by standard automated checks. This leverages domain expertise for ensuring models are trained on data that is not only technically sound but also contextually accurate and relevant.
3. **Model Training & Debugging**: The solution enables visualizing the model training process, allowing you to simultaneously observe the model's behavior and output across all data points and key dimensions for both training and test datasets at each epoch. It renders a comprehensive view that illuminates the model's interaction with the entire dataset over time, facilitating a deeper understanding of its learning dynamics. This is crucial for swiftly pinpointing and addressing issues, thereby streamlining and improving the AI model training and debugging process.
4. **Model Selection**: Visually compare how multiple models process each data point within a dataset, particularly focusing on the top dimensions. By providing a comprehensive visualization that displays the entire dataset in the context of model outputs, it allows for a deeper understanding of how different models behave with specific data characteristics. This innovative approach aids in selecting the most appropriate model based on a holistic view of model performance across the entire dataset, ensuring a more informed and effective model choice.
5. **Model Ensembling**: Expanding on the Model Selection feature, this illustrates the interactions and combined effects of various AI models on a dataset, highlighting how different models process each data point. By visually representing the diversity and correlation among models, it enables users to explore and understand the potential synergies in model ensembling. This approach aids in constructing robust ensemble models, as users can make informed decisions based on a clear understanding of how different models complement each other in handling the entire dataset.

Large Language Model Solution

Immersion Analytics plays a key role in the development and governance of large language models (LLMs). Envision a data space where points represent prompts and LLM responses. Imagine visualizing numerous moderation scores including hate speech, profanity, self-harm and violence onto each data point using the Dimensional Engine, to reveal underlying patterns of model behavior, including potential biases, anomalies, and areas requiring guard railing. While ineffective for multiple moderation scores, PCA can be used to reduce vector embeddings of the text onto X, Y and Z axes for visually organizing the data.

This visual framework both enhances interpretability and enables you to swiftly observe and address areas in the LLM's outputs that diverge from desired ethical and safety standards. It can be applied during research and development as well as for monitoring batches of prompt / response pairs in production.

Use Cases spanning AI Modalities

1. **Hyperparameter Optimization**: One of the more costly aspects of building and fine-tuning models is the selection and tuning of hyperparameters. The hyperparameter space is typically vast and multidimensional, so searching this space is both computationally expensive and time-consuming. While various techniques such as grid search and Bayesian optimization are currently used, they often require a significant computational budget and do not always find the optimal solution. This can potentially become an $O(n^m)$ problem (in the brute force case), where m is the number of hyperparameters, and n is the number of data points. By instead randomizing a tractable set of hyperparameter combinations then visualizing model error for each, domain experts engage with AI developers thoughtfully on pragmatic ways to best narrow the search space, reducing training costs by potentially an order of magnitude or more.
2. **Real-Time Performance Monitoring**: This leverages human expertise for live anomaly detection in production AI systems. It renders a multidimensional visualization of data and model output in batches, enabling users to intuitively understand and monitor the AI's behavior. This approach emphasizes the use of human insight to complement existing methods, allowing for more nuanced and context-aware detection of anomalies and deviations in AI behavior.
3. **Education**: Instructors leverage the following capabilities to employ interactive visualizations to enhance the teaching of AI and machine learning concepts to a diverse group of students. These visual tools demystify complex topics like neural networks and decision trees, making them accessible to students with varying backgrounds. During lectures, these visualizations dynamically illustrate AI processes, such as how weights and biases evolve during neural network training. Students engage with these tools in hands-on activities and assignments, manipulating parameters to see real-time effects on model behavior. This interactive approach not only deepens their intuition for AI principles but also increases engagement. This use of visualization bridges the gap between theoretical learning and practical application, resulting in a more effective and engaging learning experience, and preparing students for more advanced studies or careers in AI and data science.

Architecture

Interoperability across teams and tools is especially critical. AI developers and data scientists appreciate flexibility to choose the libraries, hardware, and tools that best fit their needs. Our Immersive Data Visualization Engine sits on a foundation of our Runtime Platform – including a library of API’s and software integrations. These give each team freedom to visualize data from any model, algorithm, dataset, and system and then see it or collaborate via a multitude of deployment options including Browser-based, PC, Mac, mobile phones, tablets, AR/VR headsets and even the world’s first glasses-free 3d laptop.

Next Steps

Reach out to contact@immersionanalytics.com to see how our pioneering visualization technologies upgrade AI workflows, and join us in shaping a safer, more inclusive future for AI.