Dimension: From curse to effectivity

Alexandre DEROUX*

July 18, 2025

*École polytechnique, ESILV Paris and De Vinci Research Center

Contents

Part I
Introductory part

1 Context

1.1 Socio-technical and economic context

Since the 1990s, financial data has been the subject of extensive study, and financial institutions have adopted quantitative models for portfolio and risk management. Tools such as logistic regression and neural networks have become essential elements in the banking sector, introducing a level of accuracy to models. However, this increasing use of financial prediction models also comes with risks.

One major risk is related to the accuracy of the models. While these models can provide accurate predictions in many cases, there is always a possibility of generalisation error. This means that the models may not perform optimally when applied to new datasets or situations different from those on which they were initially trained. This generalisation error can lead to incorrect and potentially costly predictions for financial institutions.

Furthermore, financial regulation has become stricter, necessitating explainability of the algorithms used in financial prediction models. Financial institutions need to be able to understand how predictions are generated and which variables most influence these predictions. This is particularly important in sensitive areas such as credit granting or fraud detection, where significant decisions are made based on the model's outcomes.

Overall, the current socio-technical and economic context highlights the importance of understanding complex, large-scale datasets used in financial prediction models. It is crucial to evaluate the generalisation error of these datasets using different prediction algorithms and compare their performance. This will help characterize the risk associated with prediction models for different types of financial data, highlighting the factors that influence the outcomes.

1.2 Scientific context

The research project focuses on studying the risks associated with the use of financial prediction models by analysing commonly used datasets. These datasets typically include data related to credit risk, churn risk, fraud, and market data.

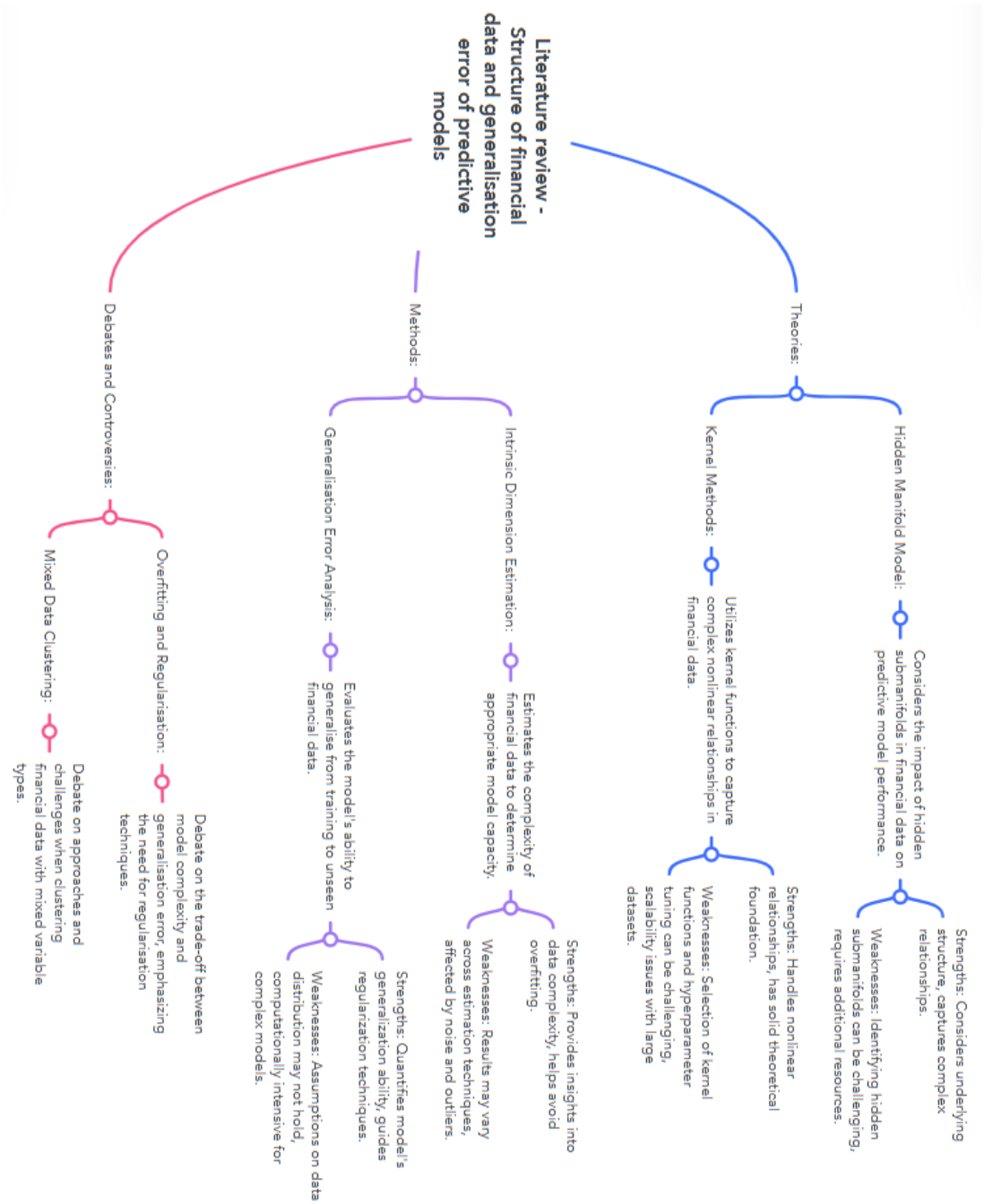
The primary objective of the research is to evaluate the generalisation error of these datasets using various prediction algorithms and compare their perfor-

mance. The goal is to determine how the models perform when applied to new datasets or situations different from those on which they were initially trained. This evaluation will provide a better understanding of the limitations of financial prediction models and identify potential vulnerabilities that could lead to erroneous predictions.

The ultimate objective of the research is to characterise the risk associated with prediction models for different types of financial data. This involves identifying the factors that influence model performance, such as data quality, dataset size, variables used in the models, etc. By identifying these factors, financial institutions will be able to make more informed decisions regarding the use of prediction models and mitigate potential risks.

In summary, the scientific context of the research project aims to evaluate the generalisation error of financial prediction models, compare their performance, and characterise the risk associated with these models for different types of financial data.

2 Literature review



2.1 Theories

- **Hidden Manifold Model**: This theory explores how the underlying hidden structure of financial data can impact the performance of predictive models. It suggests that considering the presence of hidden submanifolds in the data can improve the model's generalisation capabilities.
  Strengths: It considers the underlying structure of financial data, potentially improving model performance. It can capture nonlinear relationships and provide insights into latent factors affecting financial variables.
  Weaknesses: Identifying and characterising hidden submanifolds can be challenging. It may require additional computational resources and domain expertise.

- **Kernel Methods**: Kernel methods, such as Support Vector Machines, utilize kernel functions to transform data into high-dimensional feature spaces. These methods have been widely used in finance for capturing complex nonlinear relationships.
  Strengths: They can handle nonlinear relationships and high-dimensional feature spaces. They have a solid theoretical foundation and can capture complex patterns in financial data.
  Weaknesses: Selecting appropriate kernel functions and tuning hyperparameters can be nontrivial. Kernel methods may suffer from scalability issues with large datasets.

2.2 Methods

- **Intrinsic Dimension Estimation**: Estimating the intrinsic dimension of financial data can provide insights into its complexity and help determine the appropriate model capacity. Various techniques, including minimal neighbourhood information and dimensionality estimation algorithms, can be employed.
  Strengths: Estimating the intrinsic dimension provides insights into the complexity of financial data. It aids in determining suitable model capacities and avoiding overfitting.
  Weaknesses: Different estimation techniques may yield varying results. The estimation accuracy can be affected by noise and outliers present in the data.

- **Generalisation Error Analysis**: Evaluating the generalisation error of predictive models is crucial in assessing their performance. Techniques like Rademacher complexity and Gaussian equivalence can provide insights into the model's ability to generalise from the training data to unseen financial data.
  Strengths: It quantifies the model's ability to generalize to unseen financial data. It provides a theoretical understanding of model performance and can guide the selection of appropriate regularisation techniques.
  Weaknesses: Generalisation error analysis often assumes certain data distribution assumptions, which may not hold in real-world financial data. The analysis can be computationally intensive for complex models.

2.3 Debates and controversies

- **Overfitting and Regularisation**: There is an ongoing debate regarding the trade-off between model complexity and generalisation error. Some argue that complex models can capture intricate patterns in financial data, while others emphasise the risk of overfitting and advocate for regularisation techniques.

- **Mixed Data Clustering**: The clustering of financial data that contains mixed types of variables (e.g., numerical and categorical) is a topic of debate. Different approaches and challenges exist when clustering such heterogeneous data.

2.4 Research contribution and integration

In our research on the structure of financial data and generalization error of predictive models, we will build upon existing work and contribute in the following ways:

- **Integration of Hidden Manifold Model**: We will explore the impact of hidden submanifolds in financial data on the generalisation error of predictive models. By incorporating the hidden manifold model into existing frameworks, we aim to enhance model performance and uncover latent factors influencing financial variables.

- **Novel Intrinsic Dimension Estimation Techniques**: We will develop and evaluate new techniques for estimating the intrinsic dimension of financial data. These techniques will account for the specific characteristics and complexities present in financial datasets, improving accuracy and robustness.
- **Comparative Analysis of Model Regularisation**: I will conduct a comprehensive comparative analysis of regularisation techniques in the context of financial data. This analysis will assess the strengths and weaknesses of different regularisation approaches and provide insights into their effectiveness in reducing generalisation error.

By building on existing theories, methods, and debates, our research aims to contribute to the understanding of the structure of financial data and its implications for predictive modeling. It seeks to address the challenges and controversies in this field and provide practical recommendations for improving the generalisation performance of predictive models in financial applications.

Part II
Experimental part

1 Methodology

In this study, we investigate the relationship between the effective dimension of the data and the generalisation error of three datasets. We also explore the choice of predictive models to minimise the generalisation error for each dataset.

1.1 Algorithms for calculating effective dimension

The effective or intrinsic dimensionality of a data set refers to the quantity of significant variables that capture most of the information in the data. For example, although a data set may have hundreds of variables, if a few of these variables capture most of the variance in the data, then the effective dimensionality could be considerably smaller than the total number of variables.

We consider the following three algorithms for calculating the effective dimension of the data:

- Method with variance ratio.
- Methods with various heuristics on eigenvalues.
- Methods with nearest neighbours.

Each algorithm provides a measure of the effective dimension, which captures the intrinsic complexity of the data.

1.2 Dataset structure

We analyse three datasets with different characteristics:

- German Credit Risk.
- S&P 500.
- Customer Churn.

We examine the structure of each dataset through exploratory data analysis, including visualisations and statistical summaries. This analysis helps us understand the underlying patterns and relationships in the data.

1.3 Linking effective dimension to generalisation error

Next, we investigate the relationship between the effective dimension and the generalisation error for each dataset. We conduct experiments using different predictive models, including linear regression, decision trees, and neural networks. We train and evaluate these models on each dataset, varying the generalisation error metric using the effective dimension.

1.4 Minimising generalisation error

Based on the results of our experiments, we identify the predictive models that minimize the generalisation error for each dataset. We assess the performance of these models using appropriate evaluation metrics such as mean squared error or accuracy.

2 Results and discussions

2.1 Dimension calculation

You can find all our codes here:

`https://github.com/Alexandre-Deroux/Dimension-Calculation`

We first selected ten datasets.
Five have well-known effective dimensions based on a consensus of various scientific articles[2][20]:

- MNIST Test.
- MNIST Train Small.
- CIFAR-10 Test.
- CIFAR-10 Train.
- CelebA.

And five in the field of finance, our area of application:

- German Credit Risk in 9 dimensions.
- German Credit Risk in 21 dimensions.
- Default of Credit Card Clients Dataset.
- S&P 500.
- Customer Churn.

After extensive research, we have implemented seven approaches to calculate the effective dimension:

1. **Method with variance ratio**: A method that calculates the number of eigenvalues of the covariance matrix that describe a certain proportion of the total variance.

2. **Method with n1**: A method based on the following eigenvalue heuristic[6]:

$$\prod_{j=1}^{K} \left(\frac{\lambda_j}{\sum_{i=1}^{K} \lambda_i} \right)^{-\frac{\lambda_j}{\sum_{i=1}^{K} \lambda_i}}$$

3. **Method with n2**: A method based on the following eigenvalue heuristic[6]:

$$\frac{\left(\sum_{i=1}^{K} \lambda_i \right)^2}{\sum_{i=1}^{K} \lambda_i^2}$$

4. **Method with infinite n**: A method based on the following eigenvalue heuristic[6]:

$$\frac{\sum_{i=1}^{K} \lambda_i}{max_i \lambda_i}$$

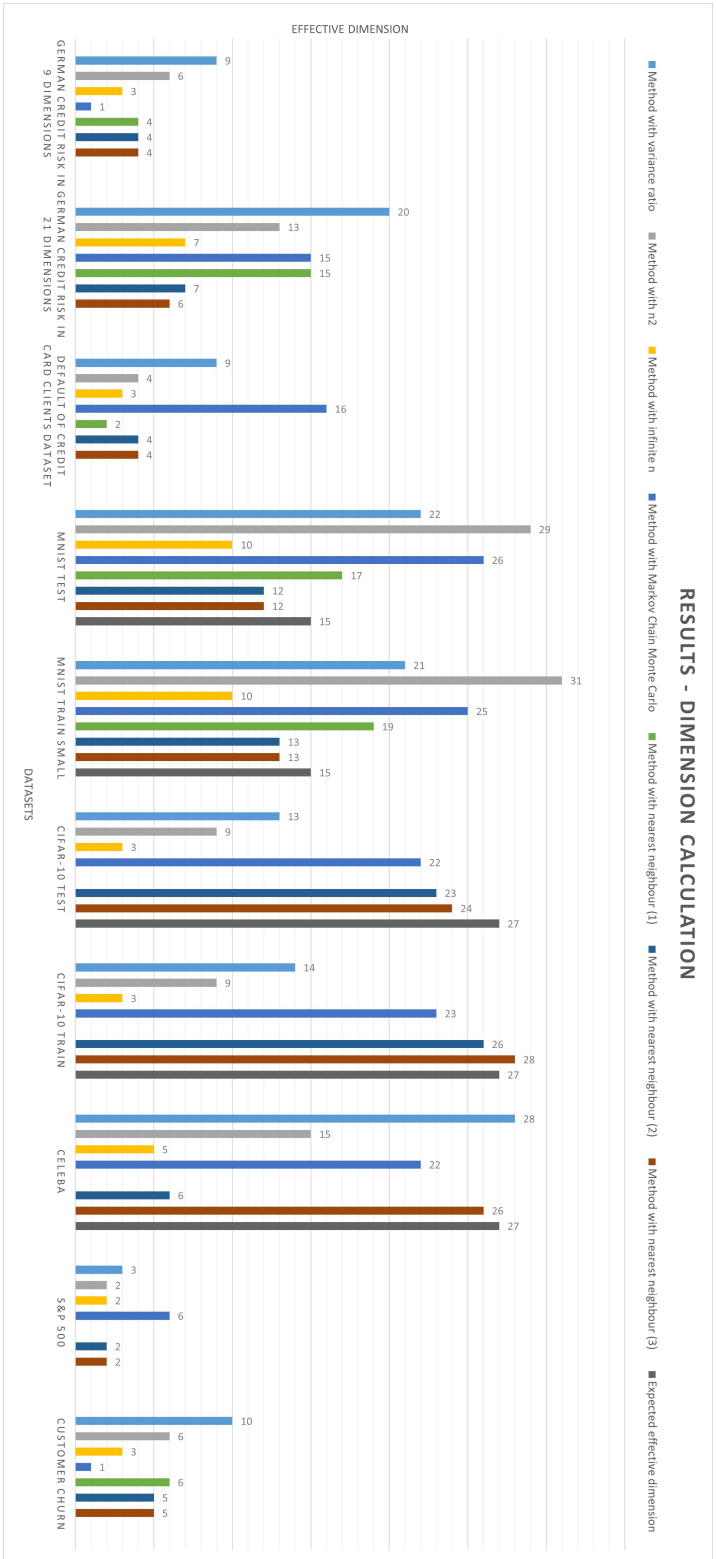
5. **Method with Markov Chain Monte Carlo**: A method that uses Markov chains and the Metropolis-Hastings criterion[16].

6. **Method with nearest neighbour (1)**: A method that uses the minimum slope of a certain line defined by a nearest neighbour heuristic[12].

7. **Method with nearest neighbour (2)**: Method based on a heuristic playing on the 30 nearest neighbours[10][20]:

$$\hat{m}_k(x) = \left[\frac{1}{k-1} \sum_{j=1}^{k-1} log \frac{T_k(x)}{T_j(x)} \right]^{-1}$$

   where $T_j(x)$ is the Euclidean (L_2) distance from x to its j^{th} nearest neighbour.


RESULTS - DIMENSION CALCULATION
EFFECTIVE DIMENSION
DATASETS
GERMAN CREDIT RISK IN 9 DIMENSIONS
9 6 3 1 4 4 4
GERMAN CREDIT RISK IN 21 DIMENSIONS
20 13 7 15 15 7 6
DEFAULT OF CREDIT CARD CLIENTS DATASET
9 4 3 16 2 4 4
MNIST TEST
22 29 10 26 17 12 12 15
MNIST TRAIN SMALL
21 31 10 25 19 13 13 15
CIFAR-10 TEST
13 9 3 22 23 24 27
CIFAR-10 TRAIN
14 9 3 23 26 28 27
CELEBA
28 15 5 22 6 26 27
S&P 500
3 2 2 6 2 2
CUSTOMER CHURN
10 6 3 1 6 5 5
Method with variance ratio
Method with n2
Method with infinite n
Method with Markov Chain Monte Carlo
Method with nearest neighbour (1)
Method with nearest neighbour (2)
Method with nearest neighbour (3)
Expected effective dimension

Of the seven methods we found, we observed that the **method with variance ratio**, which captures a significant amount of variance in the data, gave very good results for binary datasets such as CelebA. Outside this case, the **method with nearest neighbour (2)**, a method based on a heuristic that takes into account the 30 nearest neighbours gave satisfactory results.

Based on these different results, we designed an algorithm: **Method with nearest neighbour (3)**, which calculates the effective dimension by combining the estimates for binary and non-binary columns and weighting them by the respective number of columns. This algorithm gives very satisfactory results on the five datasets for which the dimensions are known.

We then evaluated these eight methods by taking a multivariate normal distribution and different independent distributions, with several dimensions between 5 and 150 and several correlation values between 0 and 1.

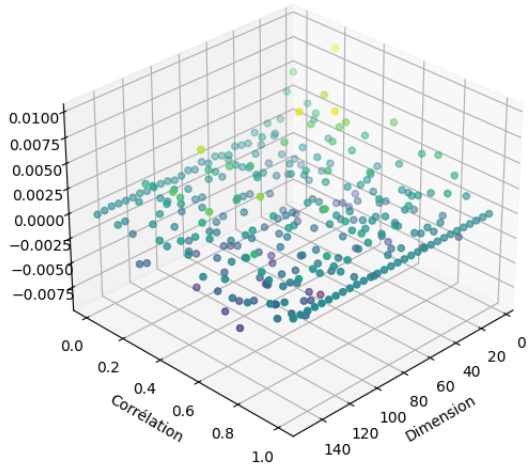
The effective dimension for Gaussian variables can be thought of as the number of independent directions that contain most of the information or variance in the data.

In theory, when the correlation is 1, the effective dimension should be equal to 1, and when the correlation is 0, the effective dimension should be equal to the dimension.

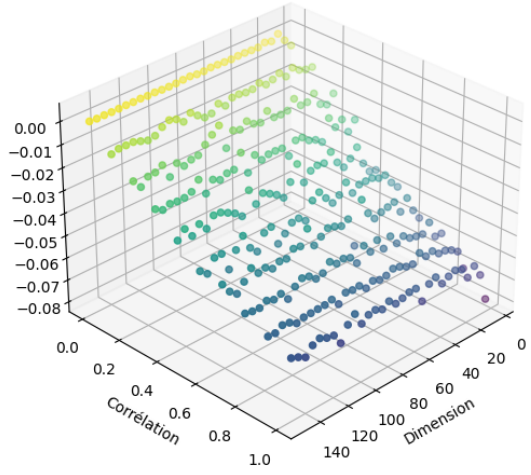
In practice, we are faced with two problems:

1. A correlation of 1 is impossible to obtain as soon as you deviate from the normal distribution, as shown in the second graph below, where the correlation error between the stipulated correlation and reality increases as the correlation increases:

For an evaluation on a multivariate normal distribution:



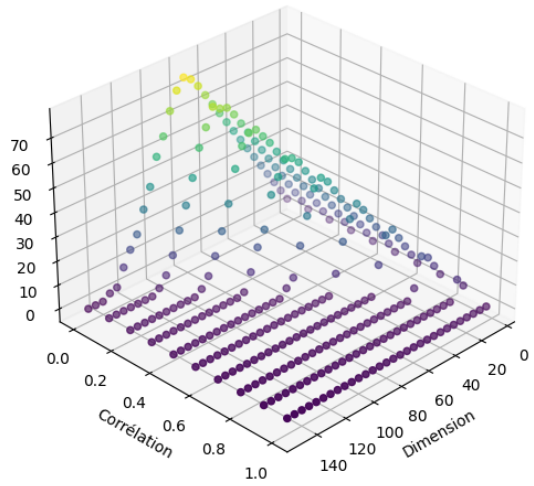
For an evaluation on different independent distributions:



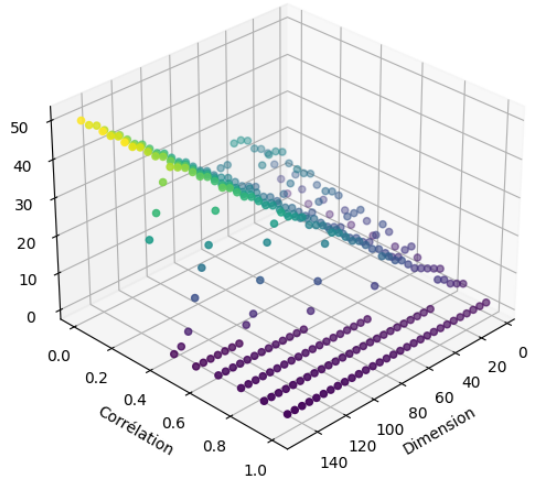
2. It also seems difficult in practice to completely decorrelate variables when the dimension is large.

Firstly, the **method with variance ratio** with a selected proportion of 99% gives inconsistent results for large dimensions, and even anarchic results for different independent distributions:

For an evaluation on a multivariate normal distribution:

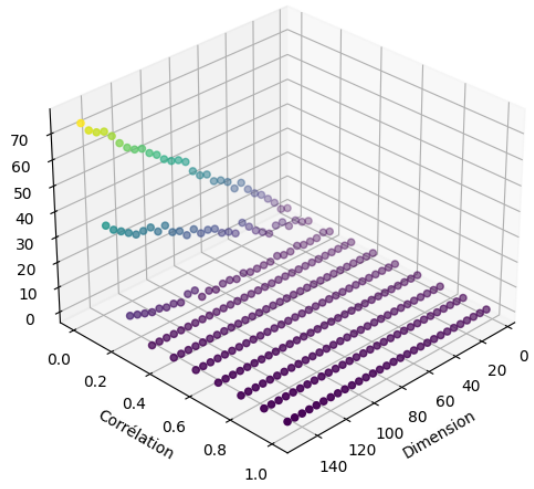


For an evaluation on different independent distributions:



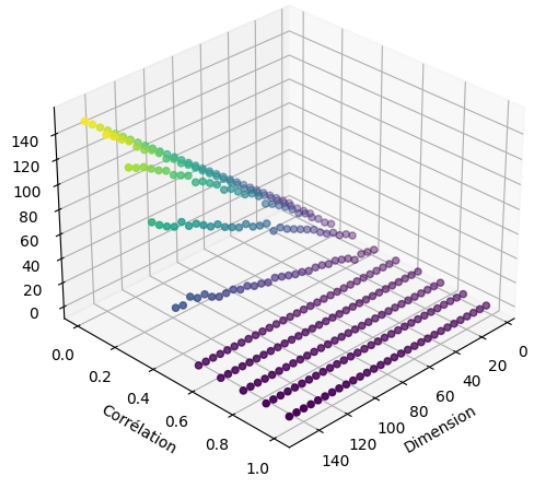
However, by changing the selected proportion to: $1/n$, we obtain more coherent results:

For an evaluation on a multivariate normal distribution:



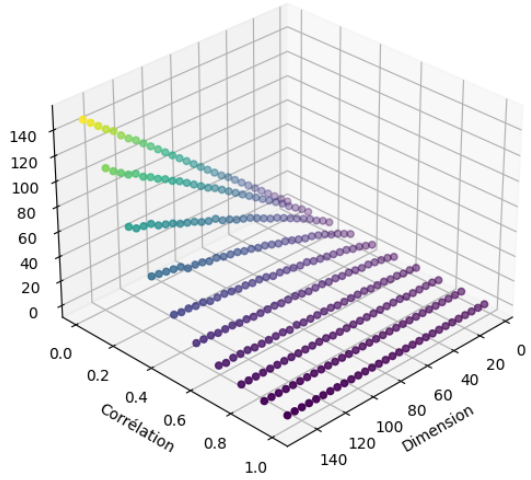
And with a selected proportion of: $1/(\varphi \cdot n)$, where φ is the golden ratio, we get good results:

For an evaluation on a multivariate normal distribution:

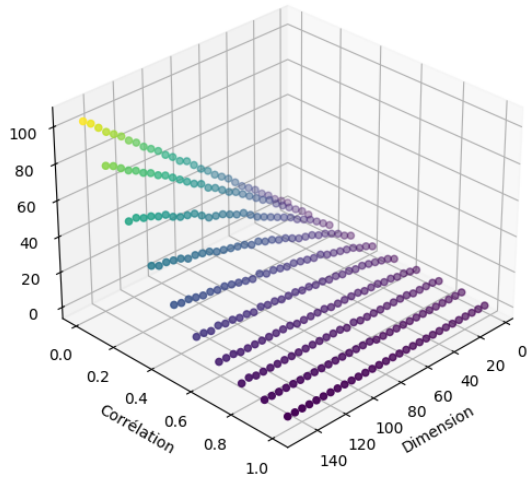


The **method with n1** gives perfectly consistent results for both distributions, but in practice it gives incongruous results such as an effective dimension of 69 instead of 15 for MNIST Train Small. We can nevertheless conclude that in this case, this method seems to be the most appropriate:

For an evaluation on a multivariate normal distribution:

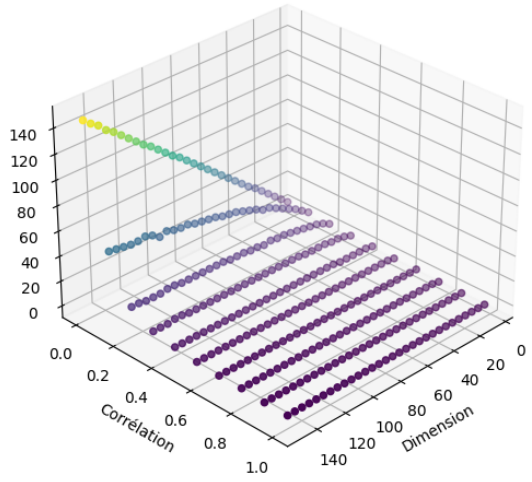


For an evaluation on different independent distributions:

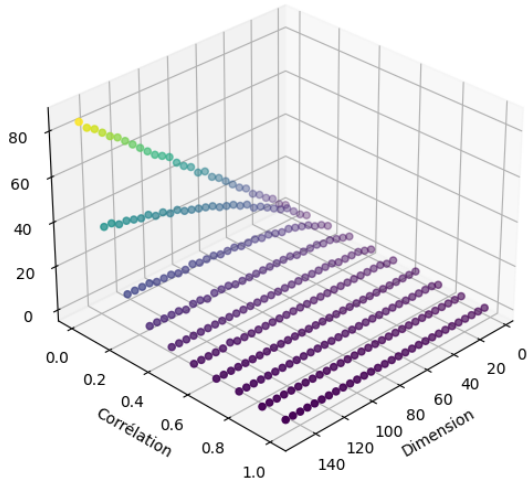


The **method with n2** gives more conservative results than method with n1:

For an evaluation on a multivariate normal distribution:

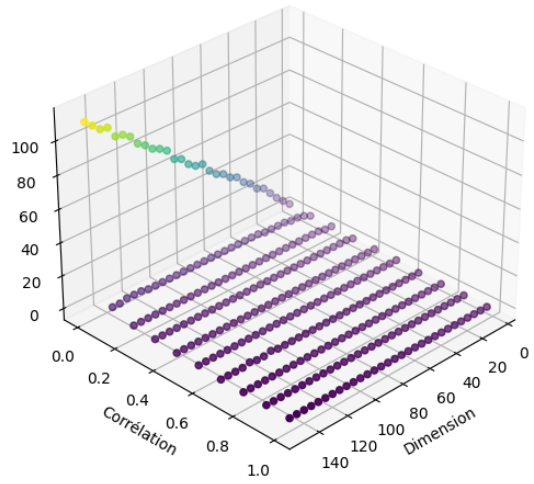


For an evaluation on different independent distributions:

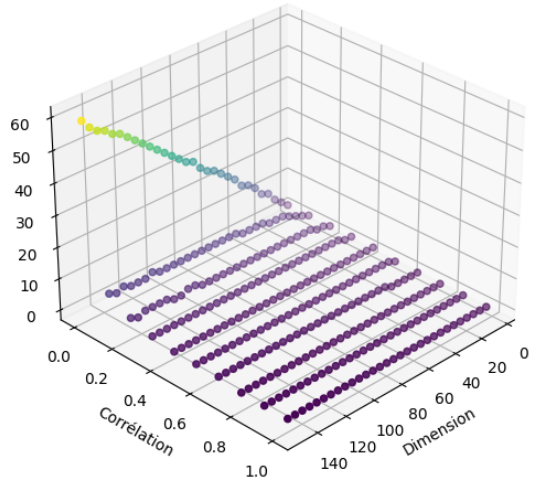


The **method with infinite n** gives even more conservative results than method with n2:

For an evaluation on a multivariate normal distribution:

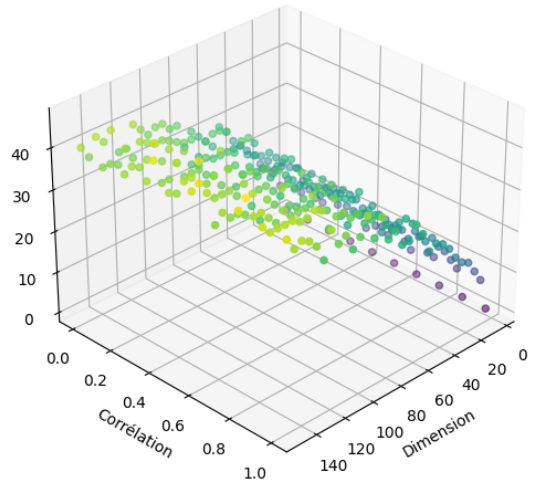


For an evaluation on different independent distributions:

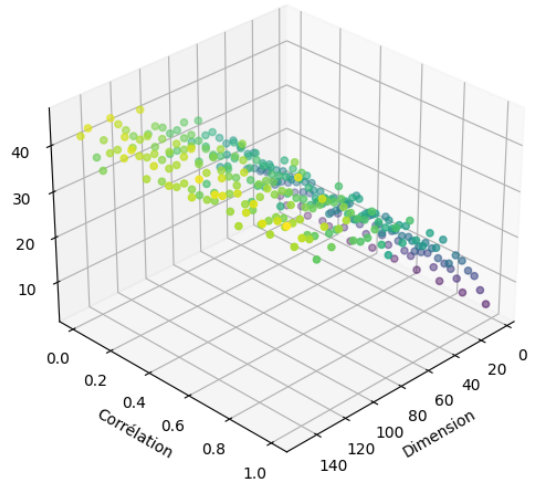


The **method with Markov Chain Monte Carlo** gives totally inconsistent results since the effective dimension does not seem to depend on the correlation:

For an evaluation on a multivariate normal distribution:



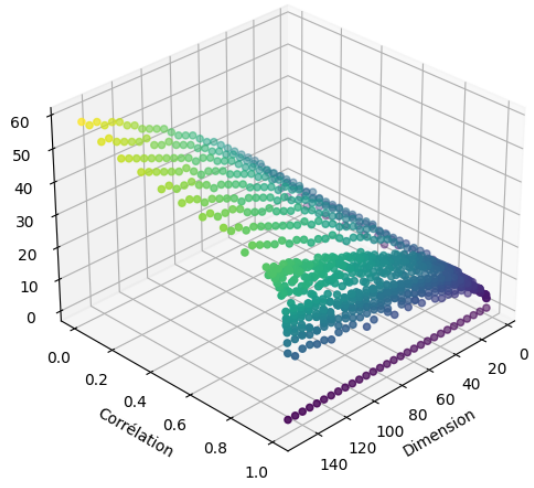
For an evaluation on different independent distributions:



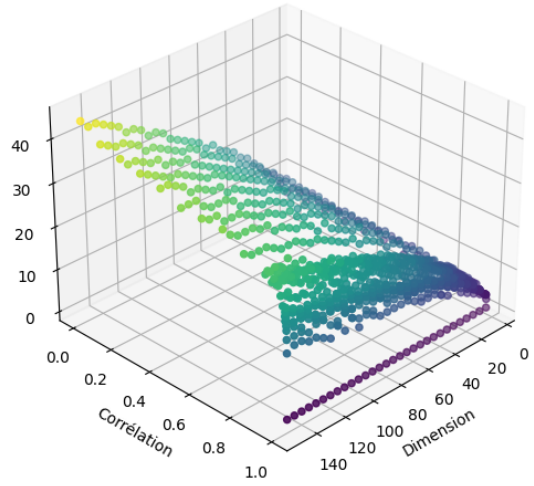
The **method with nearest neighbour (1)** takes too long to calculate, which makes these assessments impossible. Nevertheless, the results obtained in practice are not suitable.

The **method with nearest neighbour (2)** gives perfectly acceptable results, despite a virtual break in the curve when the correlation approaches 1:

For an evaluation on a multivariate normal distribution:

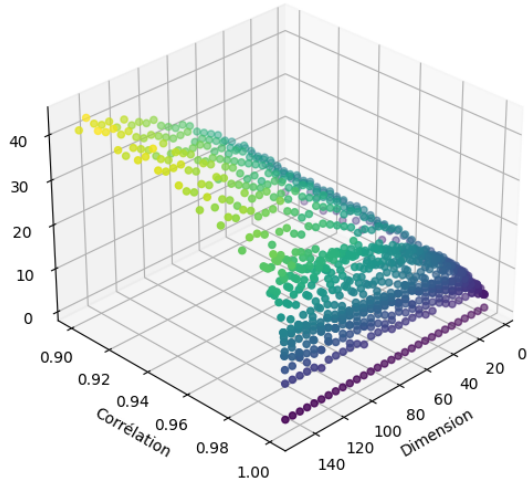


For an evaluation on different independent distributions:



If we take a closer look at this zone when the correlation approaches 1, we can see that this is not a break in the curve:

For an evaluation on a multivariate normal distribution:

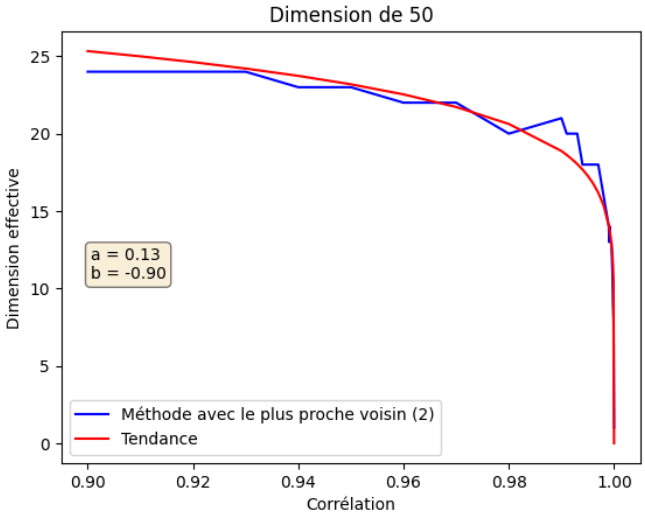


This curve can be characterised as follows:

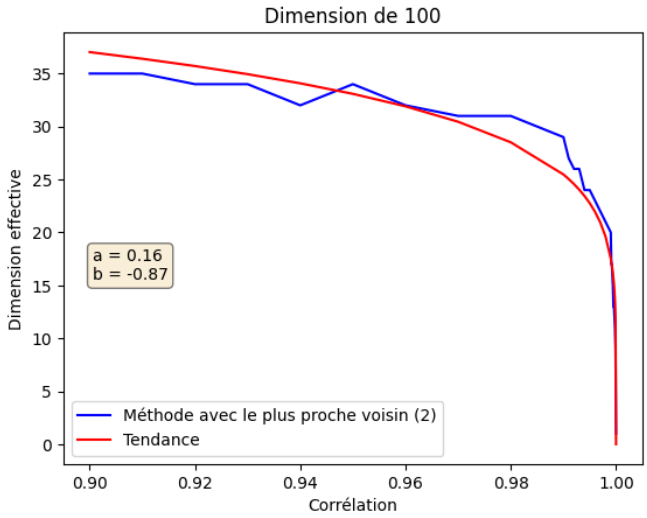
$$\frac{(1-x)^a}{dim^b}$$

Using finite-size scaling theory, we can find the coefficients a and b that best describe the trend of the curve:

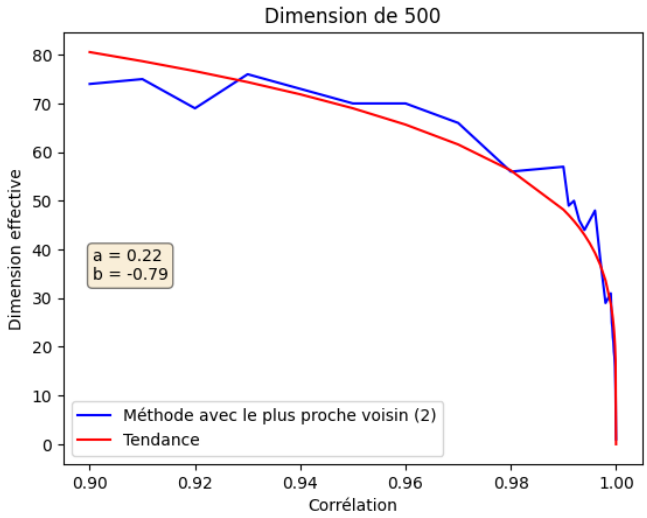
For an evaluation on a multivariate normal distribution:



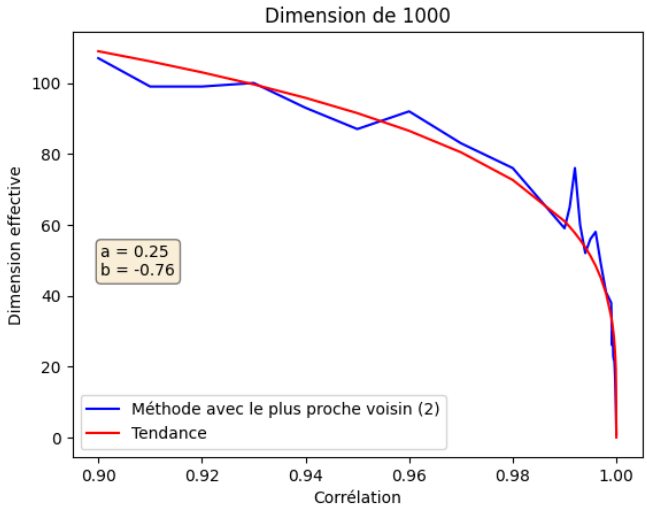
For an evaluation on a multivariate normal distribution:



For an evaluation on a multivariate normal distribution:

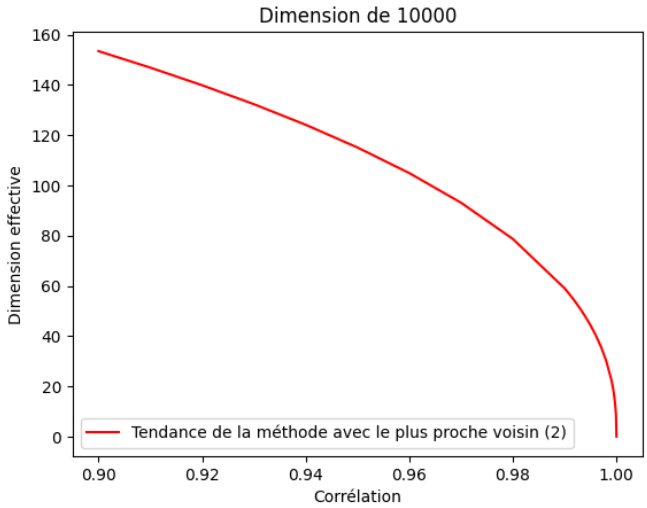


For an evaluation on a multivariate normal distribution:

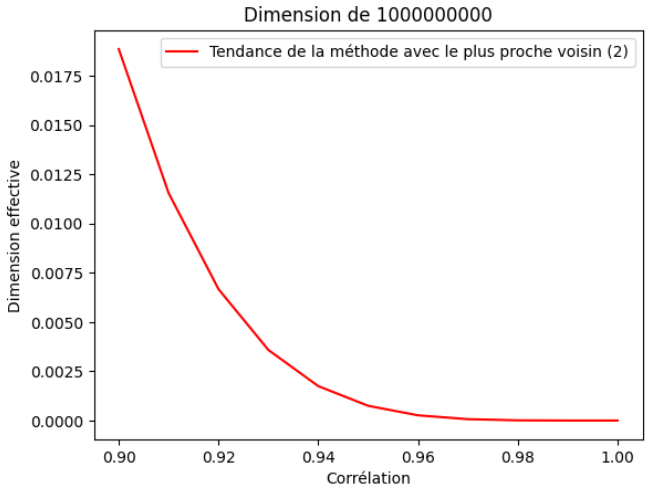


So we can extrapolate the trend of the curve for large dimensions:

For an evaluation on a multivariate normal distribution:



For an evaluation on a multivariate normal distribution:



We arrive at a surprising result, since the effective dimension seems to tend towards 0.

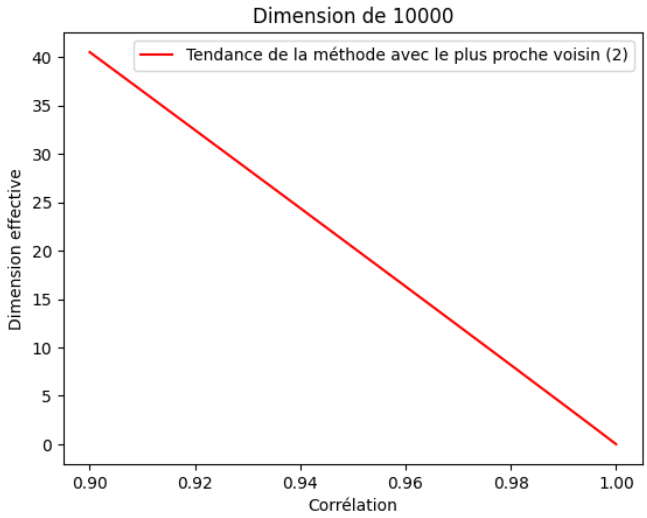
As a reminder, the k nearest neighbours heuristic is as follows[10][20]:

$$\hat{m}_k(x) = \left[\frac{1}{k-1} \sum_{j=1}^{k-1} log \frac{T_k(x)}{T_j(x)} \right]^{-1}$$

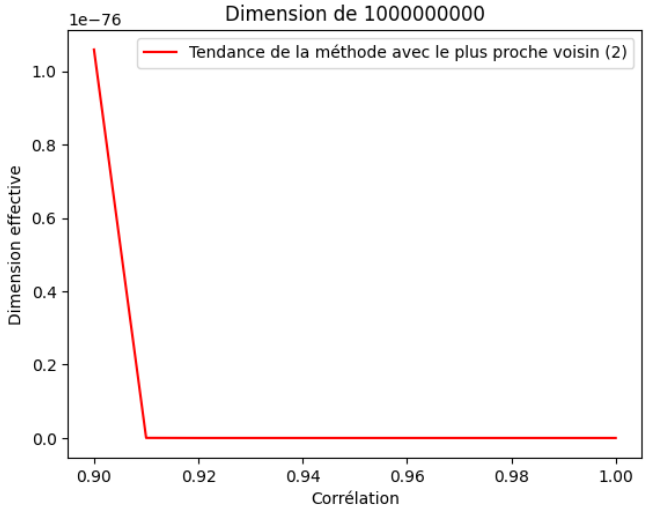
where $T_j(x)$ is the Euclidean (L_2) distance from x to its j^{th} nearest neighbour.

By increasing the number of nearest neighbours, we obtain results that tend even more rapidly towards 0. Here are the trends of the curve for large dimensions when $k = 50$):

For an evaluation on a multivariate normal distribution:

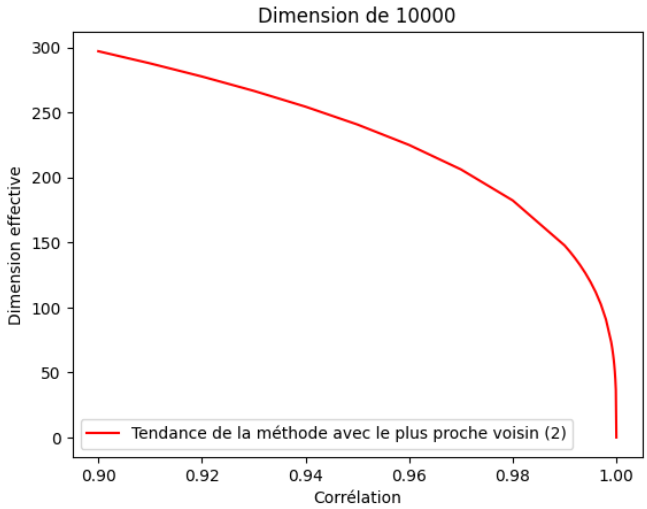


For an evaluation on a multivariate normal distribution:

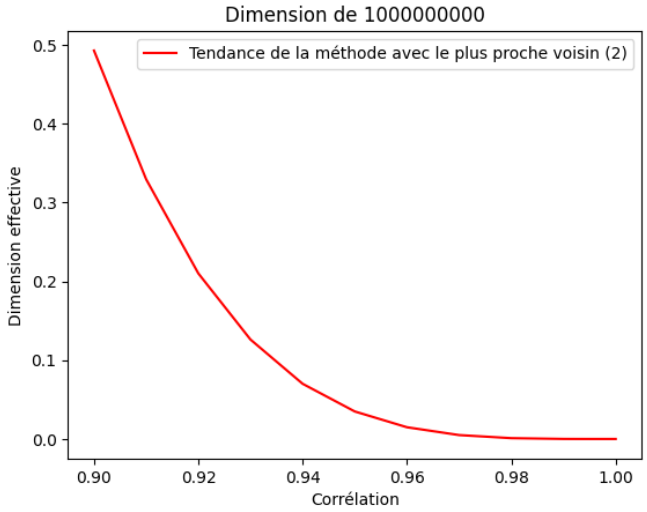


Decreasing the number of nearest neighbours gives results that tend less quickly towards 0. Here are the trends of the curve for large dimensions when $k = 10$:

For an evaluation on a multivariate normal distribution:



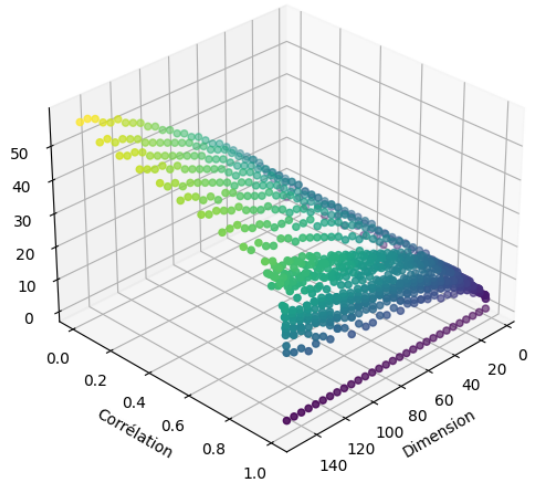
For an evaluation on a multivariate normal distribution:



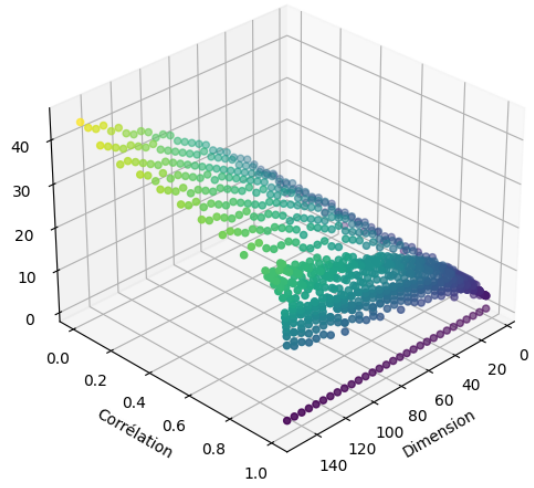
However, taking $k = 10$, the results obtained in practice are not suitable.

The **method with nearest neighbour (3)** gives identical results to method with nearest neighbour (2), since there are no binary values:

For an evaluation on a multivariate normal distribution:



For an evaluation on different independent distributions:



We are currently working on the theoretical approach of the method with nearest neighbour (2) and the geometric complexity of datasets[1][7][10].

2.2 Generalisation error

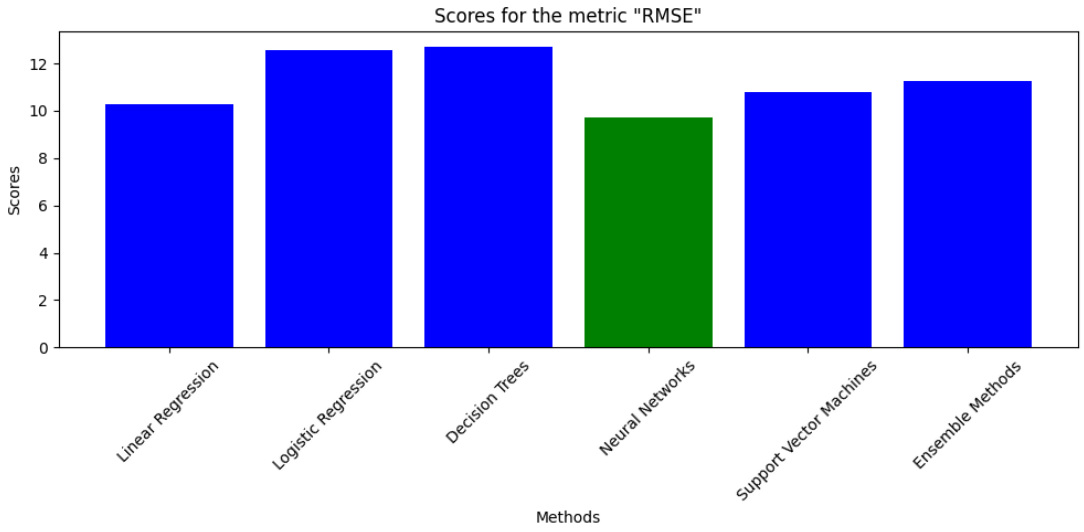
You can find all our codes here:

`https://github.com/Alexandre-Deroux/Generalisation-Error`

We have also worked on generalisation error. First, we selected six prediction methods:

1. Linear Regression.
2. Logistic Regression.
3. Decision Trees.
4. Neural Networks.
5. Support Vector Machines.
6. Ensemble Methods.

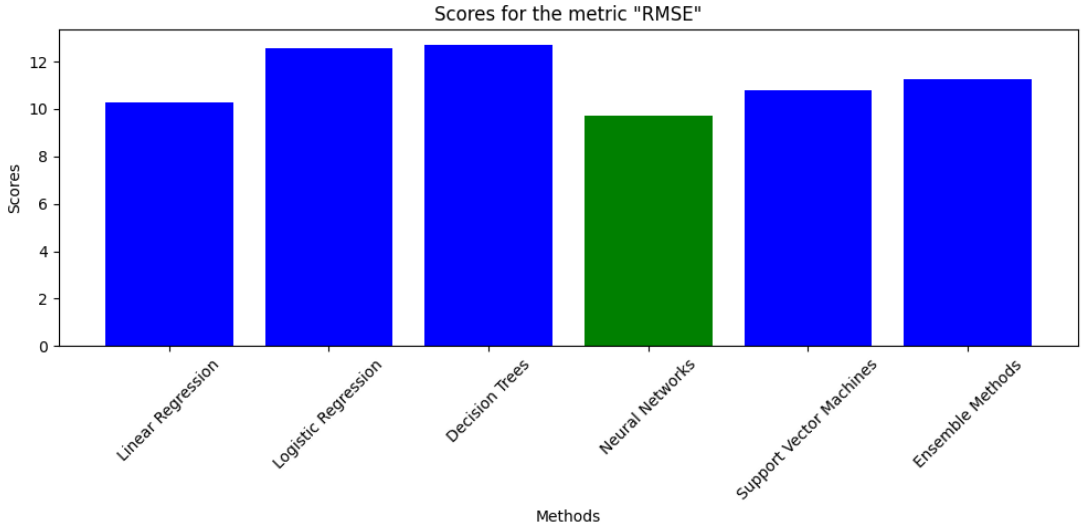
Then, we calculated different metrics for this error, such as RMSE, for the generalisation of age in the German Credit Risk dataset with 21 dimensions:



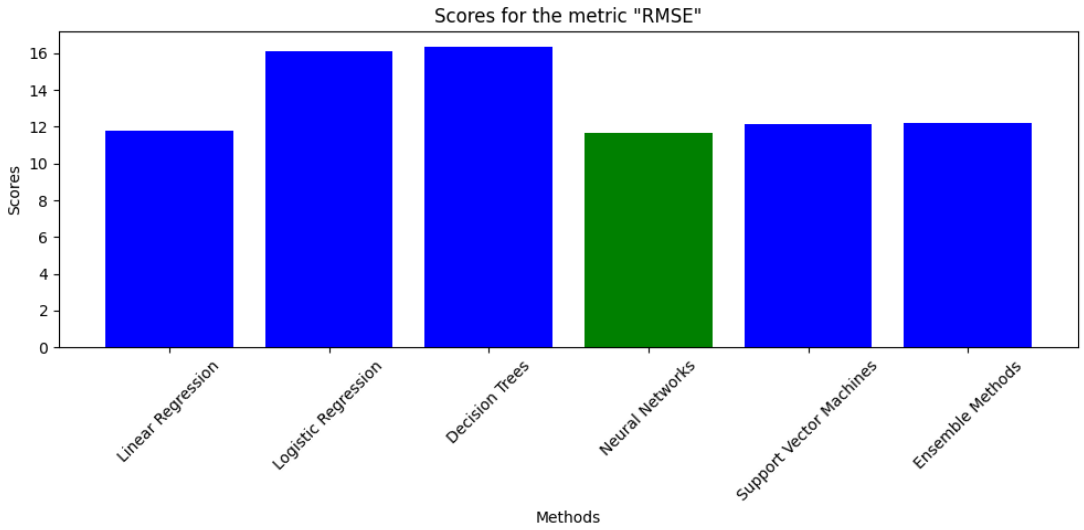
This allowed us to identify a preferred method for minimising this error, which in this case is the neural network method.

Next, we implemented these metrics by training on the dataset reduced to the effective dimension, which in this case is six. We found that there was almost no variation in the preferred methods for each measure. This constitutes an initial confirmation bias for our effective dimension:

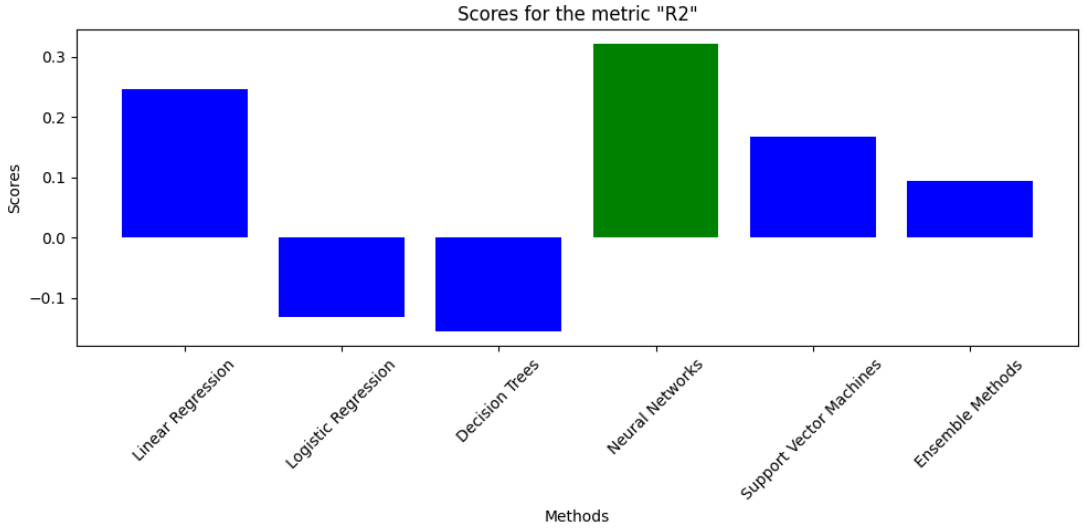
Without effective dimension:



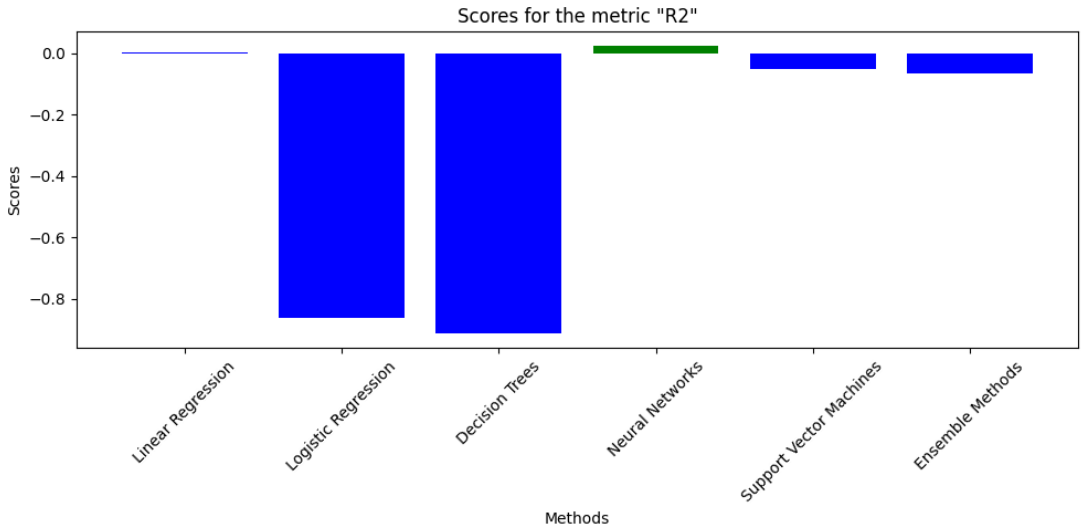
With PCA on effective dimension:



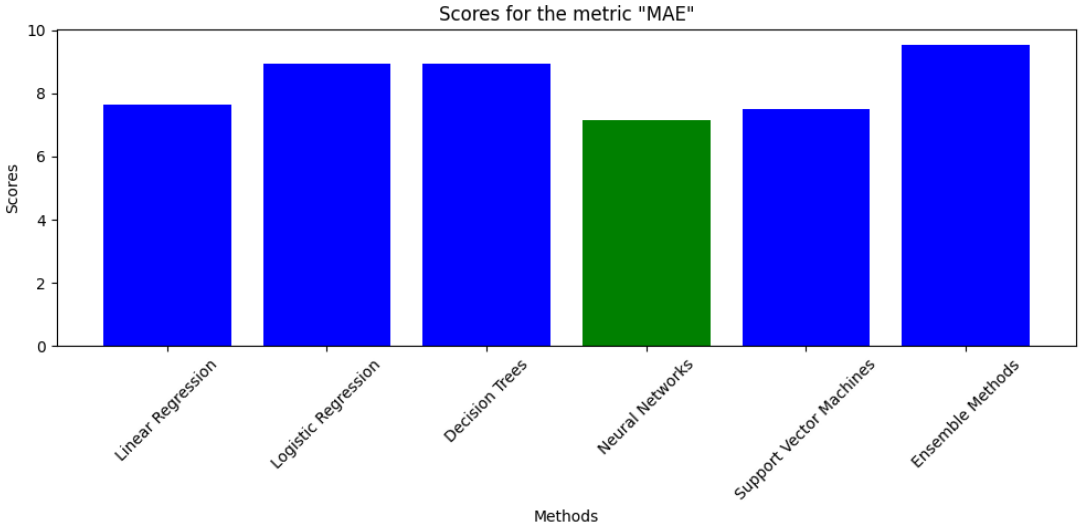
Without effective dimension:



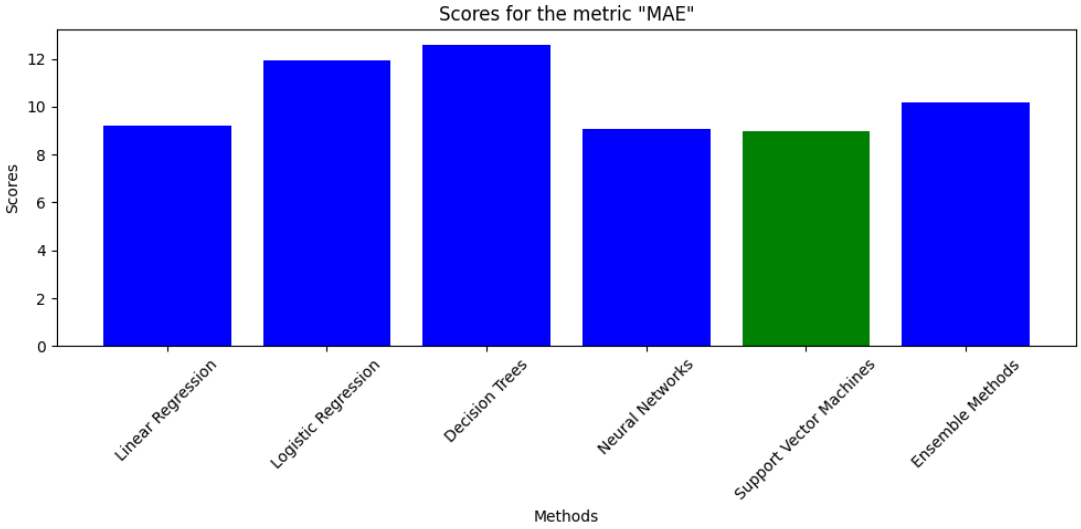
With PCA on effective dimension:



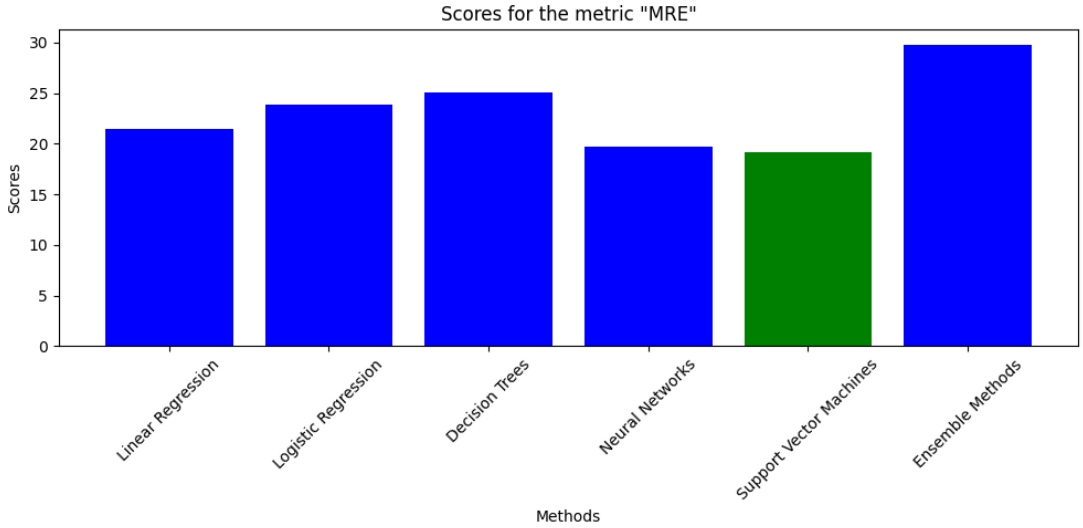
Without effective dimension:



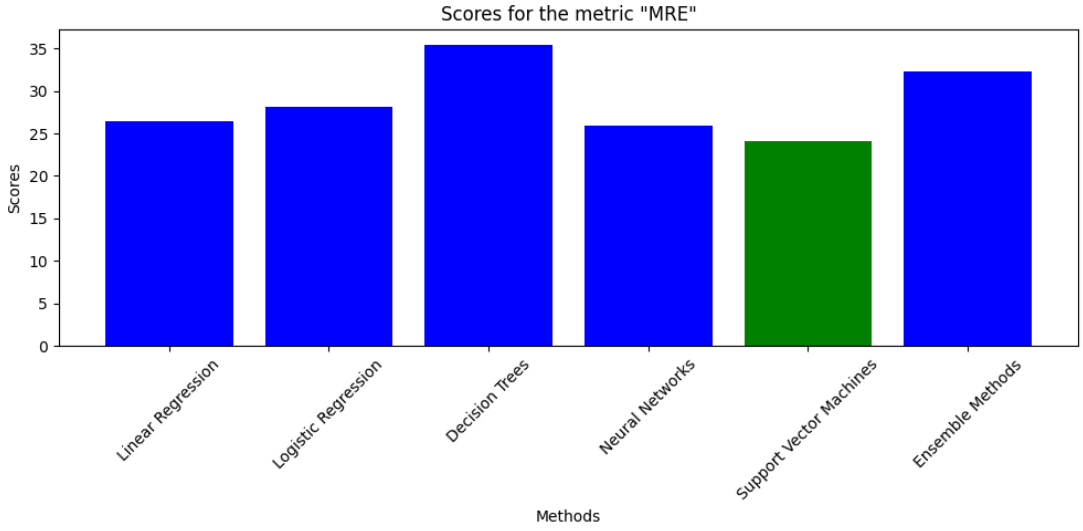
With PCA on effective dimension:



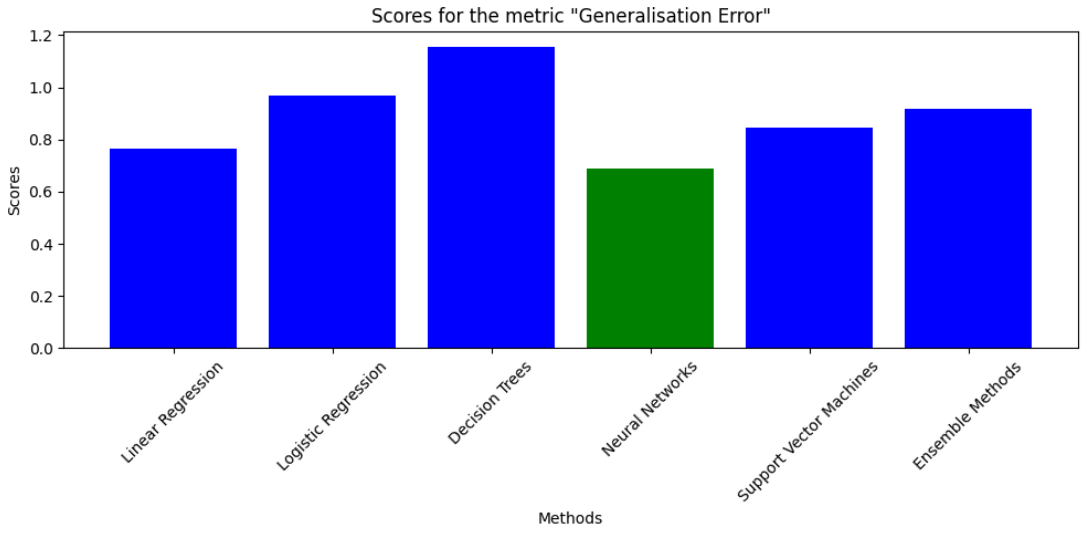
Without effective dimension:



With PCA on effective dimension:



We are currently working on a metric for generalisation error that incorporates the effective dimension based on a heuristic[3][5][17]. The results are encouraging so far. In our case, the neural network method is consistently indicated as the preferred approach:



Part III
Theoretical part

1 Explanation of the method with nearest neighbour (2)

1.1 Introduction

In this paper, we explore various methodologies for estimating the intrinsic dimension of a manifold embedded within a Euclidean space. The concept of intrinsic dimension is critical in understanding the underlying complexity and structure of data sets in high-dimensional analysis. Our approach is heavily based on the pioneering work of Levina and Bickel, who introduced a novel methodology employing a Poisson process to model data sampled from a distribution with an unknown density. This approach not only facilitates a deeper understanding of data geometry but also enhances the efficiency of data processing techniques.

1.2 Model description

Consider a data manifold of unknown dimension m:

$$X = \{x = g(b) \in \mathbb{R}^p : b \in B \subset \mathbb{R}^m\},$$

where (B, g) is a single coordinate chart embedded into an ambient p-dimensional space $\mathbb{R}^p$, such that $m \leq p$. The mapping g is a one-to-one mapping from an open bounded set $B \subset \mathbb{R}^p$ to manifold $X = g(B)$, with a differentiable inverse map $g^{-1} : X \to B$. The manifold X is unknown, and a finite data set $D = \{X_1, \ldots, X_n\} \subset X \subset \mathbb{R}^p$ is sampled from a distribution with an unknown density $f(x)$. We note that the single coordinate chart is a technical simplification, and the results are correct at least for manifolds covered with finite atlases.

1.3 Methodology

Levina and Bickel [14] suggested considering the binomial process

$$N(t, x) = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in S_x(t)\}}, \quad 0 \leq t \leq R,$$

where $S_x(t)$ is a ball of radius t centered at x. They propose to approximate this process by a Poisson process $N_\lambda(t, x)$ with rate $\lambda_{m,\theta}(t)$ where $\theta = \log f(x)$. Suppressing the dependence on x, the log-likelihood of the observed process $N_\lambda(t, x)$ is

$$L_\lambda(m, \theta) = \int_0^R \log \lambda_{m,\theta}(t)\, dN(t) - \int_0^R \lambda_{m,\theta}(t)\, dt.$$

1.4 Maximum likelihood estimation

The key idea of MLE is to fix a point x and for an unknown smooth density f on X assume that $f(z) \approx \text{const}$ in a ball $z \in S_x(R) \subset \mathbb{R}^p$ of small radius R, while the intersection of X and $S_x(R)$ is approximated by an m-dimensional ball $S_x^m(R)$. Then, the observations are treated as a Poisson process in $S_x^m(R) \subset \mathbb{R}^m$. The rate of the Poisson process for the resulting approximation is

$$\hat{\lambda}_{m,\theta}(t) = f(x) V_m m t^{m-1},$$

where V_m is the volume of the unit sphere in $\mathbb{R}^m$.

1.5 Numerical implementation

Let $T_k(x)$ be the Euclidean distance from a fixed point x to its k-th nearest neighbor in the sample D. We state the following Proposition [14]. **Proposition 1.** The intrinsic dimension estimate for a manifold X at a point x obtained by maximizing the likelihood with a rate is equal to

$$\hat{m}_R(x) = \left(\frac{1}{N(R,x)} \sum_{j=1}^{N(R,x)} \log \frac{R}{T_j(x)} \right)^{-1}.$$

For numerical calculations, it might be more convenient to fix the number of neighbors k rather than the radius of the ball R. Then the MLE reads as

$$\hat{m}_k(x) = \left(\frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{T_k(x)}{T_j(x)} \right)^{-1}.$$

1.6 Conclusion

We have introduced a robust method for estimating the intrinsic dimension of a manifold utilising the theory of Poisson processes and likelihood maximisation techniques. This method not only provides a theoretical framework but also practical tools for analysing the geometric structure of data in high-dimensional spaces. The implications of this research extend to various fields including machine learning, data compression, and complex systems analysis.

References

[1] Granata, D., & Carnevale, V. (2016). Accurate Estimation of the Intrinsic Dimension Using Graph Distances: Unraveling the Geometric Complexity of Datasets. DOI:10.1038/srep31377.

[2] Spiglera, S., Geigera, M., & Wyarta, M. (2019). Asymptotic learning curves of kernel methods: empirical data v.s. Teacher-Student paradigm. arXiv:1905.10843v8.

[3] Bartlett, P. L., Long, P. M., Lugosi, G., & Tsigler, A. (2019). Benign Overfitting in Linear Regression. arXiv:1906.11300v3.

[4] Del Ferraro, G., Wang, C., Marti, D., & Mézard, M. (2014). Cavity Method: Message Passing from a Physics Perspective. arXiv:1409.3048v1.

[5] Abbas, A., Sutter, D., Figalli, A., & Woerner, S. (2021). Effective dimension of machine learning models. arXiv:2112.04807.

[6] Del Giudice, M. (2020). Effective Dimensionality: A Tutorial. DOI:10.1080/00273171.2020.1743631.

[7] Facco, E., d'Errico, M., Rodriguez, A., & Laio, A. (2017). Estimating the intrinsic dimension of datasets by a minimal neighborhood information. DOI:10.1038/s41598-017-11873-y.

[8] Gerace, F., Loureiro, B., Krzakala, F., Mézard, M., & Zdeborová, L. (2020). Generalisation error in learning with random features and the hidden manifold model. arXiv:2002.09339.

[9] Rangan, S. (2010). Generalized Approximate Message Passing for Estimation with Random Linear Mixing. arXiv:1010.5141v1.

[10] Gomtsyan, M., Mokrov, N., Panov, M., & Yanovich, Y. (2019). Geometry-Aware Maximum Likelihood Estimation of Intrinsic Dimension. arXiv:1904.06151.

[11] Macocco, I., Glielmo, A., Grilli, J., & Laio, A. (2022). Intrinsic dimension estimation for discrete metrics. arXiv:2207.09688v2.

[12] Hein, M., & Audibert, J.-Y. (2006). Intrinsic Dimensionality Estimation of Submanifolds in R^d. DOI:10.1145/1102351.1102388.

[13] Hurlin, C., & Pérignon, C. (2023). Machine Learning and IRB Capital Requirements: Advantages, Risks, and Recommendations. DOI:10.2139/ssrn.4483793.

[14] Goldt, S., Mézard, M., Krzakala, F., & Zdeborová, L. (2021). Modeling the Influence of Data Structure on Learning in Neural Networks: The Hidden Manifold Model. DOI:10.1103/PhysRevX.10.041044.

[15] Gardner, E., & Derrida, B. (1978). Optimal storage properties of neural network models. DOI:10.1088/0305-4470/21/1/031.

[16] Fang, Y., Cao, Y., & Skeel, R. D. (2017). Quasi-Reliable Estimates of Effective Sample Size. arXiv:1705.03831v2.

[17] Abbara, A., Aubin, B., Krzakala, F., & Zdeborová, L. (2019). Rademacher complexity and spin glasses: A link between the replica and statistical theories of learning. arXiv:1912.02729v2.

[18] Unknown authors. (2022). Survey and Challenges of Mixed Data Clustering.

[19] Goldt, S., Loureiro, B., Reeves, G., & Zdeborová, L. (2020). The Gaussian equivalence of generative models for learning with shallow neural networks. arXiv:2006.14709v3.

[20] Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., & Goldstein, T. (2021). The Intrinsic Dimension of Images and Its Impact on Learning. arXiv:2104.08894v1.

[21] Gardner, E., & Derrida, B. (1979). Three unfinished works on the optimal storage capacity of networks. DOI:10.1088/0305-4470/22/12/004.