An Analysis of Dimensionality Expansion Performed on Quantum and Classical Support Vector Machines

Ryan Lam[1], Xiangyuan Ma[1]
[1]*Unionville High School, Markham, ON L3R 8G5, Canada*

March 31, 2020

Abstract

Ever since the discovery of the Support Vector Machine (SVM) by Vladimir Naumovich Vapnik in 1963, SVMs have been among one of the best supervised machine learning algorithms to classify linear and non-linear data. Unfortunately, SVMs possess kernel limitations when introduced to higher dimensional data and while working with problems containing large features.

With the introduction of quantum computation and quantum machine learning, researchers have successfully implemented a Quantum Support Vector Machine (QSVM) onto a quantum computer. Doing so has allowed classical data to be transferred into quantum states, taking advantage of the large dimensionality of quantum Hilbert space.

To study the effectiveness and accuracy of dimensionality expansion in QSVMs as opposed to SVMs, we processed dimensionally reduced data into SVMs and QSVMs. Using evaluation metrics and techniques the processed datasets are compared and analyzed.

Keywords: *Quantum Computing, Quantum Machine Learning, Quantum Support Vector Machine, Support Vector Machine, Dimensionality Expansion.*

Summary of Project: (remove this when done report)

In this project, we analyze and compare the classification accuracy of support vector machines (SVM) and quantum support vector machines (QSVM).We first artificially reduce a high-dimension dataset through performing different dimensionality reduction methods such as Principal Component Analysis (PCA). The reduced dataset is then processed by an SVM and QSVM, in-creasing the dimensionality of the data. The new processed datasets from the SVM and QSVM are then analyzed and compared using balanced accuracy and Area Under the Receiver Operating Characteristics (AUROC) techniques.

1 Introduction

example of ref: Hardaker, Richardson, Lien, and Schumann (2004)

2 Background

2.1 Dimensionality Reduction Techniques

In datasets with large volumes of data, it is believed that the data can be characterized in a lower dimensional-manifold. Dimensionality reduction is the process of reducing the training time and file size of a dataset for a supervised machine learning model, while maintaining high accuracy rates. Given a high-dimensional dataset, dimensionality reduction encodes the data into lower dimensions through removing redundant features and irrelevant data.

Dimensionality reduction is an important procedure in training machine learning models as it prevents overfitting due to *"The Curse of Dimensionality"*. The Curse of Dimensionality occurs because the sampling density is proportional to $N^{\frac{1}{p}}$, where N is the sample size and p is the number of dimensions in the input space. As the dimensionality and volume of a dataset increases, the feature space will increase, causing the model to lack well-generalization, leading to overfitting.

Dimensionality reduction is also able to provide a visualization of data in a....

In this section, we introduce the dimensionality reduction techniques used to perform our **experiment/study**.

2.1.1 Principal Component Analysis

2.1.2 Autoencoder

2.1.3 Latent Dirichlet Allocation

2.1.4 Independent Component Analysis

2.1.5 Subsubsection 5

2.1.6 Subsubsection 6

2.2 Support Vector Machines

2.3 Quantum Support Vector Machines

2.4 Evaluation Metrics

2.4.1 Balanced Accuracy

2.4.2 Area Under the Receiver Operating Characteristics

3 Methodology

text

4 Results

4.1 Subsection

text

5 Discussion

5.1 Subsection

text

6 Conclusions

text

example of a ref

References

Hardaker, J. B., Richardson, J. W., Lien, G., & Schumann, K. D. (2004, 6). Stochastic efficiency analysis with risk aversion bounds: a simplified approach. *The Australian Journal of Agricultural and Resource Economics*, *48*(2), 253–270. doi: 10.1111/j.1467-8489.2004.00239.x

Table 1: Example of a table

Variables	Categories	Unit	Rep	Mean	St. Dev.	Min	Max
Variable 1	Category A	$	8	0	0	0	0
	Category B	lb	8	22,411.20	6,325.90	13,819	31,201
	Category C	$	8	5,869.60	4,609.90	-464.1	12,744.10
Variable 2	Category A	$	8	1,777.40	144.5	1,642.30	1,912.60
	Category B	lb	8	21,444.80	5,146.90	15,096	28,032
	Category C	$	8	4,138.50	2,644.10	22.2	7,932.70
Variable 3	Category A	$	8	2,346.80	190.8	2,168.30	2,525.20
	Category B	lb	8	18,343.30	2,460.70	15,269.00	21,524.10
	Category C	$	8	3,699.20	2,549.80	1,299.10	8,709.80
Variable 4	Category A	$	8	2,288.80	186.1	2,114.80	2,462.90
	Category B	lb	8	23,450.40	4,172.50	20,045.00	32,363.00
	Category C	$	8	6,619.80	1,918.40	4,479.70	10,633.90
	CASE #1			14	6.61	6.9	27.9
	CASE #2			22.8	7.73	10.2	31.4

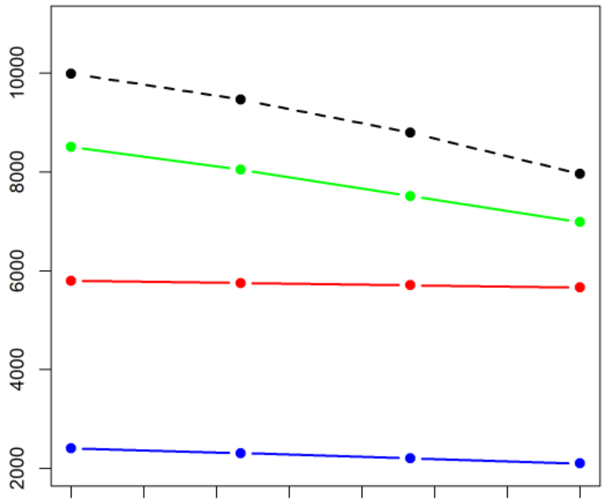

Figure 1: Example figure.