Article

Quantum Support Vector Regression for Disability Insurance

Boualem Djehiche * and Björn Löfdahl

Department of Mathematics, KTH Royal Institute of Technology, Lindstedtsvägen 25, 114 28 Stockholm, Sweden; bjornlg@kth.se

* Correspondence: boualem@kth.se

Abstract: We propose a hybrid classical-quantum approach for modeling transition probabilities in health and disability insurance. The modeling of logistic disability inception probabilities is formulated as a support vector regression problem. Using a quantum feature map, the data are mapped to quantum states belonging to a quantum feature space, where the associated kernel is determined by the inner product between the quantum states. This quantum kernel can be efficiently estimated on a quantum computer. We conduct experiments on the IBM Yorktown quantum computer, fitting the model to disability inception data from a Swedish insurance company.

Keywords: disability insurance; machine learning; support vector machines; quantum computing

1. Introduction

Support vector machines (SVM) were first introduced as part of Vapnik's Statistical Learning Framework (Vapnik 2013). Support vector classification aims to classify data, e.g., to determine if a picture contains a cat or a dog, whereas support vector regression (SVR) is used to model real-valued quantities, such as mortality rates or financial asset returns. All SVMs exploit the so-called kernel trick, where an optimization problem with data that have been mapped into a high- or even infinite-dimensional feature space may be efficiently solved by considering its Wolfe-dual (Schölkopf et al. 2000), for which the necessary input is reduced to a so-called kernel matrix consisting of inner products in the feature space between all data pairs. For cases where the kernel matrix can be readily determined, the corresponding optimization problem can be efficiently solved.

Rebentrost et al. (2014) showed that an SVM can be implemented on a quantum computer. This work was recently expanded on by Schuld and Killoran (2019) and Havlíček et al. (2019). In essence, two related methods have been proposed. The first method consists of encoding data in a high-dimensional quantum feature space, calculating a quantum kernel, and subsequently using a variational quantum circuit to find a separating hyperplane. A second approach proposes to use a quantum computer to estimate the kernel, and to implement the resulting SVM optimization on a classical computer, a so-called hybrid classical-quantum implementation. Quantum kernel methods can be efficiently used to solve some optimization problems where the kernel cannot efficiently be determined on a classical computer. This was recently demonstrated by Liu et al. (2021).

In the insurance literature, SVMs have been used as a mortality graduation technique (Kostaki et al. 2011). They could equally well be used to model other transition probabilities, such as disability inception or termination rates. These quantities are often estimated using classical techniques such as maximum likelihood or splines, see e.g., Aro et al. (2015); Djehiche and Löfdahl (2018); Christiansen et al. (2012); Renshaw and Haberman (1995, 2000). In this paper, we propose a hybrid classical-quantum SVR model for logistic disability inception probabilities, using a quantum kernel that can be estimated on a quantum computer. We conduct experiments on the IBM Yorktown quantum computer using disability inception data from a Swedish insurance company. The kernel that we use in this paper is simple enough that it can also be evaluated using classical methods. This allows us to readily compare the analytically derived kernel to the one estimated on a quantum computer. However,

the proposed approach is general and is in theory capable of estimating complex kernels based on high-dimensional data, where classical methods are intractable.

This paper is organised as follows. In Sections 2 and 3, we review kernel theory, support vector regression, and quantum kernel estimation. In Section 4, we propose a support vector regression model with a quantum kernel for disability inception rates. In Section 5, we estimate the kernel matrix associated with disability inception data from a Swedish insurance company on a quantum computer. This kernel is then used in a support vector regression to estimate disability inception rates. The results are compared to those from classical support vector regression.

2. Kernels and Support Vector Regression

In this section, we review kernel theory and support vector regression. Closely following Schuld and Killoran (2019), we let $x_i \in \mathbb{R}^d, i = 1, \ldots, n$, denote observations in a data set and let the mapping $\Phi : \mathbb{R}^d \mapsto \mathcal{F}$ be a feature map that maps a sample data point x to a feature vector $\Phi(x)$ in a (usually higher-dimensional) feature space $\mathcal{F}$, usually taken as a Hilbert space. The mapping Φ naturally gives rise to a so-called kernel through the relation

$$K(x,z) = \langle \Phi(x), \Phi(z) \rangle, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathcal{F}$. Note that, since $K(x,z)$ is determined by the inner product of $\Phi(x)$ and $\Phi(z)$, it can be seen as a similarity measure between x and z in the feature space. The reproducing kernel Hilbert space (RKHS) associated with Φ is defined by

$$\mathcal{R} = \{f : \mathbb{R}^d \mapsto \mathbb{C}; \ f(x) = \langle w, \Phi(x) \rangle \ \forall \, x \in \mathbb{R}^d, w \in \mathcal{F}\}. \tag{2}$$

Note that the functions $\langle w, \Phi(x) \rangle$ can be interpreted as linear models in the feature space $\mathcal{F}$. Now, assume that we are given a cost function $\mathcal{C}$ that measures the goodness of fit of a model by comparing predicted values $\{f(x_i)\}_i$ with observed values $\{y_i\}_i$ and that has a regularization term $g(\|f\|)$, where g is a strictly increasing function. Then, any function $f \in \mathcal{R}$ that minimizes the cost function $\mathcal{C}$ can be written as

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i), \tag{3}$$

for some parameters $\alpha_i \in \mathbb{R}, \ i = 1, \ldots, n$.

Perhaps the most famous application of the kernel approach is support vector regression (Vapnik 2013). SVR can be formulated as a convex optimization problem of the form

$$\text{P:} \min_{w,b,\xi,\xi'} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i')$$

$$\text{s.t.} \quad (w^T\Phi(x_i) + b) - y_i \leq \varepsilon - \xi_i, \qquad\qquad i = 1, \ldots, n,$$
$$y_i - (w^T\Phi(x_i) + b) \leq \varepsilon - \xi_i', \qquad\qquad i = 1, \ldots, n,$$
$$\xi_i, \xi_i' \geq 0, \qquad\qquad i = 1, \ldots, n,$$

where ε determines the error tolerance of the solution, C is a regularization parameter, and $\xi_i \in \mathbb{R}$ and $\xi_i' \in \mathbb{R}, i = 1, \ldots, n$, are slack variables. It can be shown (Schölkopf et al. 2000) that the dual formulation of P is given by

$$\text{D:} \max_{\lambda,\lambda'} -\frac{1}{2}\sum_{i,j=1}^{n}(\lambda_i - \lambda_i')(\lambda_j - \lambda_j')K(x_i, x_j) - \varepsilon\sum_{i=1}^{n}(\lambda_i - \lambda_i') + \sum_{i=1}^{n}y_i(\lambda_i - \lambda_i')$$

$$\text{s.t.} \quad \sum_{i=1}^{n}(\lambda_i - \lambda_i') = 0,$$
$$0 \leq \lambda_i \leq C, i = 1, \ldots, n,$$
$$0 \leq \lambda_i' \leq C, i = 1, \ldots, n,$$

and that the solutions of P and D coincide and are given by

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + \beta, \tag{4}$$

where $\alpha_i = \lambda_i - \lambda_i'$. In order to fit the model (4) to data, we must first determine the kernel matrix $K = \{K_{ij}\}, i, j = 1, \ldots, n$, where $K_{ij} = K(x_i, x_j)$. In the classical paradigm, we would choose a tractable kernel such as the kernel corresponding to radial basis functions (the so-called Gaussian kernel), evaluate the kernel matrix, and finally, fit the model to data by solving the optimization problem D. An alternative to classical kernels is provided by the so-called quantum kernels, which we will briefly review in the following section.

3. Quantum Kernel Estimation

In quantum kernel estimation, the kernel is determined by a quantum feature map. Following Rebentrost et al. (2014); Schuld and Killoran (2019) and Havlíček et al. (2019), we let $\Phi : x \mapsto \Phi(x)$ (or $|\Phi(x)\rangle$ using Dirac's notation) denote a quantum feature map that maps a data point to a quantum state which is an element $\Phi(x)$ of a Hilbert space $\mathcal{H}$. Any quantum state $\psi \in \mathcal{H}$ naturally satisfies the famous Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(t, x) = H\psi(t, x), \quad \psi(0, \cdot) \in \mathcal{H} \text{ is given}, \tag{5}$$

where H is the Hamiltonian operator associated with the quantum system. If H is time-independent, the solution to (5) is given by

$$\psi(t, x) = U(t)\psi(0, x), \tag{6}$$

where the operator U defined by

$$U(t) = e^{-iHt/\hbar}, \tag{7}$$

which is the unitary time evolution operator associated with H. Thus, in analogy with (6), using the characterization of reproducing kernel Hilbert spaces (RKHS), it can be shown (see Schölkopf et al. (2001); Schuld and Killoran (2019), and the references therein) that, for every pair (Φ, x), there is an operator $U_\Phi(x)$, known in the field of quantum computing as a *feature embedding circuit*, that is implicitly determined by the relation

$$\Phi(x) = U_\Phi(x)\Omega_0, \tag{8}$$

where Ω_0 (also denoted $|0 \ldots 0\rangle$ using Dirac's notation) denotes the ground state, i.e., the quantum state with the lowest energy level (associated with the smallest eigenvalue of the generator of the operator $U_\Phi(x)$). Further, let the kernel K corresponding to Φ be given by

$$K(x, z) = |\langle \Phi(x), \Phi(z) \rangle|^2. \tag{9}$$

As mentioned above, $K(x, z)$ it is essentially a similarity measure between x and z in the quantum feature space $\mathcal{H}$. It should be noted that the definition of the kernel (9) deviates from the form (1) that is common in the classical literature, in that it involves taking the absolute value squared of the inner product. This is due to the following: Using (8), the kernel can be written as

$$K(x, z) = |\langle \Phi(x), \Phi(z) \rangle|^2 = |\Omega_0^\dagger U_\Phi^\dagger(z) U_\Phi(x) \Omega_0|^2, \tag{10}$$

that is, $K(x, z)$ is given by the probability of obtaining the measurement outcome Ω_0 when measuring the quantum state $\Psi(x, z)$ defined by

$$\Psi(x, z) = U_\Phi^\dagger(z) U_\Phi(x)\Omega_0, \tag{11}$$

where $U^\dagger$ denotes the adjoint operator of U. The probability (10) can be estimated on a quantum computer by loading the state $\Psi(x,z)$ into a quantum circuit. This circuit is then run multiple times, and (10) is estimated by the frequency of Ω_0-measurements. We note that the form of the kernel (9) is what allows us to readily estimate it using a quantum computer.

Naturally, the procedure outlined above introduces a sampling error. Assuming that the measurements are independent, we have that $\widehat{p}_n$, the frequency of Ω_0-measurements after n trials (or shots), has standard error $\sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}}$. Today's noisy and rather primitive quantum computers are unfortunately quite error prone, which means that, in addition to the sampling error, there is a large and often systematic error arising from imperfect hardware. Given a large enough number of trials, these hardware errors tend to dominate the sampling error. The hardware-related errors can be partially mitigated using error correction techniques (see, e.g., Temme et al. (2017)).

4. Model Description

We consider a population of insured individuals, divided into subgroups based on some common characteristics. Let E_i be the number of healthy individuals from the population subgroup i, $i = 1, \ldots, n$, in a given disability insurance scheme. We denote by D_i the number of individuals falling ill amongst the E_i insured healthy individuals. For each population subgroup i, there is some associated data $x_i \in \mathbb{R}^d$, which may, e.g., contain information about age, gender, and other characteristics of the population subgroup at hand. We assume that the conditional distribution of D_i given E_i is binomial:

$$D_i \sim \text{Bin}(E_i, p(x_i)), \tag{12}$$

where $p(x_i)$ is the probability that an individual randomly selected from E_i falls ill. We propose to model the logistic disability inception probabilities using support vector regression:

$$\text{logit } p(x) = \log \frac{p(x)}{1 - p(x)} = \sum_{i=1}^{n} \alpha_i K(x, x_i) + \beta, \tag{13}$$

where $K \in \mathbb{R}^{n \times n}$ is a kernel matrix associated with the data $\{x_i\}_i$, , and $\beta \in \mathbb{R}$ and $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, n$, are parameters to be estimated from historical data. We propose to fit the model using a weighted support vector regression, with the weight for each sample proportional to the population subgroup size E_i, placing higher importance on large subgroups where the sampling errors of the observed inception probabilities are lower. The logistic transform guarantees that the probabilities estimated from the model lie in their natural interval $(0, 1)$.

In the classical paradigm we would choose a tractable kernel such as the kernel corresponding to radial basis functions or the linear kernel. We propose instead that this kernel be calculated using a quantum feature map, to be evaluated on a quantum computer. In order to fit the model (13) on a quantum computer, we must first choose a specific quantum feature map. This, in turn, determines the layout of our quantum circuit through (8). There are many ways to choose a suitable feature map (see, e.g., Schuld and Killoran (2019)). We suggest to choose Φ, such that it captures the richness of the data x while still being simple enough to be run on today's limited and noisy quantum computers. For simplicity, we will now assume that our data are two-dimensional, i.e., $x_i \in \mathbb{R}^2$. Then, it is enough to use a two-qubit unitary operator to obtain estimates of K. To this end, we choose the unitary operator given by
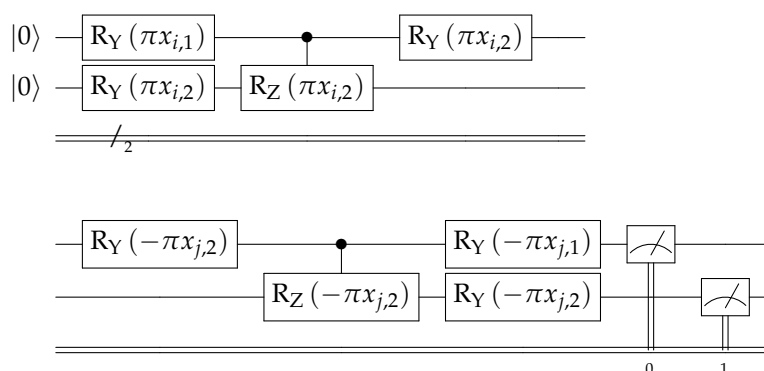
$$U_\Phi(x_i) = \left(I \otimes \mathrm{R_Y}(\pi x_{i,2}) \right) C_{\mathrm{R_Z}}(\pi x_{i,2}) \left(\mathrm{R_Y}(\pi x_{i,2}) \otimes \mathrm{R_Y}(\pi x_{i,1}) \right), \tag{14}$$

where $\mathrm{R_Y}$ and $\mathrm{R_Z}$ denote rotations around the Y and Z axes of the Bloch sphere (a.k.a. the Riemann sphere), respectively, and $C_{\mathrm{R_Z}}$ denotes a controlled $\mathrm{R_Z}$ operation on the second qubit, using the first qubit as a control. A graphical representation of the quantum circuit that implements this unitary is presented as follows:

$$q_0: \quad \boxed{R_Y(\pi x_{i,1})} \quad \bullet \quad \boxed{R_Y(\pi x_{i,2})}$$
$$q_1: \quad \boxed{R_Y(\pi x_{i,2})} \quad \boxed{R_Z(\pi x_{i,2})}$$
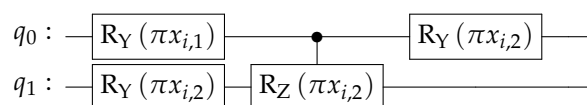
To facilitate interpretation, we let $x_{i,1}$ be a dummy variable taking the value 1 if the population subgroup is male, and 0 otherwise, and $x_{i,2}$ be the associated age of the population subgroup, measured in centuries. First, we apply a $R_Y(\pi x_{i,1})$-gate to q_0. This flips q_0 from $|0\rangle$ to $|1\rangle$ for male subgroups. Then, we perform a $R_Y(\pi x_{i,2})$ rotation on q_1. This operation rotates the state of q_1 from $|0\rangle$ towards $|1\rangle$, with the angle of rotation increasing as the age of the subgroup increases. The $C_{R_Z}(\pi x_{i,2})$ performs an additional rotation around the Z-axis of the Bloch sphere, with the angle of rotation increasing as the age of the subgroup increases. Note that this rotation is only performed if q_0 is in the state $|1\rangle$, i.e., if the subgroup is male. Finally, we perform a $R_Y(\pi x_{i,2})$ rotation on q_0.

For each data pair (x_i, x_j), we run this quantum circuit inserting the values of x_i and then run the adjoint circuit inserting the values of x_j. Finally, we perform a measurement on the two qubits. This circuit is run multiple times, and $K(x_i, x_j)$ is estimated by the frequency of obtaining the measurement $\Omega_0 := |00\rangle$. The resulting quantum circuit can be graphically represented as

$$|0\rangle \quad \boxed{R_Y(\pi x_{i,1})} \quad \bullet \quad \boxed{R_Y(\pi x_{i,2})}$$
$$|0\rangle \quad \boxed{R_Y(\pi x_{i,2})} \quad \boxed{R_Z(\pi x_{i,2})}$$

$$\boxed{R_Y(-\pi x_{j,2})} \quad \bullet \quad \boxed{R_Y(-\pi x_{j,1})} \quad \text{⊿}$$
$$\boxed{R_Z(-\pi x_{j,2})} \quad \boxed{R_Y(-\pi x_{j,2})} \quad \text{⊿}$$
$$\qquad\qquad\qquad\qquad 0 \qquad 1$$

This circuit is designed to clearly separate male and female subgroups and to gradually increase the dissimilarity between different age groups as the difference in ages increases. Note that, for $i = j$, all rotations cancel out, and the circuit will measure the state $|00\rangle$ with probability 1, i.e., $K(x_i, x_i) = 1$ as expected.

5. Numerical Results

In this section, we estimate the kernel matrix associated with disability inception data from a Swedish insurance company. This kernel is then used in a support vector regression to estimate the logistic disability inception rates. The data consist of inception counts for 81 groups of individuals as well as the associated age and gender for each group.

5.1. Estimating the Kernel Matrix

We estimate the kernel matrix using the circuit from the previous section with two different techniques. Using (10) and (14), we classically compute $K(x_i, x_j)$ for each pair (x_i, x_j) by matrix multiplication. Here, classically computing the kernel is possible due to the simple structure and low dimension of the unitary operator (14). This procedure is hereafter referred to as a state vector simulation. Figure 1 displays the estimated kernel matrix. This matrix has an interesting structure: it is block-diagonal. This is due to the fact that the second quadrant of the matrix correspond to the inner products of the female population groups. These share the common characteristic 'female', and each row is similar to its neighbours due to the encoding: similar ages are also similar in the quantum feature space. Analogously, the fourth quadrant of the matrix contains the male population

groups. The first and third quadrants contain the inner products between male and female population groups and so are dissimilar in the quantum feature space.
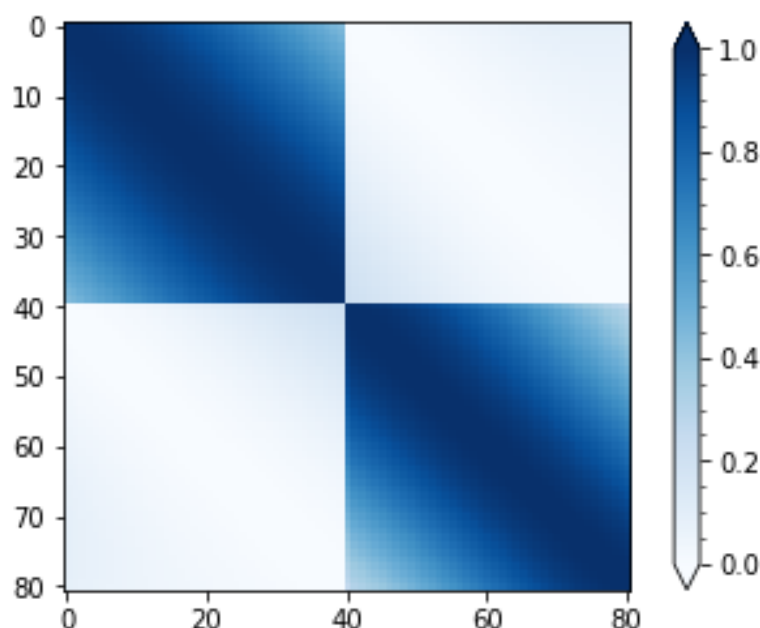

Figure 1. Kernel matrix determined by state vector simulation.

Next, we run multiple experiments on the IBM Yorktown five-qubit quantum computer (of which we use two qubits) to obtain an estimate of (10). For each data pair (x_i, x_j), we run the circuit 8192 times, measure the outcomes, and estimate $K(x_i, x_j)$ with the observed frequency of the ground state. Note, especially, that, for an ideal quantum computer without hardware related errors, this frequency converges to the probability of measuring the ground state when the number of shots goes to infinity. This means that any difference between the kernel from the state vector simulation and the kernel estimated on the quantum computer is related to sampling error and hardware-related errors. Indeed, for an ideal quantum computer using an infinite number of shots, the estimated kernels would be identical.

Figure 2 displays the estimated kernel matrix. We note that, as expected, this matrix deviates somewhat from the kernel matrix obtained by state vector simulation, but it has the same characteristics of the block-diagonal structure and an increasing dissimilarity with increasing age difference. We also note that, using 8192 shots, the maximum standard error under the assumption of ideal hardware (obtained at $\hat{p} = 0.5$) is given by $\sqrt{\frac{0.5(1-0.5)}{8192}} \approx 0.0055$. Noting that this error is small compared to the errors resulting from imperfect hardware, we will focus on the systematic error from now on.

On a five-qubit quantum computer, the natural ground state is the $|00000\rangle$ state. When running a circuit that only performs operations on, say, the first two qubits, we would expect that qubits three through five are always measured to be in their respective $|0\rangle$ state. However, due to measurement error, this is not the case. Hence, we propose to mitigate this problem through the following simple approach: Instead of measuring the frequency of the natural ground state, we measure the sum of the frequencies of all $|00m_1 m_2 m_3\rangle$ states for $m_k \in \{0, 1\}, k = 1, 2, 3$.

Figure 3 displays the first row of the kernel matrix, estimated from state vector simulation and from the Yorktown quantum computer, as well as the corresponding error-mitigated kernel. The error mitigation procedure vastly improves the kernel estimate for the first half of the columns and only slightly worsens it for the second half of the columns. The estimation error, measured as the distance from the state vector simulation kernel in the

Frobenius norm, is reduced from 15.18 to 8.07, indicating that the approach is functioning as intended.
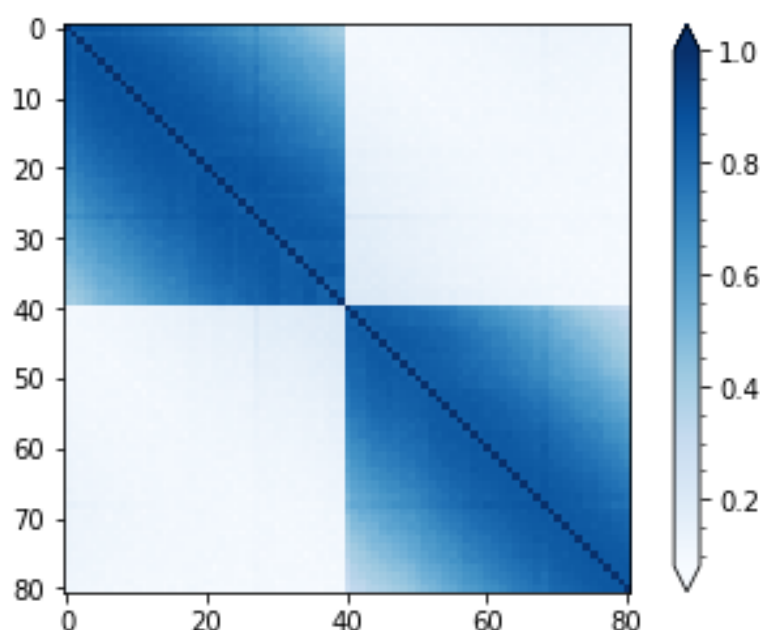

Figure 2. Kernel matrix estimated on the IBM Yorktown quantum computer.
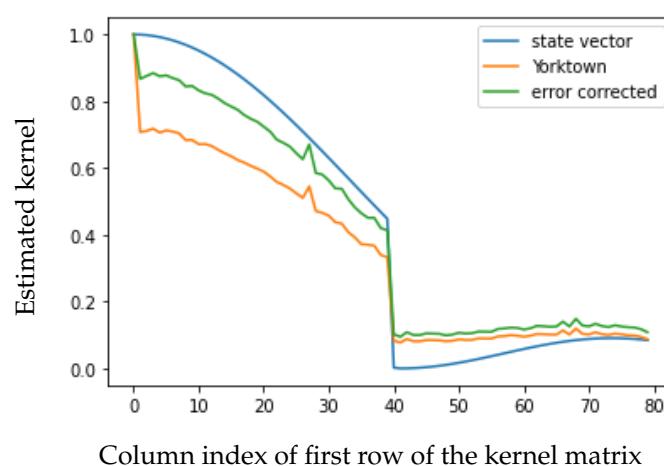

Column index of first row of the kernel matrix

Figure 3. First row of the kernel matrix, estimated from state vector simulation and from the Yorktown quantum computer, as well as the corresponding error-mitigated kernel.

5.2. Fitting Swedish Disability Inception Rates

We now fit the disability inception model (13) to data with support vector regression, using four classical kernel methods, i.e., the linear kernel, a polynomial kernel of rank 3, the radial basis functions kernel, and a sigmoid kernel. In addition, we fit the model to data using the quantum kernels based on a state vector simulation as well as from the IBM Yorktown 5-qubit quantum computer. The models are fit using leave-one-out cross-validation, so that, for each train–test split, a single out-of-sample logistic disability inception rate is estimated. After applying the inverse logistic function to obtain a disability inception rate, we then calculate the weighted R^2 statistic for the out-of-sample rates, again using the population counts as weights. The results are presented in Table 1. The state vector quantum kernel performs better than three out of the four classical kernels, the exception being the polynomial kernel. The Yorktown quantum kernel is only slightly worse compared to the state vector simulation.

Table 1. Weighted out-of-sample R^2 for the classical and quantum kernels.

Kernel	R^2
linear	0.426
polynomial	0.550
radial basis functions	0.529
sigmoid	0.494
state vector quantum kernel	0.541
Yorktown quantum kernel	0.518

The out-of-sample estimates are displayed in Figure 4. Note that, due to confidentiality, the actual values of the estimates are not reported. The support vector regression approach manages to capture the difference between the genders, as well as finding a pattern in the age dimension. The middle-aged population groups are larger than the others, meaning that the highest weights will be placed on these ages for the purposes of calibrating the model. The observations with very high or very low ages are considered as outliers by the model and so are virtually ignored.
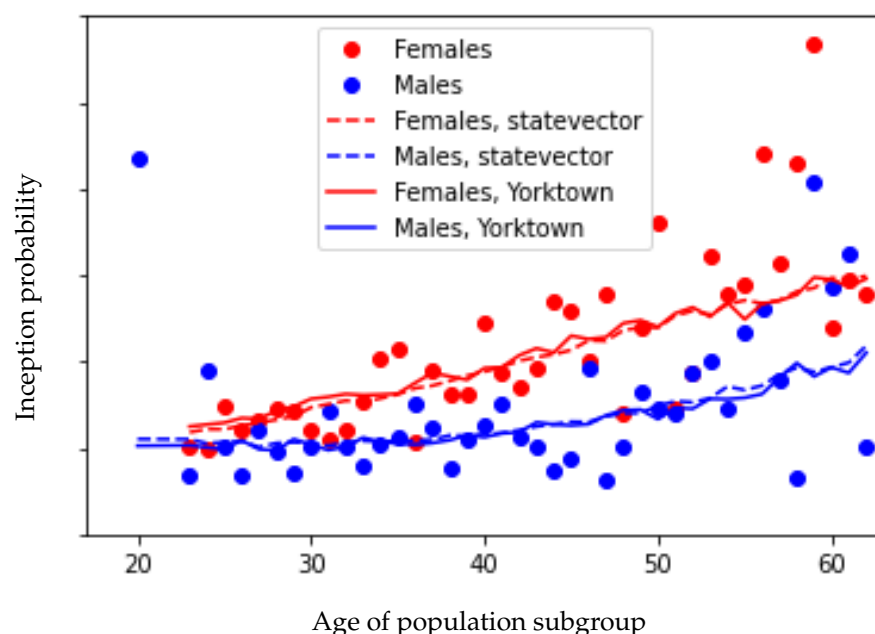
Figure 4. Out-of-sample disability inception rates estimated by state vector simulation and from the IBM Yorktown quantum computer.

We note, especially, that the estimated inception rates from the Yorktown quantum kernel are comparable to the ones obtained from the state vector simulation, even though the estimated quantum kernels were themselves quite different. We believe that this is due to the fact that the characteristics of the kernel, namely the block-diagonal structure and the increasing dissimilarity with increasing age difference, were preserved, even though the actual kernel estimates differed significantly from each other. This fact proposes some degree of robustness to the model: even though the quantum kernel is estimated with significant error, the resulting disability inception estimates are only marginally impacted.

6. Conclusions

We have proposed a hybrid classical-quantum approach for modeling transition probabilities in health and disability insurance. We formulated the modeling of logistic disability inception probabilities as a support vector regression problem. The associated kernel is determined by the inner product between the quantum states from a quantum

feature space. This procedure is made possible due to the ability to construct a 'quantum' feature map between the data points and the quantum states.

The kernel used in this paper is simple enough that it can also be evaluated using classical methods. This is what allows us to compare the analytically derived kernel to the one estimated on a quantum computer. However, the proposed approach is general and is, in theory, capable of estimating complex kernels based on high-dimensional data, where classical methods are intractable.

Since our proposed model fits the data well and produces errors that are comparable to today's classical methods, we conclude that estimating disability inception rates with quantum support vector regression is a viable statistical method, even on today's noisy quantum computers. This bodes well for the future where complex and high-dimensional data might well be modeled and fitted accurately to data in a timely fashion using quantum computers.

Author Contributions: Software, B.L.; data curation, B.L.; writing—original draft preparation, B.L.; writing—review and editing, B.D. and B.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Due to non-disclosure agreements, the data is not publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

Aro, Helena, Boualem Djehiche, and Björn Löfdahl. 2015. Stochastic modelling of disability insurance in a multi-period framework. *Scandinavian Actuarial Journal* 2015: 88–106. [CrossRef]

Christiansen, Marcus, Michel Denuit, and Dorina Lazar. 2012. The Solvency II square-root formula for systematic biometric risk. *Insurance: Mathematics and Economics* 50: 257–65. [CrossRef]

Djehiche, Boualem, and Björn Löfdahl. 2018. A hidden markov approach to disability insurance. *North American Actuarial Journal* 22: 119–36. [CrossRef]

Havlíček, Vojtěch, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. 2019. Supervised learning with quantum-enhanced feature spaces. *Nature* 567: 209–12. [CrossRef] [PubMed]

Kostaki, Anastasia, Javier M. Moguerza, Alberto Olivares, and Stelios Psarakis. 2011. Support vector machines as tools for mortality graduation. *Canadian Studies in Population* 38: 37–58. [CrossRef]

Liu, Yunchao, Srinivasan Arunachalam, and Kristan Temme. 2021. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics* 17: 1013–17. [CrossRef]

Rebentrost, Patrick, Masoud Mohseni, and Seth Lloyd. 2014. Quantum support vector machine for big data classification. *Physical Review Letters* 113: 130503. [CrossRef] [PubMed]

Renshaw, Arthur, and Steven Haberman. 1995. On the graduations associated with a multiple state model for permanent health insurance. *Insurance: Mathematics and Economics* 17: 1–17. [CrossRef]

Renshaw, Arthur, and Steven Haberman. 2000. Modelling recent time trends in UK permanent health insurance recovery, mortality and claim inception transition intensities. *Insurance: Mathematics and Economics* 27: 365–96. [CrossRef]

Schölkopf, Bernhard, Ralf Herbrich, and Alex J. Smola. 2001. A generalized representer theorem. In *International Conference on Computational Learning Theory*. Berlin and Heidelberg: Springer, pp. 416–26.

Schölkopf, Bernhard, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. 2000. New support vector algorithms. *Neural Computation* 12: 1207–45. [CrossRef] [PubMed]

Schuld, Maria, and Nathan Killoran. 2019. Quantum machine learning in feature hilbert spaces. *Physical Review Letters* 122: 040504. [CrossRef] [PubMed]

Temme, Kristan, Sergey Bravyi, and Jay M. Gambetta. 2017. Error mitigation for short-depth quantum circuits. *Physical Review Letters 119*: 180509.

Vapnik, Vladimir. 2013. *The Nature of Statistical Learning Theory*. Berlin and Heidelberg: Springer Science & Business Media.