

東京大学大学院新領域創成科学研究科  
メディカル情報生命専攻

2019 年度

修士論文

量子アニーリングを用いた特徴選択  
Feature Selection with Quantum Annealing

2020 年 1 月 17 日提出

指導教員 津田 宏治 教授

浅井 智佳子

## 要 旨

近年、機械学習分野において大規模データ解析が盛んに行われており、データの次元を絞る特徴選択は、推論時間の削減やモデルの精度、解釈性の向上、過学習抑止のために重要な手法である。特徴選択はナীবに最も良い次元の組合せを探索すると、入力次元に対して指数時間かかってしまうため、効率の良い手法を考える必要がある。一方、D-Wave 社が実用化した量子アニーリングマシンは、組み合わせ最適化問題を、量子力学を用いて高速に解くことができるため、注目を集めている。ただし、アニーリングマシンを特徴選択に使用するためには、問題をイジングモデルに帰着させる必要があるので、必ずしも容易ではない。1Qbit 社による従来手法では、特徴選択問題を、特徴量間の相関を最小化する問題として定式化しているが、汎化誤差のモデルを用いていないため、性能には限界がある。本論文では、任意の機械学習モデルの交差検証誤差を Factorization Machine によってモデリングし、特徴数を指定する罰則項を加えることで、特徴選択問題をイジングモデルに帰着させる方法を提案する。ロジスティック回帰と、線形回帰に対して提案法を適用した結果、LASSO や、1Qbit 社による手法よりも、同数の特徴量で、優れた汎化誤差を達成できることを確かめた。D-Wave の制限のため、本手法は、64 次元以内の問題にしか適用できないが、今後、量子アニーラの規模が増大すれば、生命科学分野における、より大きな問題にも適用できると考えられる。

# 目次

<b>第 1 章</b>	<b>はじめに</b>	<b>3</b>
1.1	背景と目的 . . . . .	3
1.2	主結果 . . . . .	4
1.3	本論文の構成 . . . . .	4
<b>第 2 章</b>	<b>準備</b>	<b>5</b>
2.1	基本的な定義 . . . . .	5
2.2	量子アニーリング . . . . .	6
2.3	Factorization Machine . . . . .	7
2.4	1QBit 社の手法 . . . . .	7
<b>第 3 章</b>	<b>量子アニーリングを用いた特徴選択手法</b>	<b>10</b>
3.1	FMQA . . . . .	10
3.2	ビット制約 . . . . .	11
3.3	FMQA+を用いた特徴選択アルゴリズム . . . . .	12
<b>第 4 章</b>	<b>実験</b>	<b>14</b>
4.1	実験環境 . . . . .	14
4.1.1	ニオイデータ . . . . .	14
4.1.2	German Credit Dataset . . . . .	15
4.1.3	バンドギャップデータ . . . . .	15
4.2	実験結果 . . . . .	16
4.2.1	結果 1: FMQA の最適化性能 . . . . .	16
4.2.2	結果 2: ビット制約におけるペナルティ強度とチェーン強度のチューニング . . . . .	16

4.2.3 結果 3: ビット制約をかけた時の性能比較 . . . . .	18
<b>第 5 章 おわりに</b>	<b>23</b>
<b>参考文献</b>	<b>25</b>

# 第1章

## はじめに

### 1.1 背景と目的

情報技術の発展により、機械学習分野において大規模データ解析が盛んに行われている。機械学習の目的のひとつはデータに潜む本質的な構造を明らかにすることであり、中でもデータの次元を絞る特徴選択は、推論時間の削減やモデルの精度、解釈性の向上、過学習の抑止のために重要な手法である [1]。特徴選択の手法には大きく分けて Filter と Embedded, Wrapper の 3 つがある。Filter は前処理の段階で個々の特徴量を評価し、特徴選択を行う手法であり、Embedded はモデルの学習と同時に特徴選択を行う手法である。Wrapper は可能な特徴空間を探索し、それぞれの部分集合で機械学習モデルの精度を測り、どの特徴を用いるか決定する手法であり、これは様々な機械学習モデルに適用可能な一方でナイーブに最も良い次元の組合せを探索すると、特徴数  $n$  に対し  $O(2^n)$  の計算時間がかかってしまい、実行は現実的に不可能となる。そこで、計算資源やコストの観点から効率の良い方法を考える必要がある。

近年、組合せ最適化問題を高速・高精度に処理する技術として量子アニーリングが注目を集めている。組合せ最適化問題とは、離散変数をとるコスト関数が与えられた制約のもとで、最小または最大値をとるような離散変数の組合せを求める問題であり、実世界の様々な問題が組合せ最適化問題に落とし込めることから、これを高速・高精度に解く技術は非常に重要である。量子アニーリングは 1998 年に物理学に基づく計算技術として提案された [3]。具体的には、解候補を物理的状態とみなし、イジングモデルという統計力学モデルにおける量子ゆらぎ効果による状態間の遷移を用いて解を探索するという手続きを

とる。その後、この動作原理を実装したハードウェアとして、超電導集積回路技術を用いた量子アニーリングマシンが開発された [2]。このマシンでは、イジングモデルにおける量子ゆらぎ効果を、横磁場効果として導入し、それを徐々に弱めることで量子アニーリングの動作原理を実装している。組合せ最適化問題はイジングモデルに帰着することができるので、この量子アニーリングマシンは組合せ最適化問題を解くことができる。2011 年に発表された世界初の商用量子アニーラに搭載された量子ビット数は 128 個であったが、ビットの数の増加や量子ノイズの低減など、様々な技術的向上が行われてきた。それらを踏まえてアニーリングマシンの様々な応用の研究が盛んに行われ、次世代コンピューティングとしての期待が高まっている。

これらの背景を踏まえて、機械学習モデルにおける特徴選択を組合せ最適化問題に帰着させ、量子アニーラによる特徴選択を行う方法を考える。

## 1.2 主結果

本論文では、任意の機械学習モデルの交差検証誤差を Factorization Machine によってモデリングし、特徴数を指定する罰則項を加えることで、特徴選択問題をイジングモデルに帰着させる方法を提案する。そして実際に量子アニーラの実機を用いて実験を行い、ロジスティック回帰と、線形回帰に対して提案法を適用した結果、LASSO や、1Qbit 社による手法よりも、同数の特徴量で、優れた汎化誤差を達成できることを確かめた。

## 1.3 本論文の構成

2 章では、以降の章で必要な基本的な定義を与える。また量子アニーリングと Factorization Machine, QBit 社の提案する量子アニーリングを用いた特徴選択手法を説明する。3 章では、FMQA を説明し、FMQA にビット制約を加えた FMQA+を導入し、FMQA+を用いた特徴選択手法を与える。4 章では、FMQA+を用いて機械学習モデルにおける特徴選択をした結果とその他の特徴選択手法との比較結果について述べる。5 章では、本論文をまとめる。

## 第2章

### 準備

本章では, 以降の章で必要な用語と定義を与え, 量子アニーリングと Factorization Machine, それらを組み合わせた FMQA について説明し, カナダの 1QBit 社が提案する量子アニーリングを用いた特徴選択手法 [5] について述べる.

#### 2.1 基本的な定義

集合  $\{X, Y\}$  ( $X = \{x_i : i = 1, \dots, m\}, Y = \{y_i : i = 1, \dots, m\}$ ) をデータセットという.  $x_i$  は  $n$  個の特徴をもつ入力変数  $x_{i,j}$  ( $j = 1, \dots, n$ ) からなる. このとき, データセットにおける特徴数/特徴次元とは  $x_{i,j}$  ( $j = 1, \dots, n$ ) の大きさであり,  $n$  である. またここでは, 出力変数  $Y$  をターゲットベクトルと呼び, その値域をターゲットクラス  $T$  と呼ぶ.  $T$  の実現値が  $y_i \in \{c_1, \dots, c_l\}$  のように離散化されている場合, その大きさ  $l$  をターゲットクラス数と呼ぶ. 機械学習モデルにおける教師あり学習アルゴリズムの目的とはデータセットを用いて学習を行い, 新しいデータ  $\mathbf{x}' \in \mathbb{R}^n$  に対する  $\hat{y}$  を決定する関数  $y : \mathbb{R}^n \rightarrow T$  を推定することである.  $T = \mathbb{R}$  のとき回帰問題といい,  $T = \{c_i : i = 1, \dots, l, l \in \mathbb{N}\}$  のとき分類問題という. 学習にはデータセットの一部を使用し, データセット中の学習に用いていないデータ  $\mathbf{x}^*$  をモデルに入力し, そこから予測値  $\hat{y}$  得るとする. このとき, 予測値  $\hat{y}$  と真値  $y^*$  の差を評価する関数を定義し, 得られた値をこれを予測精度や予測誤差という. 特徴選択とは  $n$  個ある特徴の集合からモデルの予測精度や予測誤差が向上するような  $k$  個の特徴の部分集合を選ぶことである.

## 2.2 量子アニーリング

量子アニーリングとは西森らによって開発された [] 量子効果を用いて組合せ最適化問題を解く手法のことである。この手法が対象とする組合せ最適化問題は、エネルギーが

$$H(\sigma) = \sum_{i<j} J_{ij}\sigma_i\sigma_j + \sum_{i=1}^N h_i\sigma_i, \quad \sigma_i \in \{-1, +1\}$$

で与えられる系 (イジング模型) の基底状態 (最低エネルギー状態) を求める問題とみなすことができる。ここで  $+1, -1$  の値をとる  $\sigma_i$  はイジングスピンと呼ばれ、 $N$  はスピンの数を表す整数であり、一般に大きな値をとる。また  $J_{ij}$  はスピン間の相互作用、 $h_i$  は局所磁場を調整するパラメータである。つまりこの問題は、 $\{J_{ij}, h_i\}$  が与えられた時、エネルギー  $H(\sigma)$  を最小にする  $\sigma$  を求める問題といえる。量子アニーリングでははじめ、量子効果 (横磁場項) を大きくし、各スピンの状態を不確定、つまり  $\pm 1$  を等確率で重ね合わせた状態に設定する。そして、量子効果を徐々に小さくすることで、相互作用  $J_{ij}$  や局所磁場  $h_i$  の影響を強くしていく。すると各スピンの状態が  $\pm 1$  のどちらかに確定した状態が自律的に選ばれ、それがイジング模型の基底状態となる。つまり、解きたい問題を  $+1, -1$  の値をとる変数を持つ (2.1) 式に落とし込み、量子アニーリングを行うと、その問題における目的関数  $H(\sigma)$  が最低エネルギーをとるような  $\sigma$  が確率的に求まる。カナダの D-Wave 社は量子アニーリングの動作原理が実装されたハードウェアを世界で初めて開発し、現在ではクラウド経由で利用できる D-Wave2000Q が提供されている [2]。現在 D-Wave2000Q で扱うことのできる最大スピン数は 2048 個である。この D-Wave2000Q で採用されているグラフ構造はキメラグラフと呼ばれ、任意の  $i, j (i < j)$  の組に対して  $J_{i,j}$  を設定できるわけではない。そのためには、複数のスピンを仮想的な 1 つのスピンとみなすことで、全結合グラフを再現する必要がある。全結合グラフを作る際に扱えるスピン数は、最大で 64 となる。



## 2.3 Factorization Machine

Factorization Machine(FM) はモデル式が

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (2.1)$$

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{k=1}^K v_{ik} v_{jk} \quad (2.2)$$

$$w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{R}^{n \times K}, \quad k \in \mathbb{N}_0^+ \quad (2.3)$$

の形で表される, 教師あり学習モデルである. ハイパーパラメータ  $K$  を与えた際に, スカラー値の  $w_i$  と  $K$  次元のベクトル  $\mathbf{v}_i$  を予測精度が最も良くなるように決める. FM のモデル式 (2.1) では, 単独の特徴  $x_i$  に対する重みは  $w_i$  を, 特徴同士の組み合わせに対する重みには内積  $\mathbf{v}_i \cdot \mathbf{v}_j$  を用いている. 学習には SGD などの最急降下法を用いることができ, 一回の更新を  $\mathcal{O}(Kn)$  で行うことができる [6].

## 2.4 1QBit 社の手法

量子アニーラを用いた特徴選択手法として 1QBit 社が開発した手法がある [5]. 本節では, この手法を紹介する. データセット  $\{U, \mathbf{V}\}$  が  $n$  個の特徴をもつ  $m \times n$  行列  $U$ ,

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ u_{21} & u_{22} & u_{23} & \dots & u_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & u_{m3} & \dots & u_{mn} \end{bmatrix}$$

$m$  個の要素からなるターゲットベクトル  $\mathbf{V}$ ,

$$\mathbf{V} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$

で表されるとき,  $\rho_{ij}$  を行列  $U$  の  $i$  列と  $j$  列の相関係数,  $\rho_{V_j}$  は行列  $U$  の  $j$  列とターゲットベクトル  $V$  との相関係数とする. ここで特徴選択をするために, 0,1 の値をとる変数  $x_j (j = 1, \dots, n)$  を要素としてもつベクトル  $\mathbf{X}$  を定義する.

$$x_j = \begin{cases} 1 & j \text{ 列を特徴として選択する} \\ 0 & \text{上記以外} \end{cases} \quad (2.4)$$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

次に, ターゲットベクトルへの影響 (**influence**) と, 特徴同士の独立性 (**independence**) を考えるために以下の二つの要素を導入する.

$$\sum_{j=1}^n x_j |\rho_{V_j}| \quad (2.5)$$

$$\sum_{j=1}^n \sum_{k=1, k \neq j}^n x_j x_k |\rho_{jk}| \quad (2.6)$$

(2.5) 式は影響 (**influence**) を表しており, ターゲットベクトルとの相関係数が高い特徴が大きな寄与をする関数である. そして, 多くの特徴を使用する場合, この項の値は大きくなる. 一方で, (2.6) 式は独立性 (**independence**) を表しており, 相関の強い特徴同士の項が大きな寄与をする. つまり, この項が大きくなると相関の強い特徴が残るため, 特徴の独立性が失われることとなる. つまり, 理想の特徴選択は (2.5) をできるだけ大きくし, (2.6) を小さくすることにより実行できる. 量子アニーラを用いて特徴選択を行うために, ハイパーパラメータ  $\alpha (0 \leq \alpha \leq 1)$  を導入し, (2.5) 式と (2.6) を用いて最小化する目的関数を次のように定義する.

$$f(\mathbf{x}) = - \left[ \alpha \sum_{j=1}^n x_j |\rho_{V_j}| - (1 - \alpha) \sum_{j=1}^n \sum_{k=1, k \neq j}^n x_j x_k |\rho_{jk}| \right] \quad (2.7)$$

ここで  $x_j$  は  $0, 1$  の値をとる変数なので  $x_j x_j = x_j$  が成り立つ. よって (2.7) 式は以下のように変形できる.

$$f(\mathbf{x}) = -\mathbf{x}^T Q \mathbf{x} \quad (2.8)$$

パラメータ  $Q$

$$Q = \begin{pmatrix} \alpha|\rho V_1| & \cdots & (1-\alpha)|\rho_{1n}| \\ & \ddots & \vdots \\ 0 & & \alpha|\rho V_n| \end{pmatrix} \quad (2.9)$$

そして (2.8) 式を最小化するような  $\mathbf{x}^*$  を量子アニーラを用いて求める. つまり,  $\mathbf{x}^*$  は以下の式で求められる.

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left[ -\mathbf{x}^T Q \mathbf{x} \right] \quad (2.10)$$

量子アニーラは (2.10) 式中の  $Q$  をパラメータとして受け取ることで,  $\mathbf{x}^*$  の近似解を返す.

## 第3章

# 量子アニーリングを用いた特徴選択手法

本章では Factorization Machine と量子アニーリングを組合せた FMQA[4] とその拡張方法であるビット制約について説明し, FMQA を拡張した FMQA+を導入する. そして FMQA+ を機械学習モデルにおける特徴選択に用いる方法を述べる.

### 3.1 FMQA

FMQA は Factorization Machine(FM) と量子アニーリングを用いてブラックボックス関数の最適化を行うためのライブラリである [4]. FMQA では D-Wave 社が提供するクラウド経由で使用可能な量子アニーラ D-Wave2000Q を用いている. D-Wave2000Q は (2.1) 式で表されるイジング模型を変形して得られる QUBO(Quadratic Unconstrained Binary Optimization) 形式

$$H(\mathbf{q}) = \sum_{i < j} Q_{ij} q_i q_j + \sum_{i=1}^N Q_{ii} q_i, \quad q_i \in \{0, 1\}$$

のエネルギー  $H$  が最小となる組合せ  $\mathbf{q}$  の近似解を求めることもできる. FM はこの QUBO パラメータ  $Q_{ij}$  を学習する為に用いられる. つまり FMQA でブラックボックス関数最適化を行う場合は, まず最適化したい関数/解きたい問題の特徴が 0 または 1 の 2 値をとるような問題に落とし込み, 以下の手順を繰り返す.

1. データセット  $\{X, \mathbf{Y}\}$  に対して (2.1) 式を用いて学習を行う.
2. 学習で得られた  $w_i, \mathbf{v}_i \cdot \mathbf{v}_j$  を QUBO((3.1) 式) のパラメータ  $Q$  に変換する.

3. 得られたパラメータ  $Q$  を量子アニーラ (D-Wave2000Q) へ渡し, エネルギーが最小となる  $\mathbf{x}$  を計算する.

4. 得られた組み合わせ  $\mathbf{x}$  を用いて  $y_i$  を得ることで, データセット  $\{X, Y\}$  を更新する.

これらを繰り返すことによって, 最適化が行われ, 目的関数を最小にするような組み合わせ  $\mathbf{x}$  を得ることが期待できる. また, FM を用いると QUBO の  $N \times N$  の相互作用行列を学習する代わりに,  $N$  個の  $K$  次元ベクトルを学習すれば良いので, パラメータ次元を  $K \times N$  まで落とすことができ, 学習が早く進み, 過学習を防ぐことができる.

### 3.2 ビット制約

QUBO における変数  $q$  は  $\{0, 1\}$  を要素に持つため, 個々の成分  $q_i$  を特にビットという. QUBO 形式の問題の解に 1 が入るビット数を  $n$  にしたい場合,

$$\left( \sum_i q_i - n \right)^2 \quad (3.1)$$

を最小化することと等しいため, ペナルティ項として係数  $\lambda$  を乗じて元の問題に加えた以下の問題を解けば良い.

$$H(\mathbf{q}) = \sum_{i < j} Q_{ij} q_i q_j + \sum_{i=1}^N Q_{ii} q_i + \lambda \left( \sum_i q_i - n \right)^2 \quad (3.2)$$

ここで  $\lambda$  をペナルティ強度と呼ぶ. D-Wave2000Q のグラフ構造はキメラグラフと呼ばれ, 全ビットが結合していないため, 複数のビットを 1 つのビットとみなすことで, 全結合グラフを構築する. この 1 つのビットを表すための複数のビット列はチェーンと呼ばれ, チェーン内のビットは全て同じ値をとることが望ましく, それを保証するためにチェーン内の相互作用には一定の大きさの強磁性相互作用 (各ビットが同じ値となる際にエネルギーが小さくなる)  $\gamma$  が付加される. これをチェーン強度と呼ぶ. つまり全体としては,

$$H(\mathbf{q}) = \sum_{i < j} Q_{ij} q_i q_j + \sum_{i=1}^N Q_{ii} q_i + \lambda \left( \sum_i q_i - n \right)^2 - \gamma \sum_{All} \sum_{chain} \sum_l \sum_{i=1}^{len(l)-1} q_{l,i} q_{l,i+1} \quad (3.3)$$

を解くことになる。よってビット制約をかけた問題を D-Wave マシン上で解く際には、ペナルティ項とチェーンの強磁性相互作用の2つの制約について考慮する必要がある。つまり、お互いが強くなり過ぎないように適切な  $\lambda$  と  $\gamma$  を設定する必要がある。

### 3.3 FMQA+を用いた特徴選択アルゴリズム

FMQA で、3.2 節のビット制約の方法を用いて、得られる解  $q$  に 1 が入るビット数を制限するようにしたものを **FMQA+** とする。機械学習モデルにおいてデータのどの特徴を用いれば精度が向上するかという問題は組合せ最適化問題に帰着できるため、FMQA+で最適化を行うことができる。具体的な方法を説明するために、以下を仮定する。

- データセット  $\{U, V\}$   
 $(U = \{u_{ij} : i = 1, \dots, m, j = 1 \dots n\}, V = \{v_i : i = 1, \dots, m\})$
- ベクトル  $\mathbf{X} = \{x_j : j = 1, \dots, n\}$ ,  $x_j = \begin{cases} 1 & j \text{ 列を特徴として用いる} \\ 0 & \text{上記以外} \end{cases}$
- 入力として  $\mathbf{u}$  を受け取り、 $v$  の予測値  $\hat{v}$  を出力する機械学習モデル  $M$

FMQA+を用いた特徴選択アルゴリズムの擬似コードを Algorithm1 に示す。このアルゴリズムは3つの手続き `Some_Costly_Function`, `FM_to_QUBO`, `FMQA_Sampling` からなる。`Some_Costly_Function` では QA によって選ばれた特徴を用いて機械学習モデル  $M$  における予測誤差/精度を評価関数  $S$  により測定している。`FM_to_QUBO` では、ある特徴を選ぶ/選ばないを決定するベクトル  $\mathbf{X}$  と `Some_Costly_Function`( $\mathbf{X}$ ) によって得られた予測精度/誤差をデータとして、Facotrization Machine の学習を行い、D-Wave2000Q でサンプリングするために必要なパラメータ  $\mathbf{Q}$  を返す。`FMQA_Sampling` は上記二つの手続きを用いて、実際に最適化を行なっていく手続きである。最初にランダムな  $\mathbf{X}$  において評価値を求めたデータ  $\{\mathbf{X}, \mathbf{y}\}$  を作り、`FM_to_QUBO` で学習し、得られた QUBO パラメータ  $\mathbf{Q}$  を用いて D-Wave2000Q によるサンプリングを行う。そして得られた新しい特徴組み合わせ候補  $x'$  とその評価値  $y'$  を既存の  $\{\mathbf{X}, \mathbf{y}\}$  に追加する。この一連の流れを **1 サンプルング** と呼び、これらを `NUM_SMPLING` で指定された回数繰り返すことによって、より良い特徴を探索する。FMQA+による特徴選択は Wrapper 法に分類されるので、機械学習モデル  $M$  や

評価関数  $S$  は問題に合わせて自由に変更することができる。また FMQA\_SAMPLING 中の D-Wave2000Q によるサンプリングは, 古典アルゴリズムのシミュレーテッドアニーリングや, Tabu サーチを用いた QBSolv というサンプラーに変えることもできる。

---

**Algorithm 1** FMQA を用いた特徴選択アルゴリズム

---

**Input:** データセット  $\{U, V\}$ , 機械学習モデル  $M$ , 評価関数  $S$ , 制約ビット数  $k$

**Output:** 0,1 の組み合わせ  $X$

```

1: procedure SOMECOSTLYFUNCTION( $X$ )
2:    $mask = []$ 
3:   for  $i = 1, \dots, n$  do
4:     if  $x_i == 1$  then  $i$  を  $mask$  に加える end if
5:   end for
6:    $U' \leftarrow U$  から  $mask$  の要素に対応する列のみを抜き出す
7:    $y \leftarrow \{U', V\}$  をモデル  $M$  用いて学習し, 評価関数  $S$  で交差検証した値
8:   return  $y$ 
9: end procedure
10: procedure FM_TO_QUBO( $X, y$ )
11:    $\{X, y\}$  で Factorization Machine を学習
12:    $Q \leftarrow$  学習した  $FM$  から重みを取り出す
13:   return  $Q$ 
14: end procedure
15: procedure FMQA_SAMPLING(NUM_SAMPLING)
16:   Initialize:
17:      $INIT\_NUM = 10$ 
18:      $X := 1$  の数が  $k$  個でランダムに 0,1 の値をとる,  $INIT\_NUM \times n$  の行列
19:      $y := \text{SomeCostlyFunction}(X)$ 
20:   for  $i = 1, \dots, NUM\_SAMPLING$  do
21:      $Q = \text{FM\_to\_QUBO}(X, y)$ 
22:      $X' \leftarrow D - \text{Wave2000Q}$  に  $Q$  渡して得られた解
23:      $y' = \text{SomeCostlyFunction}(X')$ 
24:     データセット  $\{X, y\}$  に  $x', y'$  を追加
25:   end for
26: end procedure

```

---

## 第4章

### 実験

本章では FMQA+ を用いて機械学習モデルにおける特徴選択を行い, その最適化性能や精度を評価し, 既存の特徴選択手法と比較する.

#### 4.1 実験環境

実験用のデータには, 膜型表面応力センサー (MSS: Membrane-type Surface stress Sensor[9]) を用いて得られたニオイデータと National Renewable Energy Laboratory (NREL, <https://materials.nrel.gov>) の提供するバンドギャップデータ, UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>) の提供する German Credit Data Set を用いた. 各データの詳細を表 4.1 と表 4.2 に示し, 各データについては 4.1.1 と 4.1.2, 4.1.3 で詳しく述べる. 実験用のプログラムは Python3.7.0 を用いて実装し, Intel Xeon Cpu e5-2690 v4 @ 2.60GHz, 264GB RAM の計算機上で実行した. 量子アニーラとして D-Wave 社の D-Wave2000Q をオープンソースの SDK である Ocean(<https://ocean.dwavesys.com/>) を介して利用している.

##### 4.1.1 ニオイデータ

ニオイデータは国立研究開発法人物質・材料研究機構 (NIMS) の開発した膜型表面応力センサー (MSS: Membrane-type Surface stress Sensor[9]) によって得られたデータである. MSS は中央部に塗布された感応膜に, ニオイのガス分子が吸着することによって生じる力で歪みが発生する. その歪みを, 力が加わると電気抵抗が変化する検知部で電氣的に検知することによって応答波形を得ている. ニオイデータは 12 種類の様々な物質を感応膜



に塗布したセンサー それぞれの応答波形において、傾きの変わる 4 つの部分の数値データとして取り出した、特徴次元 48 のデータとなっている。ターゲットクラスはもともと 88 種類あったニオイを果物や野菜、鉱物など 12 クラスに再度分類したものとなっている。

#### 4.1.2 German Credit Dataset

UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>) の提供する German Credit Data Set は、顧客の様々な情報(特徴)と、その顧客の信用度が組になったデータセットである。具体的には、特徴量として、顧客の信用貸金額などの 7 つの数値データ、性別や購入品などの 13 のカテゴリカルなデータを含み、顧客の信用を "Good=1" と "Bad=2" で表現したデータセットである。本論文では、扱いやすいように、カテゴリカルなデータを One-Hot 表現にし、特徴次元 48、データ数 1000 のデータとしている。

#### 4.1.3 バンドギャップデータ

National Renewable Energy Laboratory (NREL, <https://materials.nrel.gov>) は 2142 の酸化物に対して Perdue Burke Ernzerhof (PBE) という手法を用いて計算されたバンドギャップのデータを提供している [7]。これらの酸化物を数値的に表すために、材料組成から記述される特徴量 (132 次元) を Elementproperty Featurizer of Matminer[8] を用いて計算する。実験には特徴次元 132 のバンドギャップデータを D-Wave2000Q で扱えるように、特徴次元 53 に削減したデータを用いる。

表 4.1: 分類問題のデータセット

データセット	特徴次元	ターゲットクラス	データ数
ニオイデータ	48	12	440
German Credit	48	2	1000

表 4.2: 回帰問題のデータセット

データセット	特徴次元	データ数
バンドギャップデータ	53	2142

## 4.2 実験結果

以下に実験結果を示す.

### 4.2.1 結果 1: FMQA の最適化性能

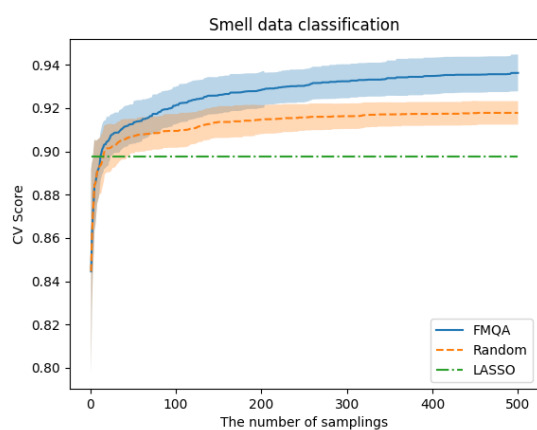
ビット制約をかけない FMQA でサンプリングを繰り返して特徴選択を行なった場合, 予測精度/誤差が最適化されていくか実験をし, 他の手法と比較した. ニオイデータと German Credit Dataset では, FMQA と LASSO, ランダムサンプリングと比較し, バンドギャップデータでは, FMQA と RFECV, ランダムサンプリングを比較した. ニオイデータと German Credit Dataset では機械学習モデルとして scikit-learn の LogisticRegression を, バンドギャップデータ (2) では scikit-learn の RandomForestRegressor を用いた. 500 回サンプリングしたものの中で最も予測精度/誤差の良かったものを比較対象とし, 全体で 50 回実験を行い, その平均を結果として用いた. 予測精度/誤差を測定する際は, データに対し 5 分割交差検証を行なって得られた交差検証誤差/精度を用いた. 結果を図 4.1a, 4.1b, 4.1c に示す. 最も予測精度/誤差が良かった際の特徴数を表 4.3 にまとめる. この実験では, 初期データ数は 10 とし, ランダムに選択した. 結果より, FMQA でサンプリングを繰り返すことにより, より良い特徴を探索できていることが分かる.

表 4.3: FMQA で 500 回サンプリングし, その中で最も良かった予測精度/誤差

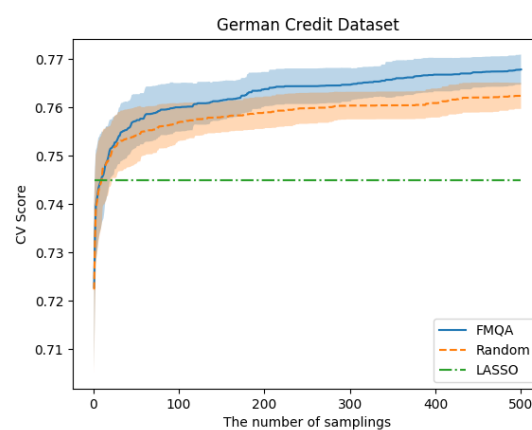
データ	とその時の特徴数 (50 回平均)		
	予測精度/誤差	データの特徴数	選択された特徴数
ニオイデータ	0.936	48	25.7
German Credit	0.768	48	22.4
バンドギャップデータ	0.701	53	22.1

### 4.2.2 結果 2: ビット制約におけるペナルティ強度とチェーン強度のチューニング

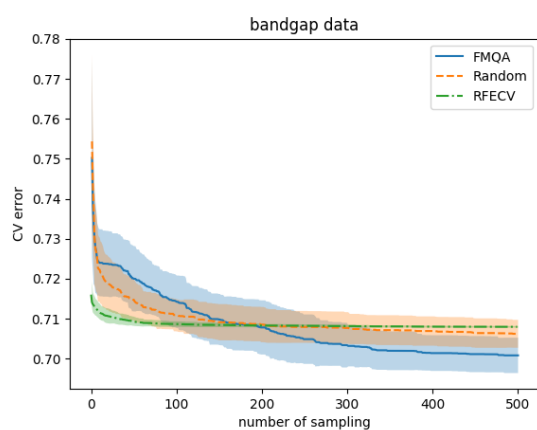
特徴選択において選ばれる特徴数を  $n$  にしたい場合, 3.2 節で述べた方法を用いて, D-Wave マシンでサンプリングされる解におけるビット数に制約をかけることができる. その際に用いるハイパーパラメータであるペナルティ強度  $\lambda$  とチェーン強度  $\gamma$  を様々に変化させた場合の結果を示す. ここではニオイデータのみに対し実験を行なった. ただし,



(a) ニオイデータ



(b) German Credit Dataset



(c) バンドギャップデータ

図 4.1: 最適化性能の比較 (50 回平均)

48 次元ある特徴を先頭 16 次元に絞り特徴選択を行った.  $\mathbf{X}$  において  $k$  個の 1 が出現するようにランダムに生成した 10 個のビット列で Factorizaion Machine を訓練し, 得られた QUBO パラメータのもと,  $\lambda$  と  $\gamma$  を 0.01 から 10 まで変化させ, 100 回サンプリングした際に制約を満たす解が得られた回数とチェーンの壊れた割合を記録した. チェーンが壊れるというのは, チェーンに含まれる複数のビット列が同じ値になっていない場合を指す.  $k = 2, 4, 8$  とした時にチェーンの壊れた割合を図 4.2a, 4.2b, 4.2c に, 制約を満たす解が得られた回数を図 4.3a, 4.3b, 4.3c に示す. 図 4.2 より,  $\gamma$  が  $\lambda$  より大きい値をとる範囲ではチェーンはほとんど壊れていないことが分かる. また, 図 4.3 より  $\frac{\lambda}{\gamma} = 0.215$  の比率を維持すると制約を満たす解が得やすいことが分かる.

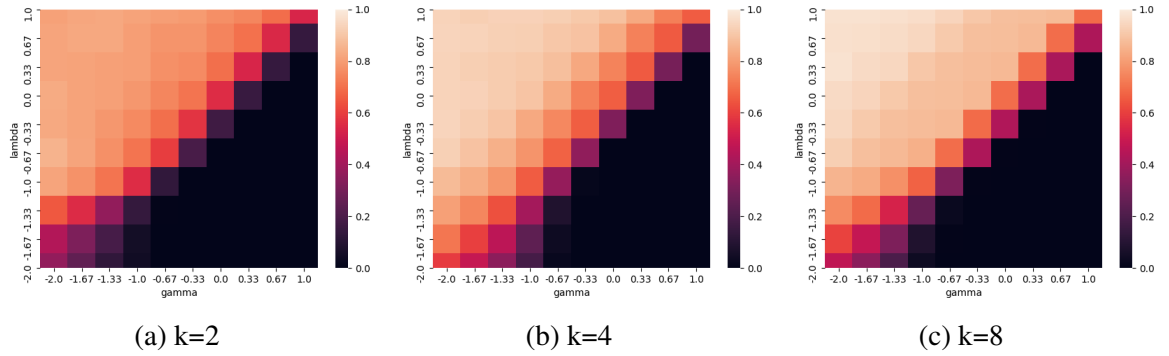


図 4.2: チェーンの壊れた割合 (縦横軸は log10 スケール)

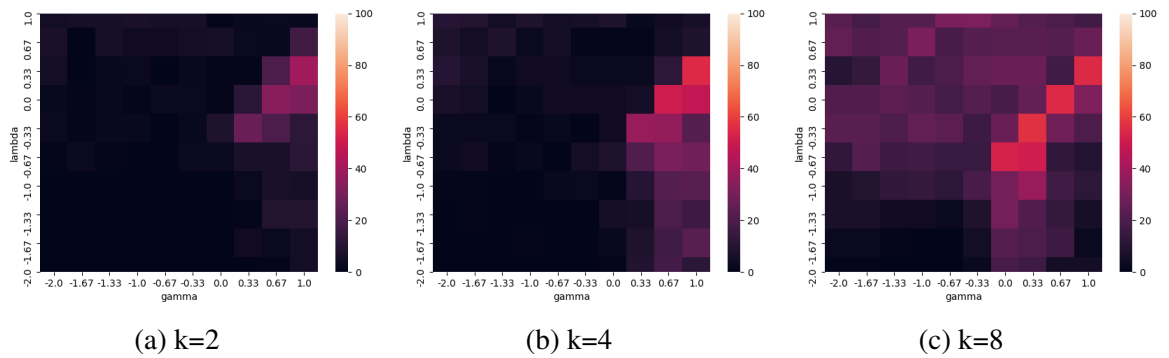


図 4.3: 制約を満たす解が得られた回数 (縦横軸は log10 スケール)

#### 4.2.3 結果 3: ビット制約をかけた時の性能比較

実験 1 の結果から FMQA+において  $\lambda = 0.21, \gamma = 1.0$  のもと, ビット制約をかけ, ニオイデータと German Credit Dataset における, 選択された特徴数毎の予測精度を LASSO

と 1QBit 社の手法, 全探索と比較した. 比較対象の LASSO は正則化項にスパースな解を得やすい L1 正則化項を用いたもので, scikit-learn の LogisticRegression を用いる. またバンドギャップデータにおける選択された特徴選択毎の予測誤差を 1QBit 社の手法と全探索, scikit-learn の RFE という Wrapper 法に分類される手法と比較した. 具体的には, 再帰的に特徴を削除してその都度, 特徴量の重要度を測り, 指定した特徴数になるまで処理を繰り返す方法である. バンドギャップデータにおける機械学習モデルには scikit-learn の RandomForestRegressor を用いた. ここで, LASSO の正則化パラメータ  $C$  を様々に変化させた場合に選ばれる特徴数と予測精度の関係を図 4.4 に示す. この結果より, 正則化パラメータをニオイデータでは 0.01 から 60, German Credit Dataset では 0.005 から 15 の範囲で 100 個決め打ちし, そのときの特徴数と予測精度を比較対象とした. また 1Qbit 社の手法におけるハイパーパラメータ  $\alpha$  を変化させた時の結果を図 4.5 に示す. これらの図より,  $\alpha$  の値が大きくなるほど, 選択される特徴数が増加することがわかる. 予測精度を測定する際は, 学習データに対して, 5 分割交差検証により得られた交差検証精度を用いた. ニオイデータと German Credit Dataset における選択された特徴数毎の予測精度結果を図 4.6, 4.7 に示す. またバンドギャップデータにおける選択された特徴選択毎の予測誤差を図 4.8 に示す. 結果より, ニオイデータと German Credit Dataset におけるロジスティック回帰問題で, 各特徴数において FMQA+ は LASSO や 1Qbit 社の手法よりも高い予測精度を達成していることが分かる. また, バンドギャップデータにおける回帰問題でも, 各特徴数において FMQA+ は 1QBit 社の手法や RFE よりも良い予測誤差を達成していることが分かる.

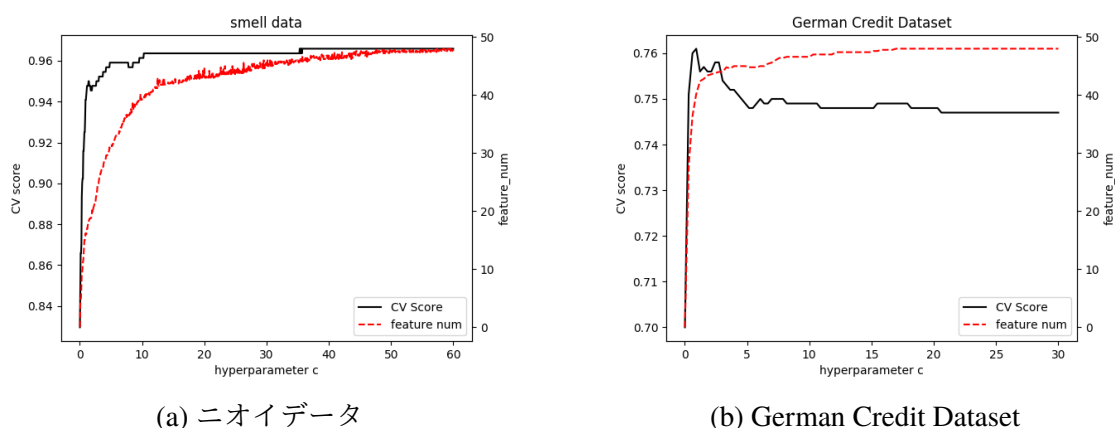
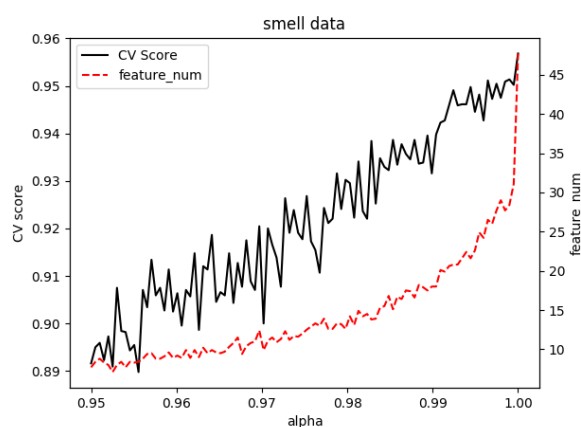
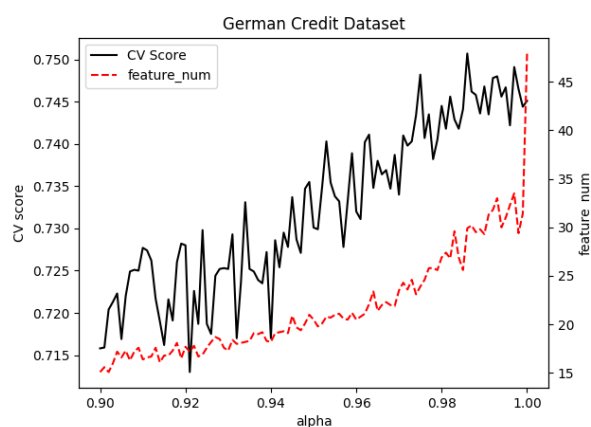


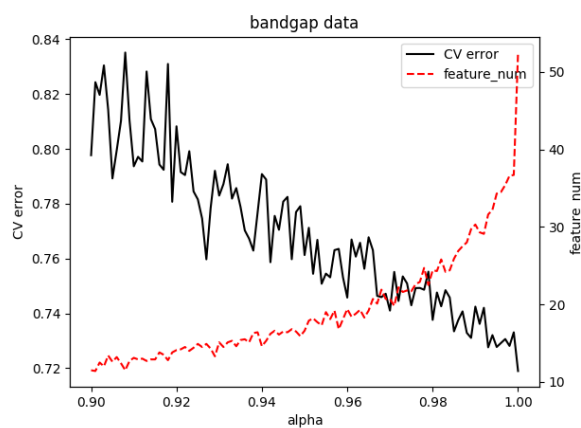
図 4.4: 正則化パラメータを変化させた場合の特徴数と予測精度の関係



(a) ニオイデータ



(b) German Credit Dataset



(c) バンドギャップデータ

図 4.5: 1QBit 社の手法における  $\alpha$  による予測精度/誤差と特徴数の関係

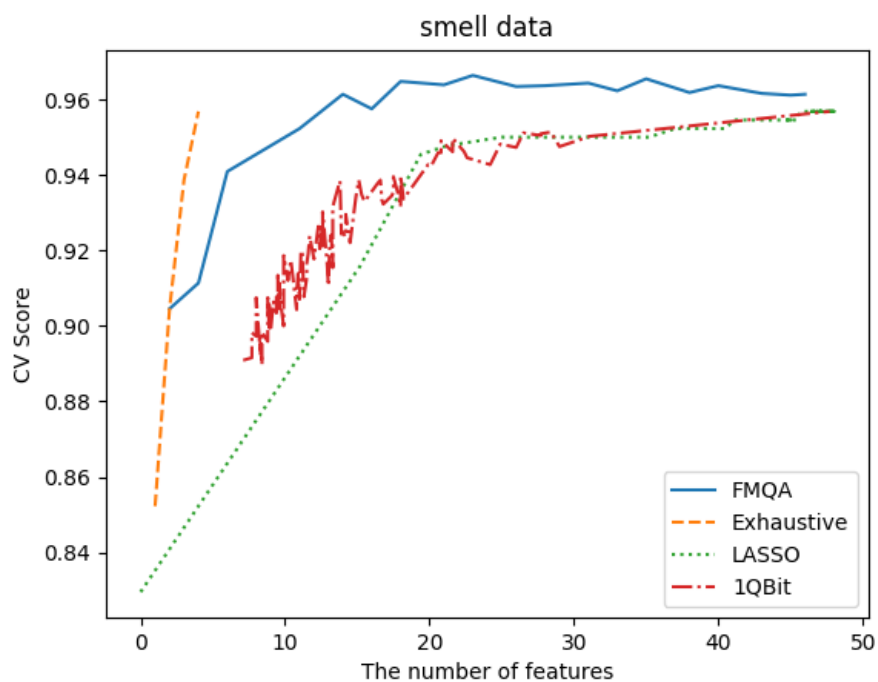


図 4.6: ニオイデータにおける特徴数毎の予測精度 (50 回平均)

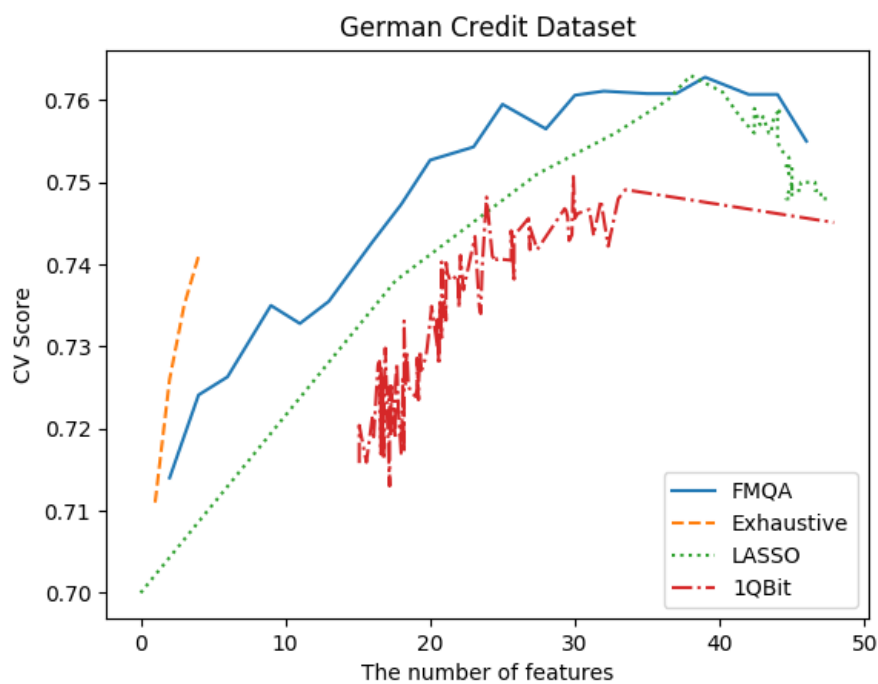


図 4.7: German Credeit Dataset における特徴数毎の予測精度 (50 回平均)

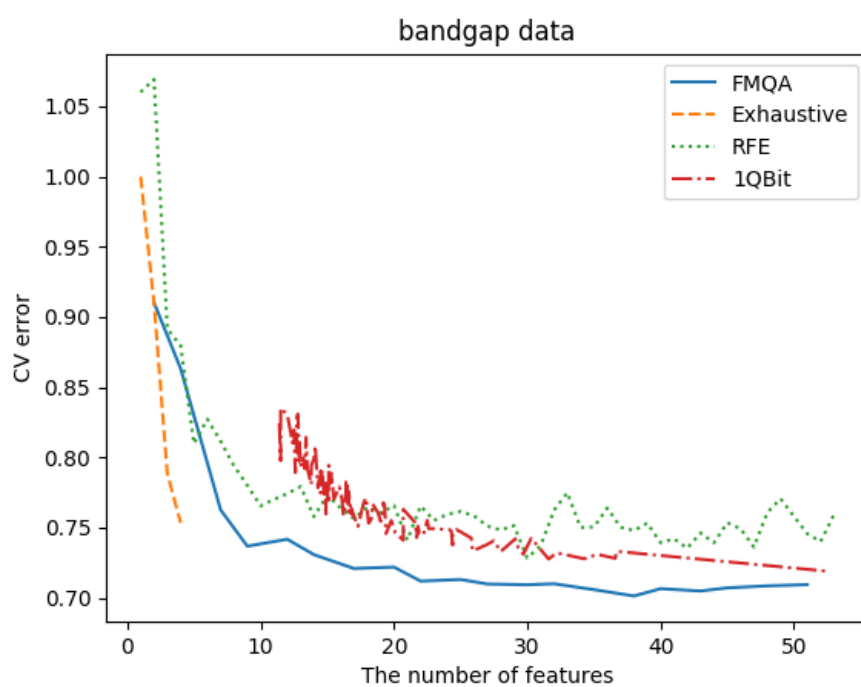


図 4.8: バンドギャップデータにおける特徴数毎の予測誤差 (10 回平均)



## 第5章

### おわりに

本論文では、任意の機械学習モデルの交差検証誤差を Factorization Machine によってモデルリングし、特徴数を指定する罰則項を加えることで、特徴選択問題をイジングモデルに帰着させる方法を提案した。その具体的な方法として第3章では、量子アニーリングと Factorization Machine を組合せた FMQA に対し、ビット制約を加えた FMQA+を導入し、FMQA+を用いた特徴選択アルゴリズムを提案した。第4章では、FMQA で特徴のサンプリングを繰り返すことによって、機械学習モデルにおける汎化性能が向上するような特徴を探索できていることを示した。また、ビット制約におけるペナルティ強度とチェーン強度のチューニングを行い、D-Wave マシンでサンプリングした解における1の数が目的の数だけ出現するか確かめた。目的の数と同じ数だけ1が含まれるような解を得られる割合は高くはないものの、実際にビット制約を行うことができた。そして実際に、量子アニーラの実機である D-Wave2000Q を用いて特徴選択を行い、特徴選択の近似手法である LASSO や scikit-learn で提供されている Wrapper 法の RFE, 1QBit 社の提案する量子アニーラを用いた特徴選択手法と比較した結果を示した。ロジスティック回帰と、線形回帰に対して提案法を適用した結果、LASSO や RFE, 1Qbit 社による手法よりも、同数の特徴量で、優れた汎化誤差を達成できることを確かめた。今後の課題としては、ビット制約における成功の割合を増やすための手法を考えることが挙げられる。また現在は D-Wave の制限のため、本手法は、64 次元以内の問題にしか適用できないが、今後、量子アニーラの規模が増大すればより大きな問題にも適用できると考えられるので、生命科学分野における大規模な実データへの応用などが挙げられる。

## 謝辞

本研究を進めるにあたり、熱心な指導をしていただいた東京大学大学院新領域創成科学研究科メディカル情報生命専攻の津田宏治教授に心より感謝します。津田先生には、研究の進め方だけでなく心構えまで学ばせていただきました。また、お忙しい中、日頃のミーティングで研究内容に関して頻繁に議論していただいたことを重ねて感謝いたします。講師の田村亮さんには、量子アニーリングに関する貴重なコメントや論文に関するアドバイスをいただきました。本論文が完成したのは先生方のおかげです。心より感謝します。研究室の先輩方、特に博士課程2年の北井孝紀さんにはFMQAをはじめ、量子アニーリングやD-Waveマシンの使い方など多くのことを教えていただきました。本研究に関するだけでなく、プログラミングのテクニックなど多くのことを学ばせていただきました。深く感謝いたします。また、津田研究室の澁谷薫さんには、出張手配から書類提出まで幅広くお世話になりました。ありがとうございました。そして、日々の研究室生活において、多くの意見を交わし、励ましあった同期に感謝いたします。

最後になりますが、大学院に入学する機会を与えてくださり、ここまで支えてくれた両親に深く感謝します。

## 参考文献

- [1] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.
- [2] Mark W Johnson, Mohammad HS Amin, Suzanne Gildert, Trevor Lanting, Firas Hamze, Neil Dickson, R Harris, Andrew J Berkley, Jan Johansson, Paul Bunyk, et al. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, 2011.
- [3] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355, 1998.
- [4] Koki Kitai, Jiang Guo, Shenghong Ju, Shu Tanaka, Koji Tsuda, Junichiro Shiomi, and Ryo Tamura. Expanding the horizon of automated metamaterials discovery via quantum annealing. *arXiv preprint arXiv:1902.06573*, 2019.
- [5] Andrew Milne, Maxwell Rounds, and Phil Goddard. Optimal feature selection in credit scoring and classification using a quantum annealer. 2017.
- [6] Steffen Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010.
- [7] Vladan Stevanović, Stephan Lany, Xiuwen Zhang, and Alex Zunger. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Physical Review B*, 85(11):115104, 2012.
- [8] Logan Ward, Alexander Dunn, Alireza Faghaninia, Nils ER Zimmermann, Saurabh Bajaj, Qi Wang, Joseph Montoya, Jiming Chen, Kyle Bystrom, Maxwell Dylla, et al. Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, 152:60–69, 2018.

- [9] Genki Yoshikawa, Terunobu Akiyama, Sebastian Gautsch, Peter Vettiger, and Heinrich Rohrer. Nanomechanical membrane-type surface stress sensor. *Nano letters*, 11(3):1044–1048, 2011.