Encoding Classical Information

Basis Encoding

$x=3 \xrightarrow{\text{binary}} 11$

$\downarrow$ quantum state

$|11\rangle$

$x=\begin{bmatrix}3\\2\end{bmatrix} \rightarrow \begin{bmatrix}11\\10\end{bmatrix}$

$\downarrow$

$|1110\rangle$

Advantage: Ease of preparation

$$\left.\begin{matrix}|0\rangle - \boxed{X} - \\ |0\rangle - \boxed{X} - \end{matrix}\right\} |11\rangle$$

Disadvantage: Qubit count.

Amblitude Encoding

$x=\begin{bmatrix}x_0\\x_1\end{bmatrix}$ $\|x\|_2=1$

$|x\rangle = x_0|0\rangle + x_1|1\rangle$

Advantage: fewer qubit

Disadvantage: Preparation & readout

Hamiltonian Encoding

① The Ising model

Map the problem to the Ising model.

$$H=-\sum_{\langle ij\rangle} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i$$

e.g. Quantum Anealing.

Advantage: easy to implement

Disadvantage: Limited Scope

② Hamiltonian simulation

Simulate $H=e^{-iHt}$

e.g. QAOA, quantum matrix inversion

Advantage: natural to encode matrices.

Disadvantage: T&C apply

Ensemble Learning

Optimization in M.L.

$S = \{(x_i, y_i)\}_{i=1}^N \quad x_i \in \mathbb{R}^d$

$h(\theta; x): \mathbb{R}^d \to \{0;1\}$ ⇒ discriminitive & supervised

$p(y|x)$ $\quad \to \{0; \dots; c-1\}$

$\min_\theta L(\theta, S), \quad \theta \in \mathbb{R}^p$ → each is a 32-bit/64-bit floating point number
↘ Deep learning: p is high

Ensembles

$h_1(\theta_1; x) \dots h_K(\theta_K; x)$ → combine many neural networks together.

$F_K(w; x) = \sum_{\ell=1}^{K} w_\ell h_\ell(\theta_\ell, x)$

AdaBoost

Sequential expansion of the ensemble $\quad m \le K$: $F_m(w; x) = F_{m-1}(w; x) + w_m h_m(\theta_m; x)$

w_m is derived from the expansion loss:

$$\sum_{i=1}^{N} e^{-y_i F_m(x_i)}$$

→ Regularization is absent.

Modern variants:

- Xgboost
- Gradient based trees

Qboost

An ensemble with a different loss function.

$S = \{(x_i, y_i)\}_{i=1}^N \quad \{h_\ell(x)\}_{\ell=1}^K$

$\min_w \left[\frac{1}{N} \sum_{i=1}^{N} \left(\sum_{\ell=1}^{K} w_\ell h_\ell(x_i) - y_i\right)^2 + \lambda \|w\|_0\right]$

Zero norm of w. measures how many elements are non-zero.

$w \in \mathbb{R}^F$ not discrete;

$\|w\|_0 = \sqrt[0]{\sum_i w_i^0}$

Hamiltonian Encoding: Ising

constant

$\min_w \left[\frac{1}{N} \sum_{i=1}^{N} \left[\left(\sum_{\ell=1}^{K} w_\ell h_\ell(x_i)\right)^2 - 2 \sum_{\ell=1}^{K} w_\ell h_\ell(x) y_i + \cancel{y_i^2}\right] + \lambda \|w\|_0\right]$

$= \min_w \left[\frac{1}{N} \sum_{\ell=1}^{K} w_{\ell,1} \left(\sum_{i=1}^{N} h_\ell(x_i) h_1(x_i)\right) - \frac{2}{N} \sum_{\ell=1}^{K} w_K \sum_{i=1}^{N} h_\ell(x_i) y_i + \lambda \|w\|_0\right]$

Clustering By Quantum Optimization

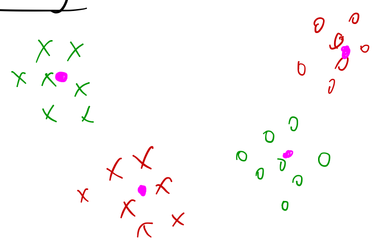
Unsupervised Learning.

$\{x_i\}_{i=1}^{N} \quad x_i \in \mathbb{R}^d$

$p(y|x) \rightarrow$ Discrimitive

$p(x) \rightarrow$ Generative

Clustering



Kmeans:

- central points

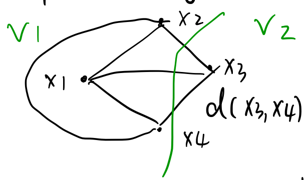
density-based clustering

Mapping the Ising model

Calculate the Gram matrix.

$K_{ij} = d(x_i, x_j)$

Define a weighted graph.



$\sigma_i = -1, \ x_i \in V_1$

$\sigma_i = +1, \ x_i \in V_2$

Max-cut: NP-hard.

Coupling strength

spin variables.

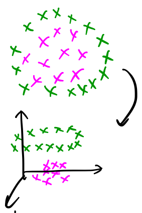
Cost of cut: $\sum_{i \in V_1, j \in V_2} \overbrace{d(x_i, x_j)}^{w_{ij}} = \frac{1}{4} \sum w_{ij} - \frac{1}{4} \sum_{ij} w_{ij} \sigma_i \sigma_j$ ⇒ Ising model

constant offset

Kernel Methods.

Non-linear Embedding

$S = \{(x_i, y_i)\}_{i=1}^{N}$



higher dimension

Kernel Function

Gram / Kernel matrix.

$k_{ij} = x_i^T x_j \Rightarrow$ inner product of the original space

$k_{ij} = \underbrace{\phi(x_i)^T \phi(x_j)}_{k(x_i, x_j)} \Rightarrow$ " " Embedded space

You don't actually need ϕ!

e.g. kenerlized k-means

SVM

An Interference Circuit.

Start from Hardware

(QC)

① Think of what a Quantum Computer can do

② Create a new learning algorithm

State preparation in amplitude Encoding

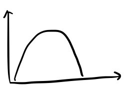
$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \rightarrow |x\rangle = x_0|0\rangle + x_1|1\rangle$

$\|x\|_2 = 1$

Kernel as Interference

Calculate $1 - \frac{1}{4N} \sum_{i=0}^{N-1} \|\tilde{x} - x_i\|_2^2 \Rightarrow$ Easy for a gate model QC.

$\Downarrow$



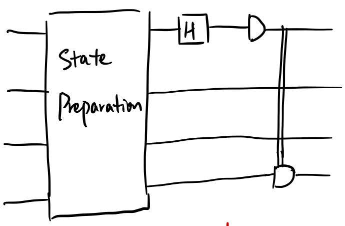
Circuit:

Be entangled with test instance $\tilde{x}$ → Ancilla qubit

keep track of the index → Index qubit

data → Data qubit

label → Class qubit



⇒ If we have 3 data points, a single-qubit index is not enough.

Prepare:

test instance

data instance

$|\psi\rangle = \frac{1}{C} \sum_{i=0}^{N-1} \left(|0\rangle|i\rangle|\tilde{x}\rangle + |1\rangle|i\rangle|x_i\rangle \right) |y_i\rangle$

$\Downarrow$ Hadamard

Hadamard on ancilla:

$\frac{1}{C} \sum_{i=0}^{N-1} \left(\boxed{|0\rangle|i\rangle|\tilde{x} + x_i\rangle} + |1\rangle|i\rangle|\tilde{x} - x_i\rangle \right) |y_i\rangle$

$\downarrow$

measure ancilla qubit ⇒ if successful case ⇒ measure the class qubit

Probabilistic Graphic models

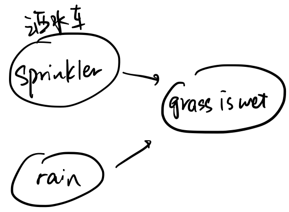
Markov Network

Discriminative models: $P(y|x)$ → Deep Learning

Generative models: $p(x), p(x,y)$ → The output of the learner has same dimension with input space, making it hard to create a concise representation

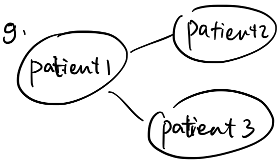
Probabilistic Graphic models: sparse modeling of $p(x)$ or $p(x,y)$

↘ Directed models: Bayesian Network e.g.



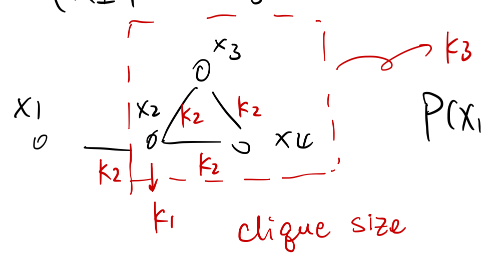
↓ Undirected models: Markov Network

e.g.



Factoring a joint distribution

$(X \perp Y \mid Z)$ if $P(X=x, Y=y \mid Z=z) = P(X=x \mid Z=z) P(Y=y \mid Z=z) \quad \forall x, y, z$



$$P(x_1 \dots x_4) = \frac{1}{Z} \exp\left(-\sum_k E(C_k)\right)$$

Markov Network ⟷ Boltzmann Dist.

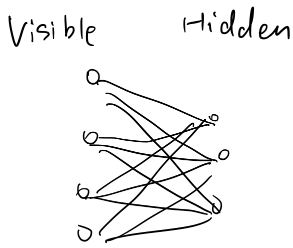
Speeitial Cases

① Ising Model

② Bolzmann machines

$p(v,h) = \frac{1}{Z} \exp(-E(h,v))$

$p(v) = \sum_h P(v,h)$

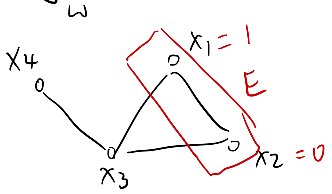


Optimization and Sampling in PGMs

Probalistic Queries

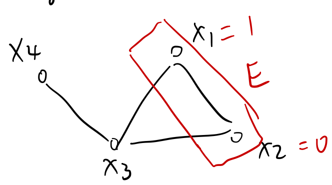
① Most probable explanation

$\arg\max_w P(w, E=e)$



② Maximize a posterior

$\arg\max_r \sum_z P(r, z \mid E=e)$



Only interested in the subset.

Approximate Inference

Sampling on a digital computer Markov Chain Monte Carlo

$\rightarrow P(X_1 \dots X_m) = \frac{1}{Z} \exp\left(-\sum_{\xi_\ell} E(\xi_\ell)\right)$

Quantum-Enhanced inference

① Set Ising model

② Run Ⓐ Quantum annealing

→ More qubits

→ Estimate effective temperature.

Ⓑ Quantum Approximation thermalization.