Case studies in machine learning via quantum annealing

Presented by: Richard Li, University of Southern California
September 26, 2018
Overview

- Machine learning
- Higgs boson
- TF-DNA binding
- Cancer classification

Machine learning

• Spam detection

• Natural language processing

• Recommender systems
Machine learning

Give computers ability to learn without being explicitly programmed

http://cs231n.github.io/classification/
(Supervised) Machine learning

**THIS IS YOUR MACHINE LEARNING SYSTEM?**

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.

https://xkcd.com/1838/
Overview

- Machine learning
- Higgs boson
- TF-DNA binding
- Cancer classification

Authors: Alex Mott¹, Joshua Job², Jean-Roch Vlimant¹, Daniel Lidar³, Maria Spiropulu¹

Associations: 1. Department of Physics, California Institute of Technology 2. Department of Physics, Center for Quantum Information Science & Technology, University of Southern California 3. Departments of Electrical Engineering, Chemistry, and Physics, Center for Quantum Information Science & Technology, University of Southern California
Higgs boson – Standard Model
Signal and background

• Signal: a Higgs decaying into two photons; i.e. the $H \rightarrow \gamma\gamma$ process

• Background processes two gluons producing two photons
  $gg \rightarrow \gamma\gamma$
The data

• <5% of events are signal of Higgs, >95% are background events of other Standard Model processes

• (Use data simulated from event generators)
The data

• Start with 8 kinematic variables, mostly involving momenta of output photon pairs

• Take various products, end with 36 final variables

• Predict whether event is Higgs or background

• Construct data so that equal number of signal and background events
The approach

- Construct weak classifiers $h(x_i)$ based on the 36 variables

- Use *boosting* to construct a strong classifier (learn the w’s)

- Minimize error $\delta(\tilde{w})$ between prediction and output

\[ p(y_i = 1|h(x_i) > p(y_i = 1) > 0 \]
\[ -1 \leq h(x_i) \leq 1, \quad \forall x_i \]

\[ O(x) = \sum_j w_j h_j(x) \]

\[ O(x) < O_c \rightarrow B \]
\[ O(x) > O_c \rightarrow S \]

\[ \delta(\tilde{w}) = \sum_{x \in T} (y(x) - \sum_i w_i h_i(x))^2 \]
Construction of the Weak Classifiers

- Truncate Distributions
- Assign overflowing events to high/low bins

- Invert distribution
- Center at 0

- Rescale distribution so it fills [-1,1]
Results

• Interpretability: found variables that appear the most important for classification

\[ [p_T^2, (\Delta R p_T^{\gamma \gamma})^{-1}, \text{and } \frac{p_T^2}{p_T^{\gamma \gamma}}] \]

• Classification performance (ROC)
Overview

- Machine learning
- Higgs boson
- TF-DNA binding
- Cancer classification

“Quantum annealing versus classical machine learning applied to a simplified computational biology problem”, npjQI (2018)

Authors: Richard Li, Rosa Di Felice, Remo Rohs, Daniel Lidar
Background

Central dogma of molecular biology

The data

• Three PBM datasets: Mad, Max, Myc

• Get sequences and log signal intensities\{ (\vec{x}_n, y_n) \}

\[
\vec{x}_n = (\{x_{n,1}, x_{n,2}, \ldots, x_{n,L}\})
\]

\[x_{n,j} \in \{A, C, G, T\}, j = 1, \ldots, L\]

• Encode sequences with one-hot encoding

\[\phi_n \equiv \phi(\vec{x}_n) = (\phi_1(\vec{x}_n), \phi_2(\vec{x}_n), \ldots, \phi_M(\vec{x}_n))\]
(A slide of math)

\[ f_w(\phi_n) = w^\top \phi_n = \sum_{m=1}^{M} w_m \phi_m(\bar{x}_n) \]

\[ w_{\text{opt}} = \arg \min_w \sum_{n=1}^{N} \left( y_n - f_w(\phi_n) \right)^2 + \lambda \|w\|_1 \]

\[ = \arg \min_w \sum_{n} y_n^2 - 2y_n f_w(\phi_n) + f_w(\phi_n)^2 + \lambda \|w\|_1 \]

\[ = \arg \min_w \sum_{m} \sum_{n} -2y_n \phi_m(\bar{x}_n) w_m + \sum_{m,m'} \left( \sum_{n} \phi_m(\bar{x}_n) \phi_{m'}(\bar{x}_n) \right) w_m w_{m'} + \sum_{m} \lambda w_m \]

\[ = \arg \min_w \sum_{m} h_m w_m + \sum_{m,m'} J_{m,m'} w_m w_{m'}, \quad \text{where} \]

\[ h_m = \lambda - 2 \sum_{n} y_n \phi_m(\bar{x}_n) \quad \text{and} \quad J_{m,m'} = \sum_{n} \phi_m(\bar{x}_n) \phi_{m'}(\bar{x}_n) \]
Workflow

1. Calibrate
   - Evaluate different calibrations
   - \( \lambda^{(1)} \), \( \lambda^{(2)} \), ..., \( \lambda^{(22)} \)
   - Use best calibration (train on 50 random subsets)
   - 100-fold Monte Carlo cross-validation

2. Train
   - Select 2% and 10% of training data
   - \( \lambda^{(2)} \)

3. Evaluate performance
   - Test data
   - Test AUPRC
   - Average validation AUPRC
   - \( w_1^* \), \( w_2^* \), ..., \( w_{50}^* \)
Results

- Area under the precision-recall curve (AUPRC)
- Kendall's $\tau$, a rank correlation coefficient
- Agreement with previously known biological information
Overview

• Machine learning
• Higgs boson
• TF-DNA binding
• Cancer classification

“Classifying cancer multi-omics data on a quantum computer” (in preparation)

Omar E. Gamel, Richard Y. Li, Nicholas A. Cilfone, Daniel A. Lidar, Thomas W. Chittenden
The data

• An ensemble multi-omics cancer data pipeline
  • A variety of different signals
  • Construct metagenes

• Used cancer types from the cancer genome atlas

Cilfone et. al, under review
The approach

- Assign probability to each class by using softmax

\[
Pr(y_i = c) = \frac{\exp w^T_c x_i}{1 + \sum_{k=1}^{K-1} \exp w^T_k x_i}, \quad c = 1, \ldots, K - 1
\]

\[
Pr(y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} \exp w^T_k x_i}
\]

\[
\{w_k^*\} = \arg \min \{w \mid \{w_k\}\} \mathcal{L}
\]

\[
\mathcal{L} = -\log \prod_i Pr(y_i) = -\sum_i \log Pr(y_i)
\]

\[
= -\sum_{i : y_i \in 1 \ldots K-1} w^T_{y_i} x_i + \sum_i \left[ \log K + \frac{1}{K} \sum_k z^{(k)}_i + \frac{1}{2} \frac{K-1}{K^2} \sum_k \left( z^{(k)}_i \right)^2 - \frac{1}{2K^2} \sum_{j \neq k} z^{(j)}_i z^{(k)}_i \right]
\]

\[
= \sum_{k=1}^{K-1} w^T_k (b_k + h') + \sum_{k=1}^{K-1} w^T_k J' w_k - \sum_{k=1}^{K-1} \sum_{j \neq k} w^T_j J'' w_k
\]

\[
b_k = \sum_{i : y_i = k} -x_i, \quad h' = \sum_i x_i, \quad J' = \frac{K-1}{2K^2} \sum x_i x_i^T, \quad \text{and} \quad J'' = \frac{1}{2K^2} \sum x_i x_i^T
\]
Results

• Breast cancer tumor vs normal
• Estrogen receptor + vs –
• Kidney clear vs papillary cell carcinoma
• Lung squamous cell vs adenocarcinoma
• Luminal A vs Luminal B
• 6 cancer types
Results

• Comparison on one dataset using “raw” meta-genes

• DW has a slight advantage for some performance metrics, not for some others
Summary

• Higgs Boson
  • Construct weak classifiers on kinematic variables
  • Use boosting to predict class

• TF-DNA binding
  • sequence-based features
  • Predict binding affinity

• Cancer classification
  • Multi-omics data
  • Predict probabilities of classes
Summary

• Seems to do slightly better on small training sizes
• Not making a claim or a quantum advantage in general (usually SA does as well)
• Enforcing binary weights means a simpler model
• Fairly interpretable
THANKS FOR LISTENING!
Math sketch:

- $S$ is the signal distribution, $B$ background, $v$ is the variable
- $v_{low}$ and $v_{high}$ are the 30th and 70th percentiles of $S$, $b_{low}$ and $b_{high}$ the percentiles on $B$ at those values
- If $b_{high} < 0.7$ then define $v_{shift} = v_{low} - v$, elseif $b_{low} > 0.7$ then $v_{shift} = v - v_{high}$, else reject $v$
- Define $v_{+1}$ and $v_{-1}$ as the 10th and 90th percentile of the transformed $S$ distribution
- With this formulation, the weak classifier is given by
- Do this for all the variables and products (or, if flipped flipped, the ratio)

\[
h(v) = \begin{cases} 
+1 & \text{if } v_{+1} < v_{shift}(v) \\
\frac{v_{shift}(v)}{v_{+1}} & \text{if } 0 < v_{shift}(v) \leq v_{+1} \\
\frac{v_{shift}(v)}{|v_{-1}|} & \text{if } v_{-1} < v_{shift}(v) \leq 0 \\
-1 & \text{if } v_{shift}(v) < v_{-1}
\end{cases}
\]
The diagram shows the Max and TCF4 sequences with different training conditions.

- **Max**
  - DW: CACGTG
  - SA: CACGTG
  - MLR: CACGTG

- **TCF4**
  - DW: CACCTG
  - SA: CACCTG
  - MLR: CACCTG

**Training Conditions**
- Training with 1% of training data
- Training with 5% of training data
- Training with 2% of training data
- Training with 10% of training data
Results
Higgs Boson Pizza Day at CERN

What's happening on my Ham & Cheese pizza?

A few micrometers (micron, or millionth of a meter) thick slice of the Higgs boson, which... (text cut off)

What's happening on my Vegetarian pizza?

A few micrometers (micron, or millionth of a meter) thick slice of the Higgs boson, which... (text cut off)

Liberal-arts majors may be annoying sometimes, but there's nothing more annoying than a physicist first encountering a new subject.