Minimum Precision Requirements for the SVM-SGD Learning Algorithm

Charbel Sakr, Ameya Patil, Sai Zhang, Yongjune Kim, Naresh Shanbhag

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Hyperplane Classification Example

Can we control this behavior?

Geometry of classifier and data

Fixed-point Quantization

Geometry in fixed-point
Prior Work

**Fixed-point Training**
- Stochastic Rounding for regularization
  - Deep Learning with Limited Numerical Precision (Gupta et al., ICML, 2015)
  - BinaryConnect (Courbariaux et al., NIPS, 2015)
  - BinaryNet (Courbariaux et al., arXiv, 2016)
- Bitwise Operations (Kim & Smaragdis, arXiv, 2016)
- XNOR-net (Rastegari et al., ECCV, 2016)

Training in a discrete space is harder

**Fixed-point Quantization After Training**
- Exhaustive Search (Hwang & Sung, SiPS, 2014)
  - Trial-and-error may be hard
- SQNR based precision allocation (Lin et al., ICML, 2016)
  - No theoretical guarantees on accuracy
The SVM-SGD learning algorithm

Classification

\[ \hat{y}_n = \begin{cases} 1 & \text{if } \mathbf{w}^T \mathbf{x}_n + b > 0, \\ -1 & \text{otherwise}. \end{cases} \]

Training

\[ \mathbf{w}_{n+1} = (1 - \gamma \lambda) \mathbf{w}_n \]

\[ + \begin{cases} 0 & \text{if } y_n (\mathbf{w}_n^T \mathbf{x}_n + b) > 1, \\ \gamma y_n \mathbf{x}_n & \text{otherwise}. \end{cases} \]
The SVM-SGD Architecture

Classifier

Weight Update

$\mathbf{X} \xrightarrow{\mathbf{w}} \mathbf{w}^T \mathbf{X} \xrightarrow{b} \mathbf{y} \xrightarrow{\pm} \hat{y}$

$\mathbf{B}_X \xrightarrow{\mathbf{w}} \mathbf{B}_F \xrightarrow{\mathbf{w}} \mathbf{B}_W \xrightarrow{\mathbf{y}} (1 - \gamma \lambda)$
Effects of Quantization

• Decision equation is modified to:

\[(w + q_w)^T(x + q_x) + b + q_b \geq 0\]

• Decision equation is modified to:

\[q_x \in \mathcal{R}^N, \ q_w \in \mathcal{R}^N, \ \text{and} \ q_b \in \mathcal{R}\]

\[q_x \sim (U[-\frac{\Delta_x}{2}, \frac{\Delta_x}{2}])^N; \ \Delta_x = 2^{-(B_x - 1)}\]

\[q_w \sim (U[-\frac{\Delta_f}{2}, \frac{\Delta_f}{2}])^N; \ \Delta_f = 2^{-(B_f - 1)}\]

\[q_b \sim U[-\frac{\Delta_f}{2}, \frac{\Delta_f}{2}]\]

• Total quantization noise can be simplified to:

\[q_w^T x + w^T q_x + q_b\]
Geometric Bound

\[ B_X > \log_2 \left( \frac{\sqrt{N} \|w\|}{1 - (1 + \sqrt{N} \|x\|) 2^{-B_F}} \right) \]

- Dependence on margin
- Trade-off between weight and data precision
- Dependence on dimensionality
Probabilistic Bound

\[ p_m \leq \frac{1}{24} \left( \Delta_x^2 \mathbb{E} \left[ \frac{||w||^2}{|w^T X + b|^2} \right] + \Delta_f^2 \mathbb{E} \left[ \frac{||X||^2 + 1}{|w^T X + b|^2} \right] \right) \]

\[ \Delta_x = 2^{-(B_X - 1)} \quad \Delta_f = 2^{-(B_F - 1)} \]

- Data dependence
- Exponential trade-off between precision and accuracy
- Compute once and reuse
To ensure non-zero updates:

\[ B_W \geq B_X - \log_2(\gamma) \]
Simulation Results

• Dataset: Breast Cancer Dataset from UCI Machine Learning Repository.

• Classification: Fix $B_F$ and sweep $B_X$. We compare fixed-point simulations to analysis.

• Training: Fix $B_X$ and $B_F$ and sweep $B_W$. We compare floating-point convergence curves to fixed-point simulations.

• Energy estimation: We use the methodology from (Abdallah & Shanbhag, TVLSI, 2014) on a 45 nm CMOS process to estimate the energy savings of reducing precision.
Classification

Probabilistic upper bound

Geometric Bound

Fixed-point simulations

$\mathcal{P}_e$ vs $B_X$ (bits)

$B_F = 6$
Training

\[ B_W = B_X - \log_2(\gamma) \]

\[ B_W \leq B_X - \log_2(\gamma) \]

Floating-point simulations

\[ B_X = 6, B_F = 6, \gamma = 2^{-5} \]
Energy Savings

![Energy Savings Graph]

- Minimum precision
- Minimum precision + 2 bits
- Uniform 16-bit precision

$E(J)$ vs. $V_{dd}(V)$

5.3x energy savings
Conclusion

• We presented analytical requirements on the fixed-point precision in the context of the SVM-SGD algorithm.

• Ongoing Work: Extension of results for complete deep learning systems.

Acknowledgments

This work was supported in part by Systems on Nanoscale Information fabriCs (SONIC), one of the six SRC STARnet Centers, sponsored by MARCO and DARPA.
Thank you!