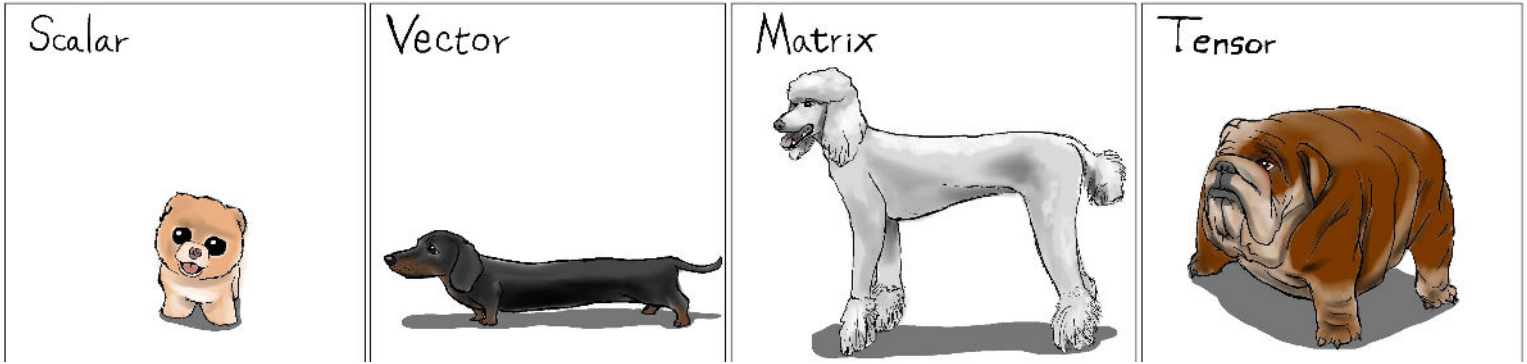
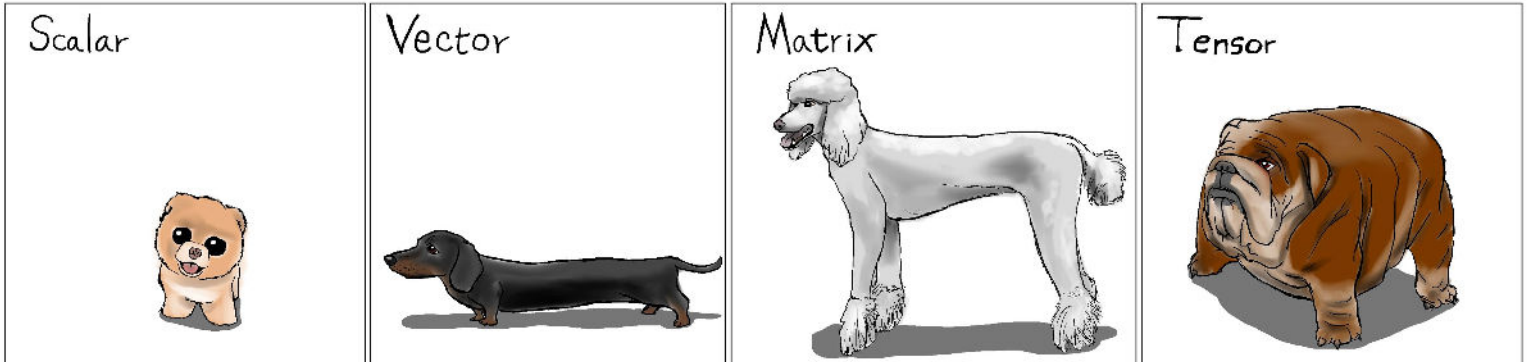
Role of Tensors in Large-Scale Machine Learning

Anima Anandkumar
Principal Scientist, AWS
Bren Professor, Caltech

Tensors are Multi-Dimensional



Tensors are Multi-Dimensional



Modern data is inherently multi-dimensional

Tensors to encode multi-dimensional data

Images: 3 dimensions

Tensors to encode multi-dimensional data

Images: 3 dimensions

Videos: 4 dimensions

Tensors to encode multi-dimensional data

Images: 3 dimensions

Videos: 4 dimensions

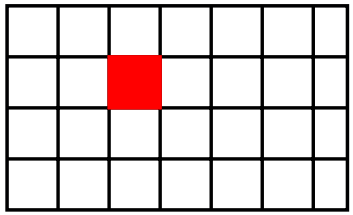
Are there trees in the image?

Visual q&a: ? dimensions

Tensors to encode higher order moments

Pairwise correlations

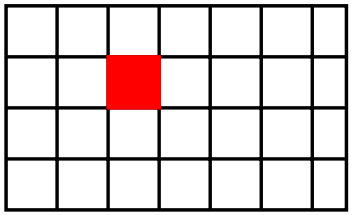
$$E(x \otimes x)_{i,j} = E(x_i x_j)$$



Tensors to encode higher order moments

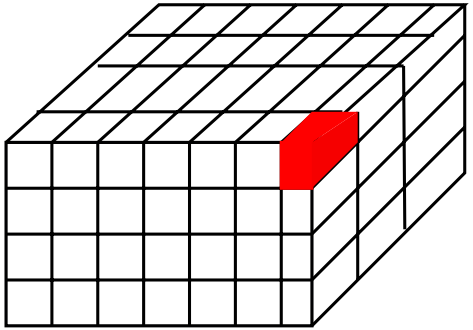
Pairwise correlations

$$E(x \otimes x)_{i,j} = E(x_i x_j)$$



Third order correlations

$$E(x \otimes x \otimes x)_{i,j,k} = E(x_i x_j x_k)$$

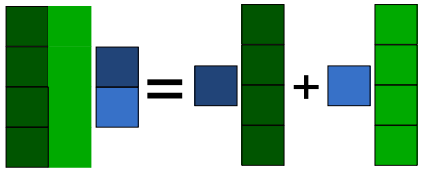


Operations on Tensors: Tensor Contraction

Extends the notion of matrix product

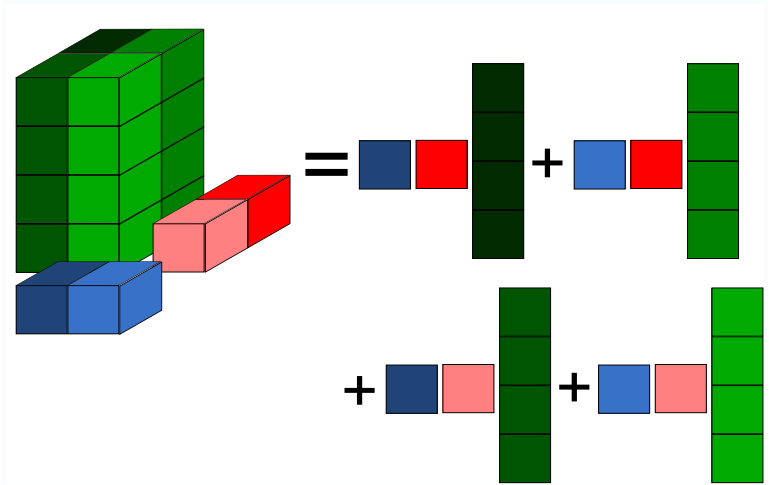
Matrix product

$$Mv = \sum_j v_j M_j$$



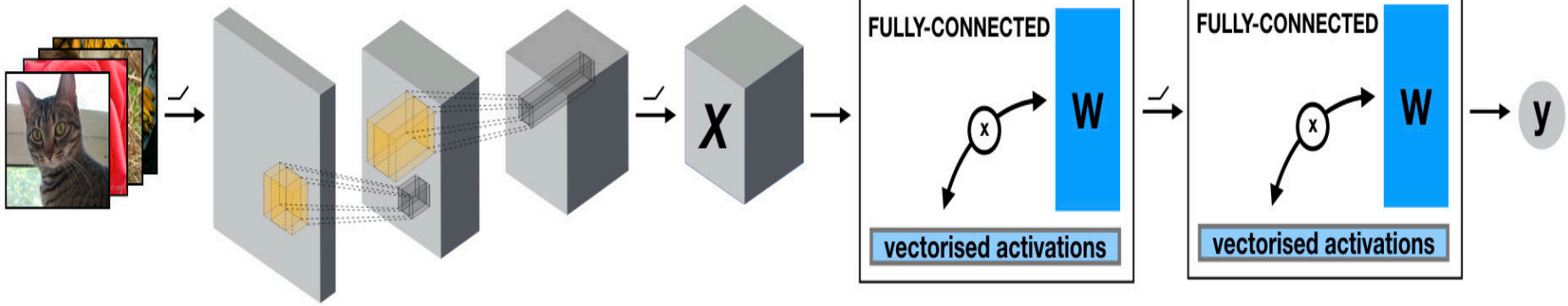
Tensor Contraction

$$T(u, v, \cdot) = \sum_{i,j} u_i v_j T_{i,j,:}$$

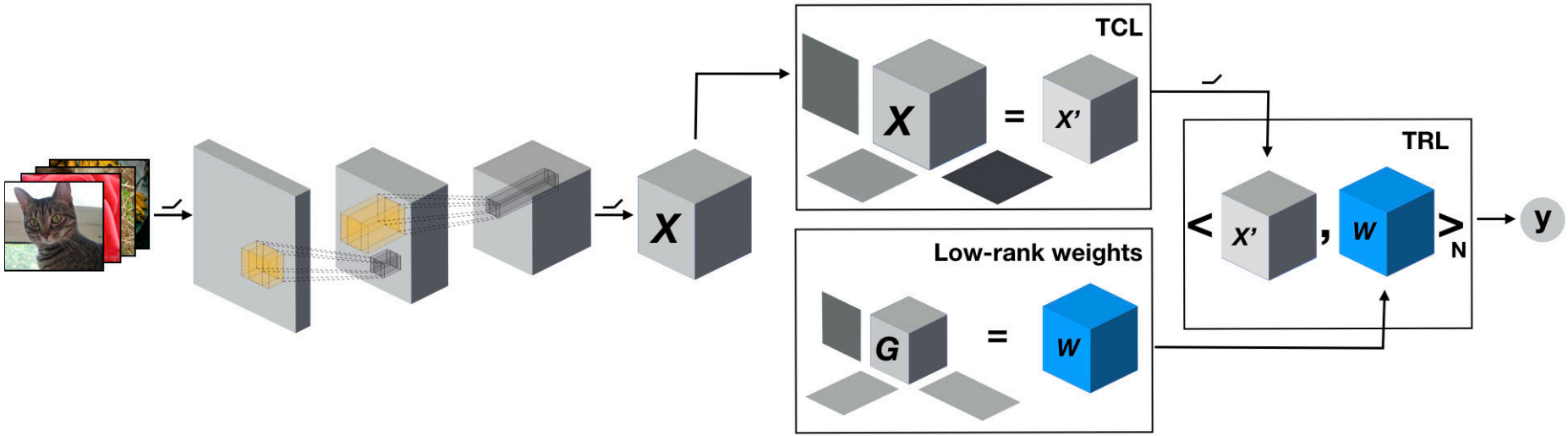


Tensorized Neural Networks

Deep Neural Nets: Transforming Tensors

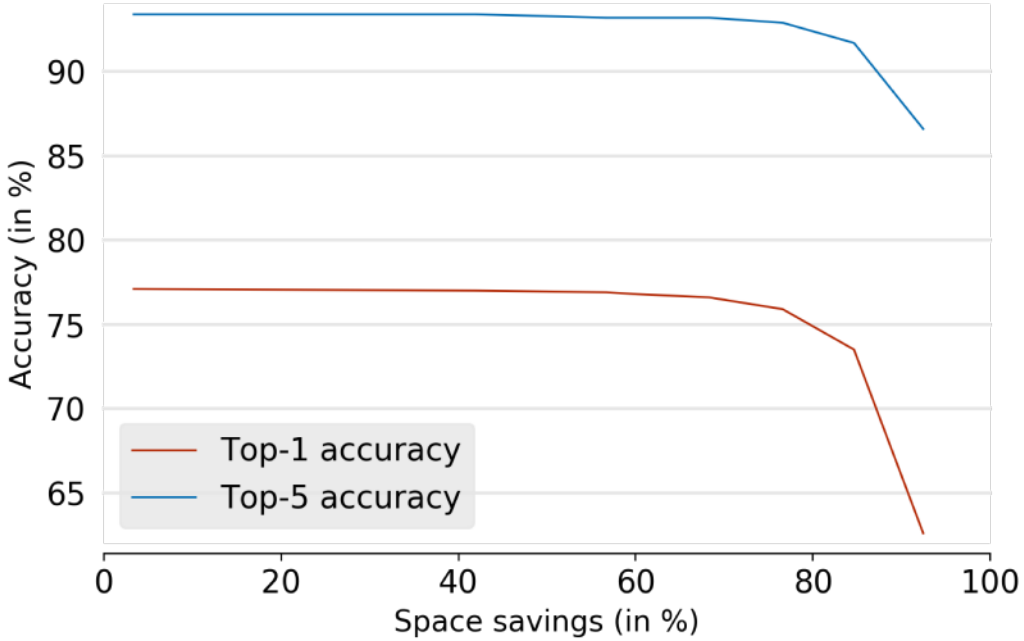


Deep Tensorized Networks

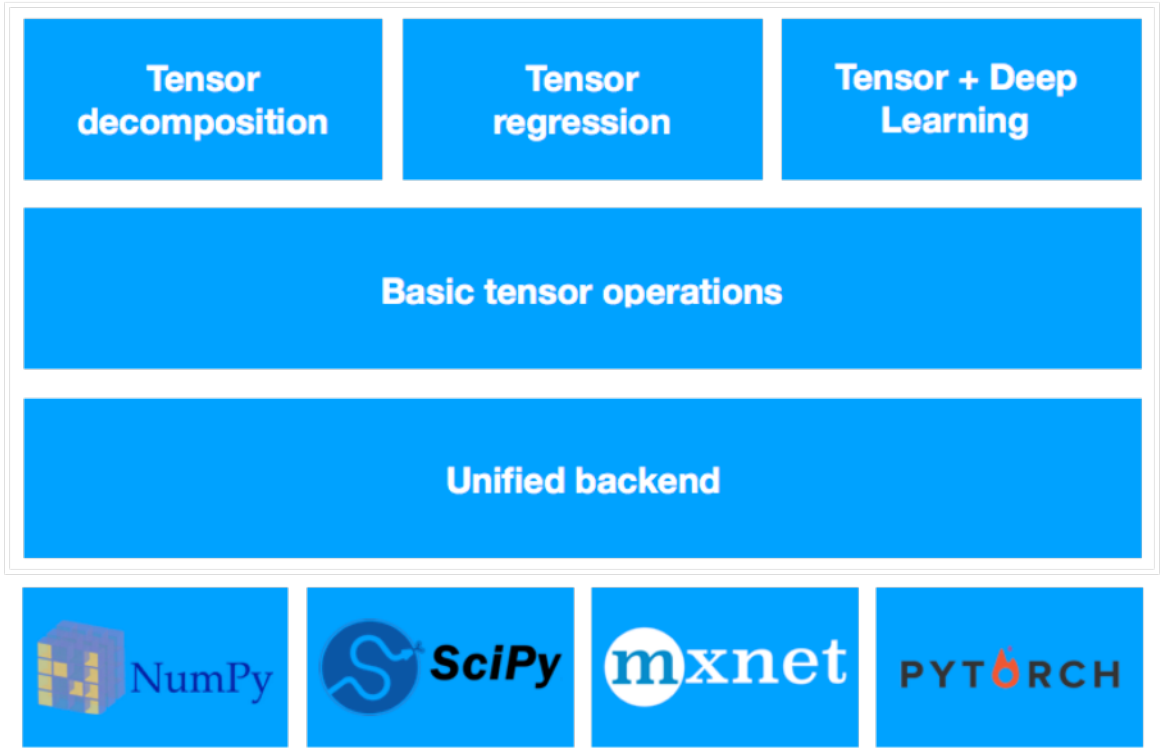


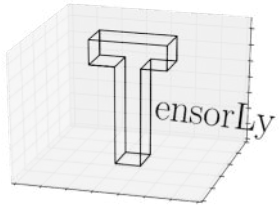
Tensor Contraction & Regression Networks by J. Kossaifi, Z. C. Lipton, A. Khanna, T. Furlanello, **A,** 2017.

Space Saving in Deep Tensorized Networks



Tensorly: Framework for Tensor Algebra





- Python programming
- User-friendly API
- Multiple backends: flexible + scalable
- Example notebooks in repository

Tensorly with Pytorch backend

```
import tensorly as tl
from tensorly.random import tucker_tensor

tl.set_backend('pytorch')
core, factors = tucker_tensor((5, 5, 5),
                              rank=(3, 3, 3))
core = Variable(core, requires_grad=True)
factors = [Variable(f, requires_grad=True) for f in factors]

optimiser = torch.optim.Adam([core]+factors, lr=lr)

for i in range(1, n_iter):
    optimiser.zero_grad()
    rec = tucker_to_tensor(core, factors)
    loss = (rec - tensor).pow(2).sum()
    for f in factors:
        loss = loss + 0.01*f.pow(2).sum()

    loss.backward()
    optimiser.step()
```

Set Pytorch backend

Tucker Tensor form

Attach gradients

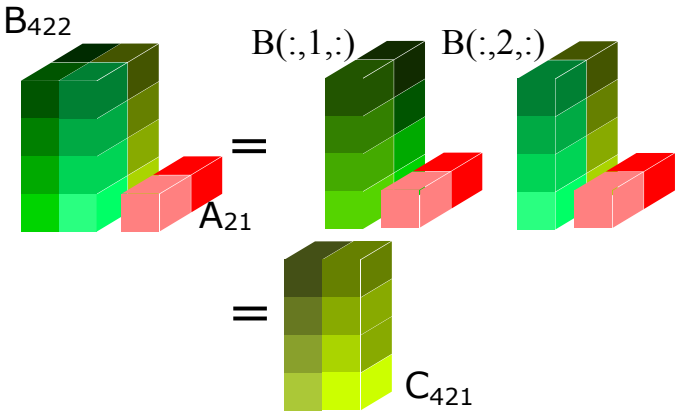
Set optimizer

Speeding up Tensor Contractions

Speeding up Tensor Contractions

1. Tensor contractions are a core primitive of multilinear algebra.
2. BLAS 3: Unbounded compute intensity (no. of ops per I/O)

Consider single-index contractions: $C_{\mathcal{C}} = A_{\mathcal{A}}\, B_{\mathcal{B}}$



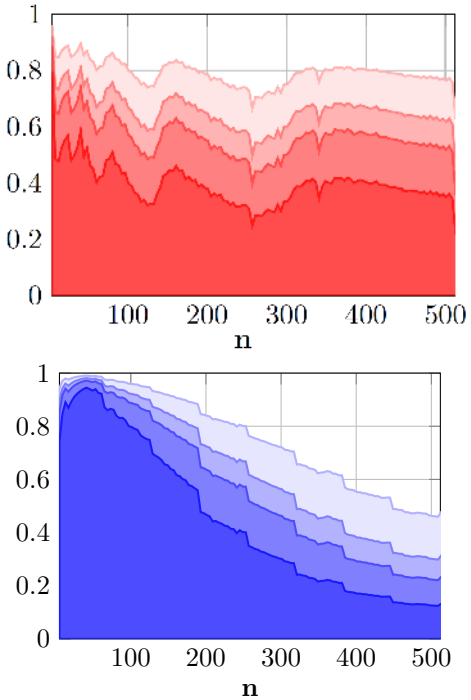
e.g. $C_{mnp} = \mathrm{B}_{mnk}\, \mathrm{A}_{kp}$

Speeding up Tensor Contractions

Explicit permutation dominates, especially for small tensors.

Consider $C_{mnp} = A_{km}\, B_{pkn}$.

1. $A_{km} \rightarrow A_{mk}$
2. $B_{pkn} \rightarrow B_{kpn}$
3. $C_{mnp} \rightarrow C_{mpn}$
4. $\boxed{C_{m(pn)} = A_{mk}\, B_{k(pn)}}$
5. $C_{mpn} \rightarrow C_{mnp}$



(Top) CPU. (Bottom) GPU. The fraction of time spent in copies/transpositions. Lines are shown with 1, 2, 3, and 6 transpositions.

New Primitive for Tensor Contractions

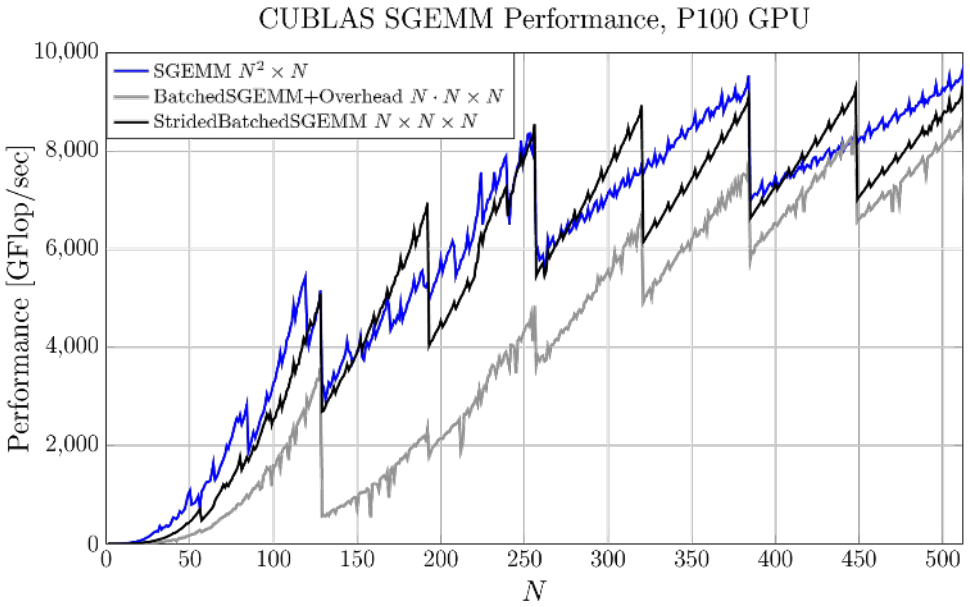
$$C[p] = \alpha\,\mathsf{op}(A[p])\,\mathsf{op}(B[p]) + \beta\,C[p]$$

- Pointer-to-Pointer BatchedGEMM requires memory allocation and pre-computation.
- Solution: StridedBatchedGEMM with fixed strides.
  - Special case of Pointer-to-pointer BatchedGEMM.
  - No Pointer-to-pointer data structure or overhead.

```
cublas<T>gemmStridedBatched(cublasHandle_t handle,
                            cublasOperation_t transA, cublasOperation_t transB,
                            int M, int N, int K,
                            const T* alpha,
                            const T* A, int ldA1, int strideA,
                            const T* B, int ldB1, int strideB,
                            const T* beta,
                            T* C, int ldC1, int strideC,
                            int batchCount)
```

Performance of StridedBatchedGEMM

- Performance on par with pure GEMM (P100 and beyond).

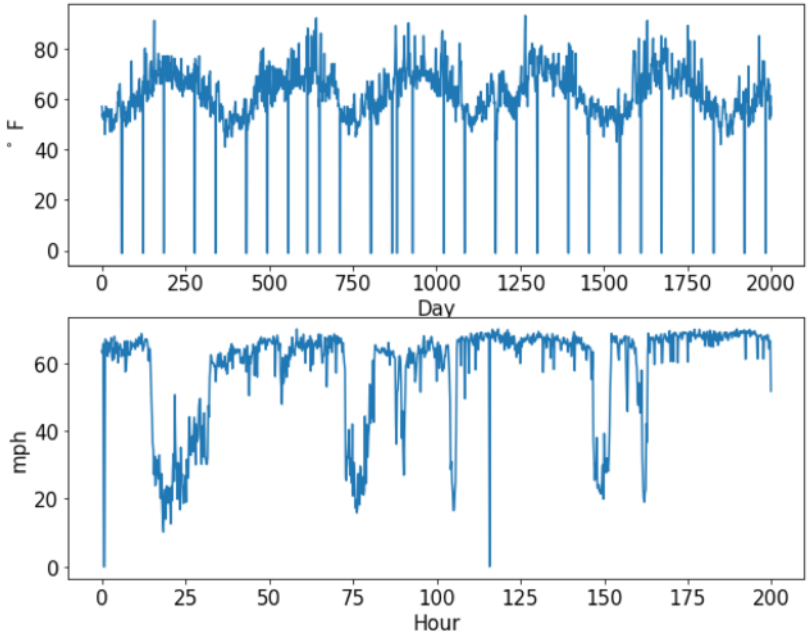


Tensors in Time Series

Tensors for long-term forecasting

Difficulties in long term forecasting:

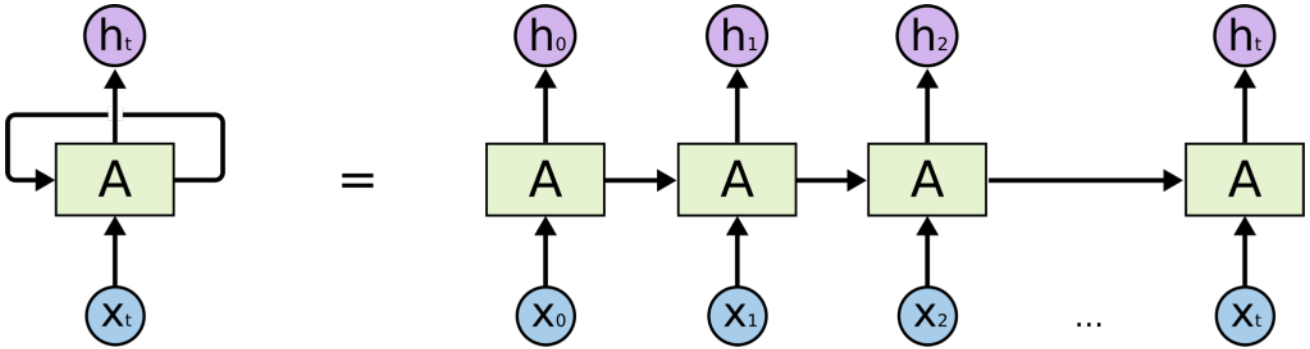
- Long-term dependencies
- High-order correlations
- Error propagation



RNNs: First-Order Markov Models

Input state x_t, hidden state h_t, output y_t,

$$h_t = f(x_t, h_{t-1}; \theta); \; y_t = g(h_t; \theta)$$



Tensorized RNNs

L-order Markov

$$h_t = f(x_t, h_{t-1}, h_{t-2}, \dots h_{t-L}; \theta);\ y_t = g(h_t; \theta)$$

Polynomial Interactions

$$s_{t-1} = [1, h_{t-1}, h_{t-2}, \dots h_{t-L}]$$

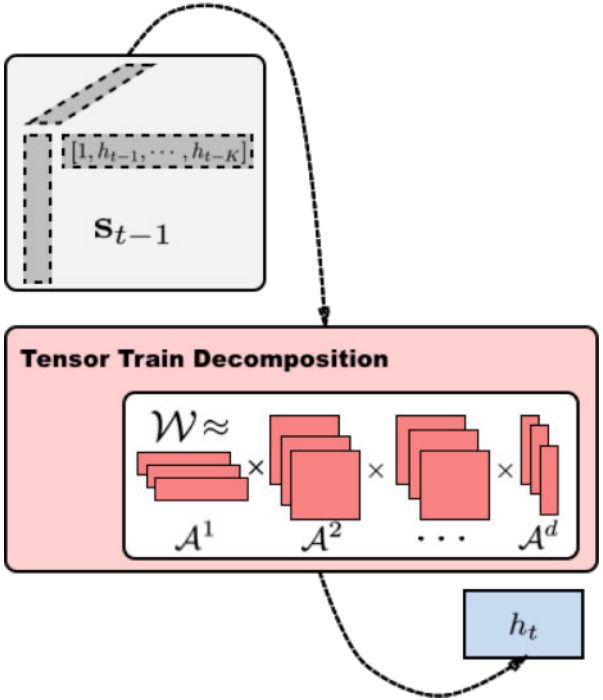
$$h_{t,\alpha} = f\left(W^{hx} x_t + \mathcal{W}_{\alpha, i_1, \dots i_P} s_{t-1;i_1} \otimes \dots \otimes s_{t-1;i_P}; \theta\right);$$

Tensor-Train Decomposition

Reduce # of parameters

Linear tensor network

Dimension-free decomposition

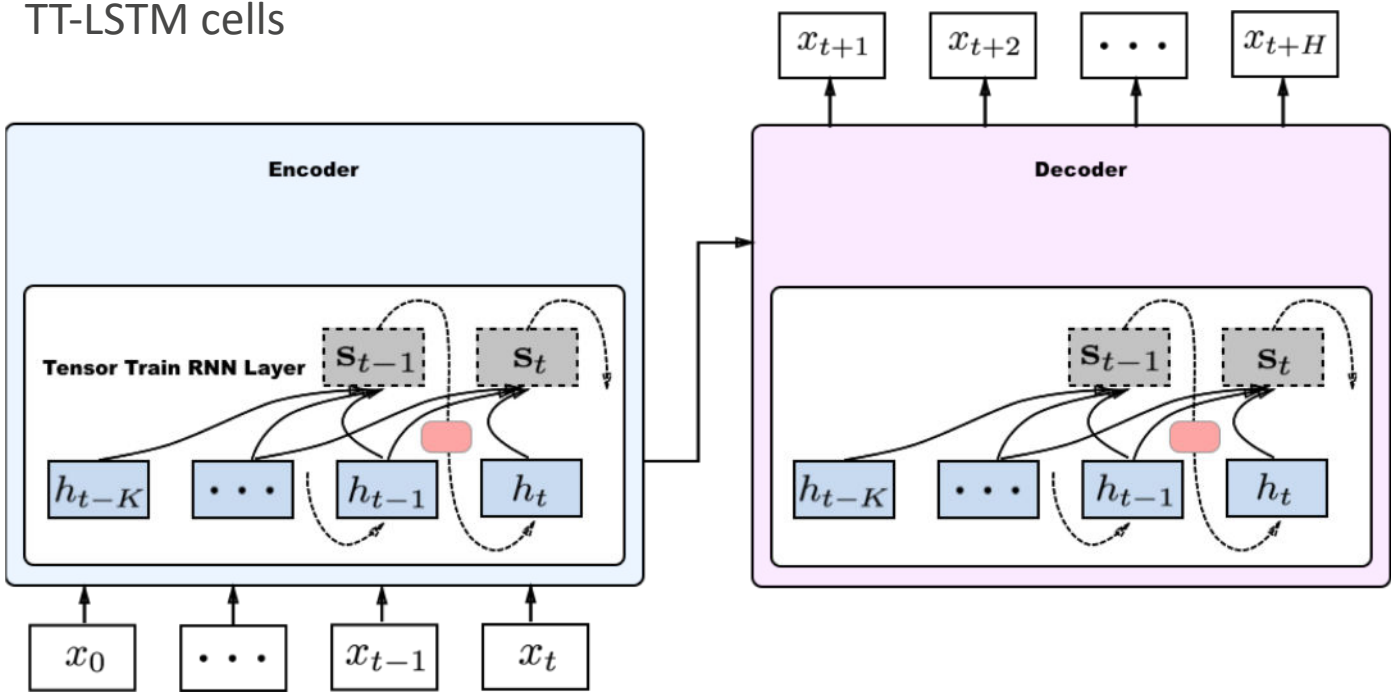


$$\mathcal{W}_{i_1 \ldots i_P} = \sum_{\alpha_0 \ldots \alpha_P} \mathcal{A}_{\alpha_0 i_1 \alpha_1} \cdots \mathcal{A}_{\alpha_{P-1} i_P \alpha_P}$$

Tensor-Train RNNs and LSTMs

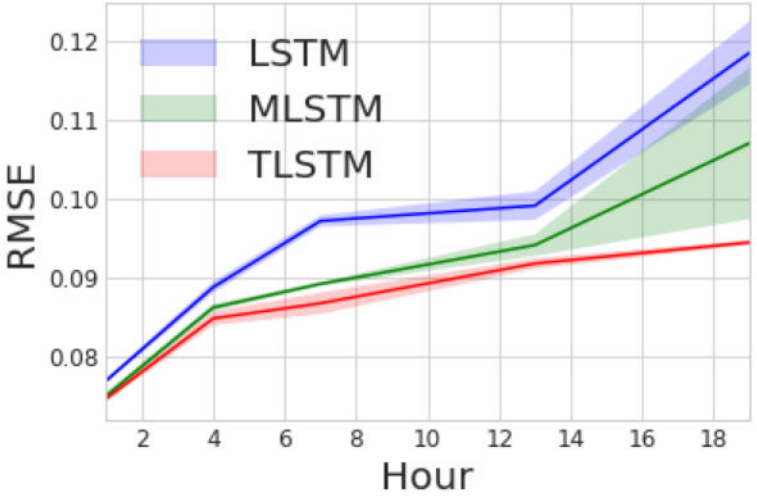
Seq2seq architecture

TT-LSTM cells

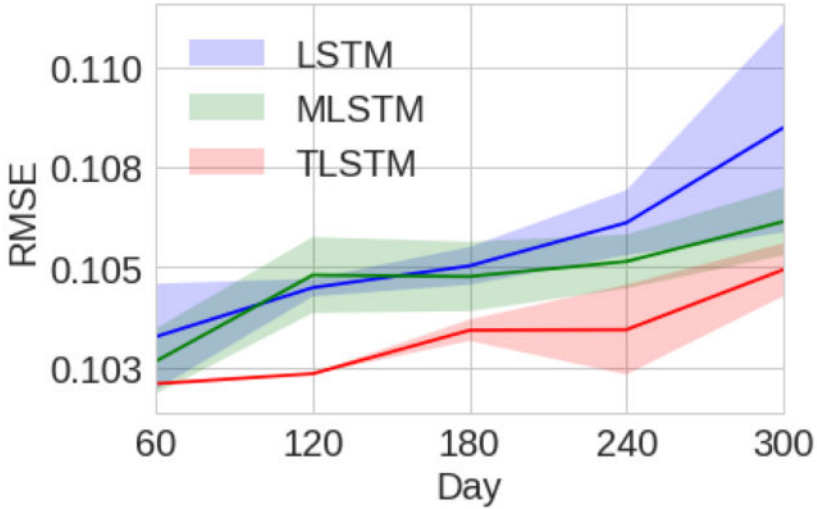


Tensor LSTM for Long-term Forecasting

Traffic dataset



Climate dataset



Approximation Guarantees

Theorem 1 [Yu et al 2017]: *Let the state transition function $f \in \mathcal{H}_{\mu}^{k}$ be a Hölder continuous, with bounded derivatives up to order k and finite Fourier magnitude distribution C_f. Then a single layer Tensor Train RNN can approximate f with an estimation error of ε using with h hidden units:*

$$h \leq \frac{C_f^2}{\varepsilon}(d-1)\frac{(r+1)^{-(k-1)}}{(k-1)} + \frac{C_f^2}{\varepsilon}C(k)p^{-k}$$

Where d is the size of the state space, r is the tensor-train rank and p is the degree of high-order polynomials i.e., the order of tensor.

- Easier to approximate if function is **smooth**
- **Polynomial interactions** are more efficient

Long-term Forecasting using Tensor-Train RNNs by R. Yu, S. Zheng, **A**, Y. Yue, 2017.

Tensors for multiple modalities

Visual Question & Answering

Feature Extraction

Pooling Methods

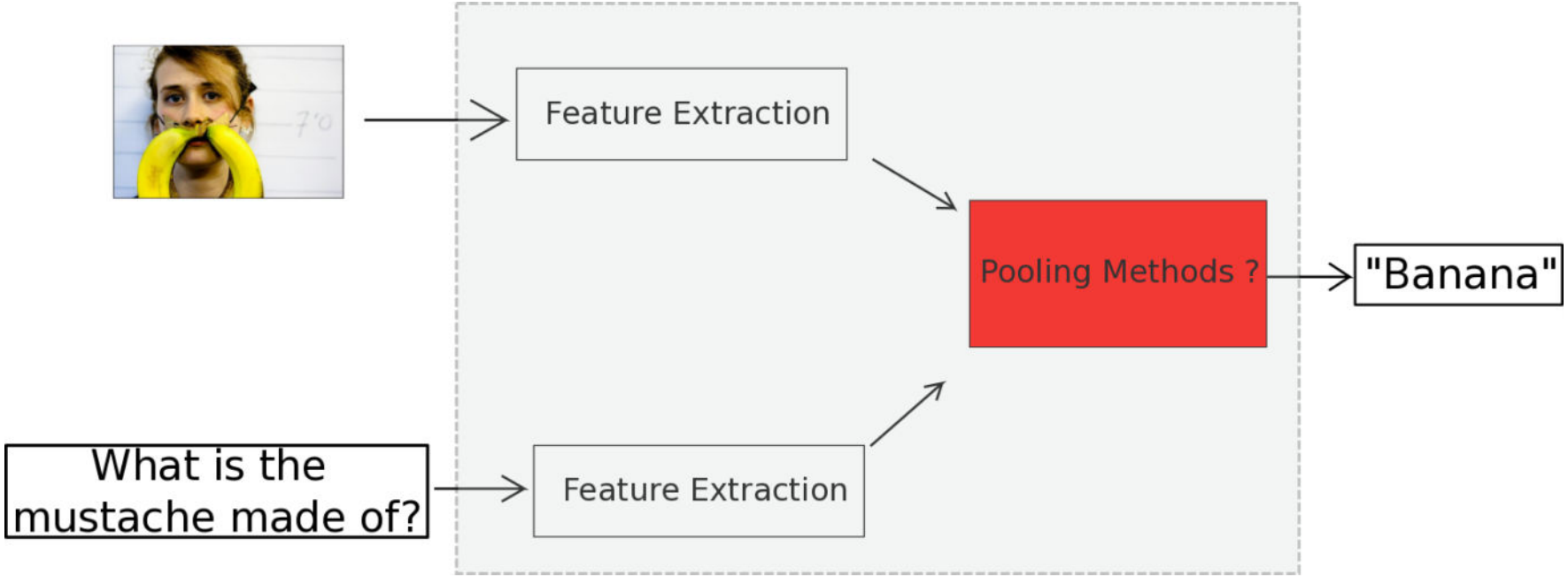
"Banana"

What is the mustache made of?

Feature Extraction

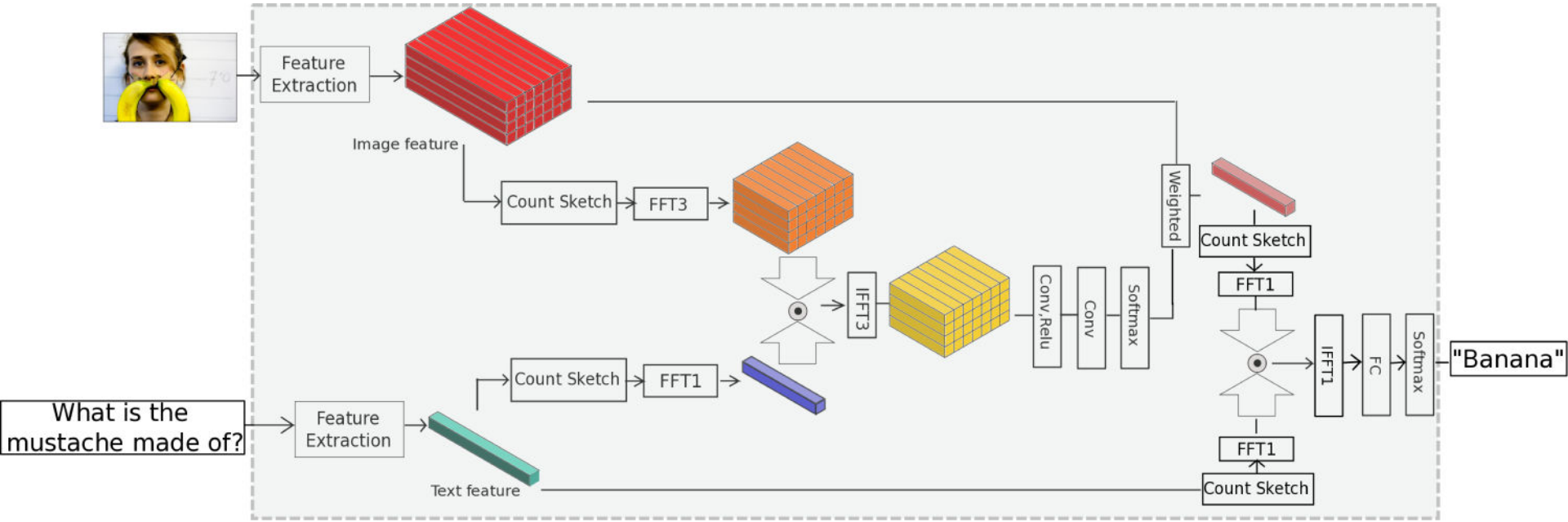
Tensors for multiple modalities

Visual Question & Answering



Tensor Sketching Algorithms

Visual Question & Answering



Tensors for Topic Modeling

Tensors for Topic Detection

The New York Times

U.S.

High-Tech Industry, Long Shy of Politics, Is Now Belle of Ball

By LIZETTE ALVAREZ DEC. 26, 1999

Correction Appended

At a time when Congress is bitterly divided and unable to reach consensus on issues like gun control and health care, Democrats and Republicans are happily reaching across party lines to pass legislation backed by high-tech companies.

The high-tech industry, at the same moment, is lavishing new attention on Washington and changing its once-aloof posture toward the federal government.

Republicans and Democrats are both eager to win the loyalties of high-tech companies and executives, knowing that they represent untold jobs, wealth and ultimately votes and campaign contributions.

For its part, the industry has realized that the federal government can do its members as much harm as good. Microsoft, and its battle with the Justice Department, along with a spate of other threatened legal problems, drilled this point home.

"Microsoft was a poster child for our industry," said Connie Correll, director of communications for the Information Technology Industry Council, a trade organization that represents America Online, Dell and I.B.M., among others.

In the House, Representative David Dreier, Republican of California, Robert W. Goodlatte, Republican of Virginia, and Thomas M. Davis III, a Virginia Republican whose district includes a growing number of high-tech companies, among others, are viewed as industry experts.

The New Democrats are often willing to buck their own leadership to live up to the label as technology boosters. In the past two years, their numbers have grown, to 64 this year from 41 in 1997.

These New Democrats have breakfast with high-tech executives every week and visit Silicon Valley routinely. "I think they are trying to create a mini high-tech party in a way," said Wade Randlett, a co-founder of TechNet and now an executive at Red Gorilla, an Internet company. "It's a smart political approach."

The bulging docket of high-tech issues had led to dramatic growth in the industry's lobbying on Capitol Hill. Internet and software companies have snagged some of Capitol Hill's best talent this year -- former Congressional aides whose expertise blends technology and politics. While Microsoft built a large battery of lobbyists, beginning two years ago, few other high-tech companies had Washington lobbying operations. Suddenly, companies, including Yahoo and Gateway, are busily opening Washington outposts, hiring lobbyists and starting trade associations. And in yet another coming-of-age gesture, executives are beginning to dip into their coffers.

"It's so critical that Silicon Valley be involved in Washington right now," said Chris Larsen, the 39-year-old founder of E-Loan who has made six trips here this year to meet with members. "The stakes are really high."

High-tech lobbying has become so profitable that some high-powered lobbyists are carving out specialty niches and crossing party lines to do it. Edward W. Gillespie, a former policy and communications director for Representative Dick Armey of Texas, the House majority leader, and Jack Quinn, Vice President Al Gore's former chief of staff, have teamed up to start a firm. The two have worked together on legislation to ease encryption export restrictions, a high priority for the high-tech industry. Tony Podesta, the brother of John D. Podesta, White House chief of staff,

Tensors for Topic Detection

Topics

- Government
- Information Technology
- Politics

The New York Times

U.S.

High-Tech Industry, Long Shy of Politics Is Now Belle of Ball

By LIZETTE ALVAREZ DEC. 26, 1999

Correction Appended

At a time when Congress is bitterly divided and unable to reach consensus on issues like gun control and health care, Democrats and Republicans are happily reaching across party lines to pass legislation backed by high-tech companies.

The high-tech industry, at the same moment, is lavishing new attention on Washington and changing its once-aloof posture toward the federal government.

Republicans and Democrats are both eager to win the loyalties of high-tech companies and executives, knowing that they represent untold jobs, wealth and ultimately votes and campaign contributions.

For its part, the industry has realized that the federal government can do its members as much harm as good. Microsoft, and its battle with the Justice Department, along with a spate of other threatened legal problems, drilled this point home.

"Microsoft was a poster child for our industry," said Connie Correll, director of communications for the Information Technology Industry Council, a trade organization that represents America Online, Dell and I.B.M., among others.

In the House, Representative David Dreier, Republican of California, Robert W. Goodlatte, Republican of Virginia, and Thomas M. Davis III, a Virginia Republican whose district includes a growing number of high-tech companies, among others, are viewed as industry experts.

The New Democrats are often willing to buck their own leadership to live up to the label as technology boosters. In the past two years, their numbers have grown, to 64 this year from 41 in 1997.

These New Democrats have breakfast with high-tech executives every week and visit Silicon Valley routinely. "I think they are trying to create a mini high-tech party in a way," said Wade Randlett, a co-founder of TechNet and now an executive at Red Gorilla, an Internet company. "It's a smart political approach."

The bulging docket of high-tech issues had led to dramatic growth in the industry's lobbying on Capitol Hill. Internet and software companies have snagged some of Capitol Hill's best talent this year -- former Congressional aides whose expertise blends technology and politics. While Microsoft built a large battery of lobbyists, beginning two years ago, few other high-tech companies had Washington lobbying operations. Suddenly, companies, including Yahoo and Gateway, are busily opening Washington outposts, hiring lobbyists and starting trade associations. And in yet another coming-of-age gesture, executives are beginning to dip into their coffers.

"It's so critical that Silicon Valley be involved in Washington right now," said Chris Larsen, the 39-year-old founder of E-Loan who has made six trips here this year to meet with members. "The stakes are really high."

High-tech lobbying has become so profitable that some high-powered lobbyists are carving out specialty niches and crossing party lines to do it. Edward W. Gillespie, a former policy and communications director for Representative Dick Armey of Texas, the House majority leader, and Jack Quinn, Vice President Al Gore's former chief of staff, have teamed up to start a firm. The two have worked together on legislation to ease encryption export restrictions, a high priority for the high-tech industry. Tony Podesta, the brother of John D. Podesta, White House chief of staff,

LDA Topic Model

Topics

Justice

Education

Sports

SECTIONS HOME SEARCH The New York Times

COLLEGE FOOTBALL

At Florida State, Football Clouds Justice

By MIKE McINTIRE and WALT BOGDANICH OCT. 30, 2014

Now, an examination by The New York Times of police and court records, along with interviews with crime witnesses, has found that, far from an aberration, the treatment of the Winston complaint was in keeping with the way the police on numerous occasions have soft-pedaled allegations of wrongdoing by Seminoles football players. From criminal mischief and motor-vehicle theft to domestic violence, arrests have been avoided, investigations have stalled and players have escaped serious consequences.

In a community whose self-image and economic well-being are so tightly bound to the fortunes of the nation's top-ranked college football team, law enforcement officers are finely attuned to a suspect's football connections. Those ties are cited repeatedly in police reports examined by The Times. What's more, dozens of officers work second jobs directing traffic and providing security at home football games, and many express their devotion to the Seminoles on social media.

Most recently, university officials suspended Mr. Winston for one game after he stood in a public place on campus and, playing off a running Internet gag, shouted a crude reference to a sex act. In a news conference afterward, his coach, Jimbo Fisher, said, "Our hope and belief is Jameis will learn from this and use better judgment and language and decision-making."

On Jan. 10, 2013, a female student at Florida State reported the man she believed had raped her the previous month. After learning his name, Jameis Winston, she reported him to the Tallahassee police.

In the 21 months since, Florida State officials have said little about how they handled the case, which is no investigated by the federal Depart

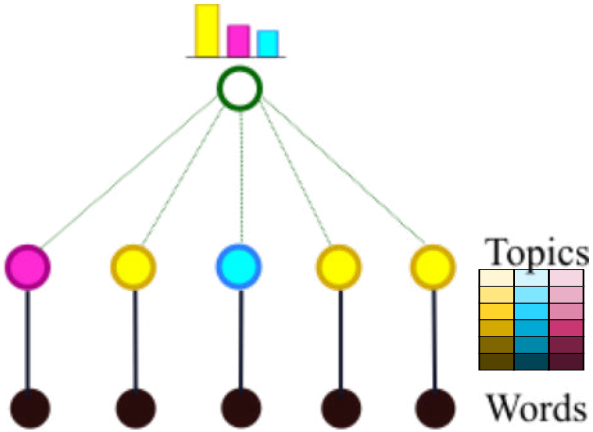
TMZ, the gossip website, also requested the police report and later asked the school's deputy police chief, Jim L. Russell, if the campus police had interviewed Mr. Winston about the rape report. Mr. Russell responded by saying his officers were not investigating the case, omitting any reference to the city police, even though the campus police knew of their involvement. "Thank you for contacting me regarding this rumor I am glad I can dispel that one!" Mr. Russell told TMZ in an email. The university said Mr. Russell was unaware of any other police investigation at the time of the inquiry. Soon after, the Tallahassee police belatedly sent their files to the news media and to the prosecutor, William N. Meggs. By then critical evidence had been lost and Mr. Meggs, who criticized the police's handling of the case, declined to

lson after the Seminoles' first game; five am's second leading receiver.

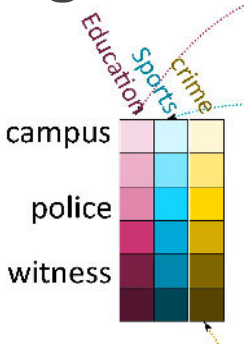
As The Times reported last April, the Tallahassee police also failed to aggressively investigate the rape accusation. It did not become public until November, when a Tampa reporter, Matt Baker, acting on a tip, sought records of the police investigation.

Upon learning of Mr. Baker's inquiry, Florida State, having shown little curiosity about the rape accusation, suddenly took a keen interest in the journalist seeking to report it, according to emails obtained by The Times.

"Can you share any details on the requesting source?" David Perry, the university's police chief, asked the Tallahassee police. Several hours later, Mr.

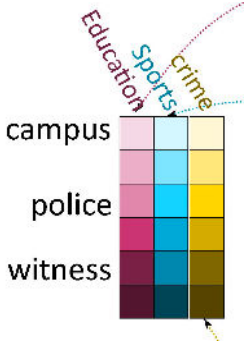


Learning LDA Model



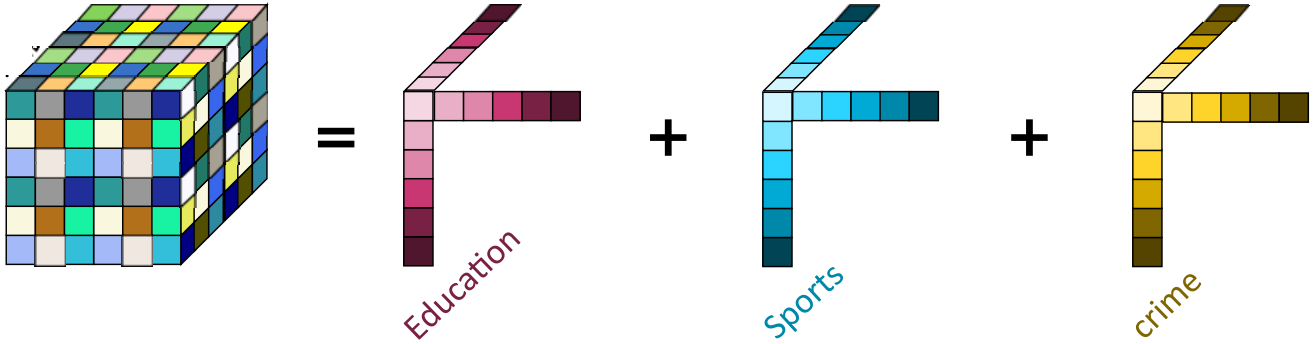
- Topic-word matrix $P[\text{word} = i | \text{topic} = j]$
- Topic proportions $P[\text{topic} = j | \text{document}]$

Learning LDA Model

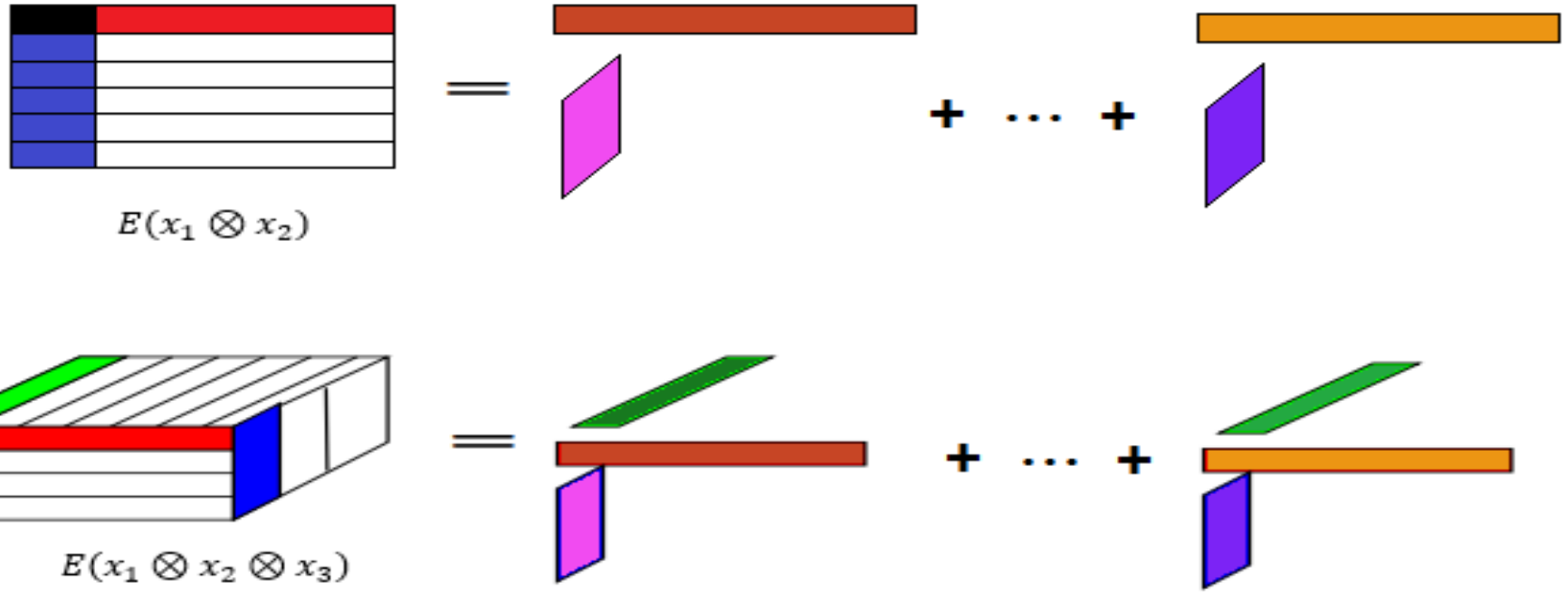


- Topic-word matrix $P[\text{word} = i | \text{topic} = j]$
- Topic proportions $P[\text{topic} = j | \text{document}]$

Moment Tensor: Co-occurrence of Word Triplets



Spectral Decomposition



Why Tensors?

Statistical reasons:

- Incorporate **higher order** relationships in data
- Discover hidden topics (not possible with matrix methods)

A. Anandkumar etal,Tensor Decompositions for Learning Latent Variable Models, JMLR 2014.

Why Tensors?

Statistical reasons:

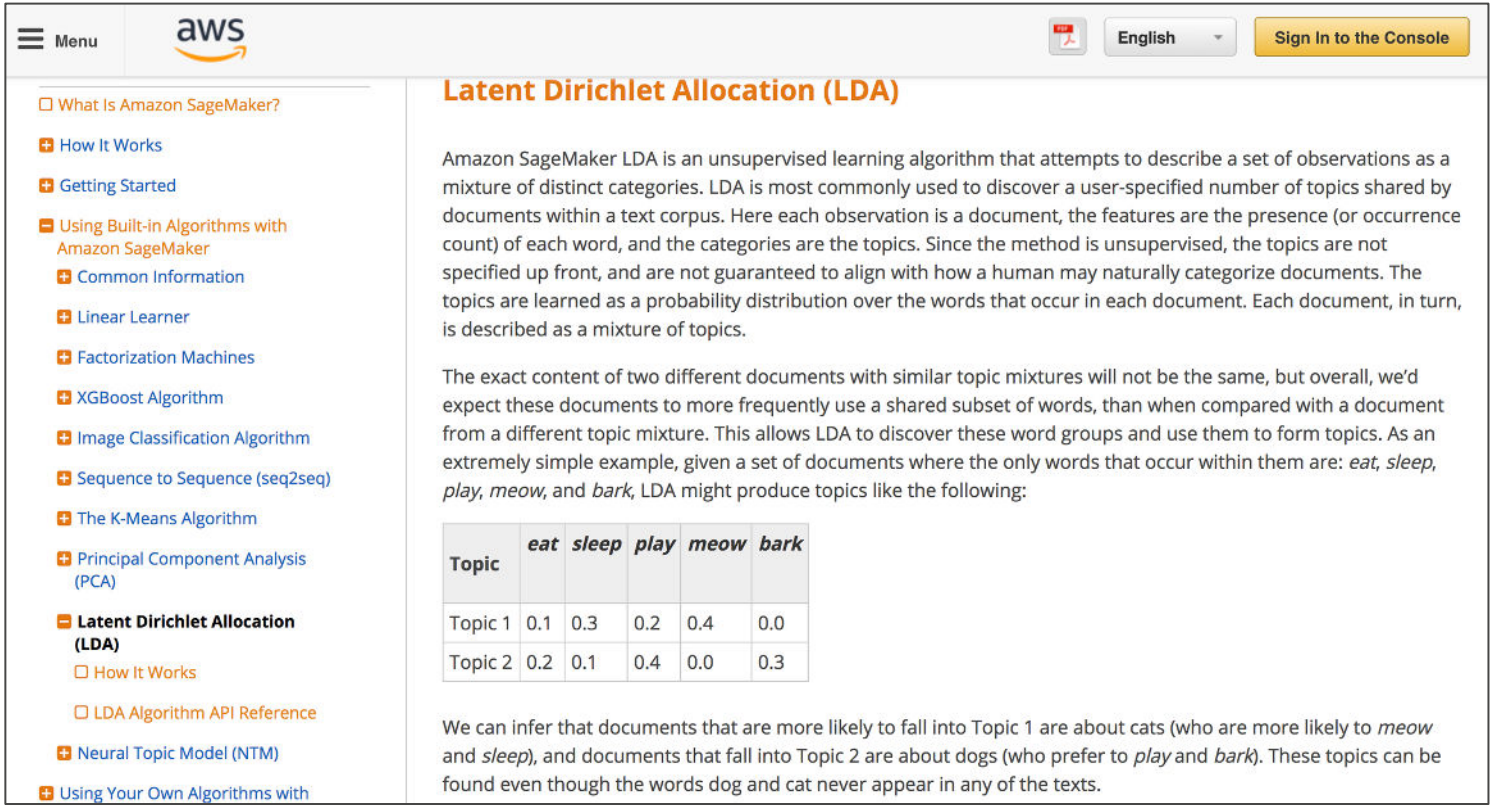
- Incorporate **higher order** relationships in data
- Discover hidden topics (not possible with matrix methods)

Computational reasons:

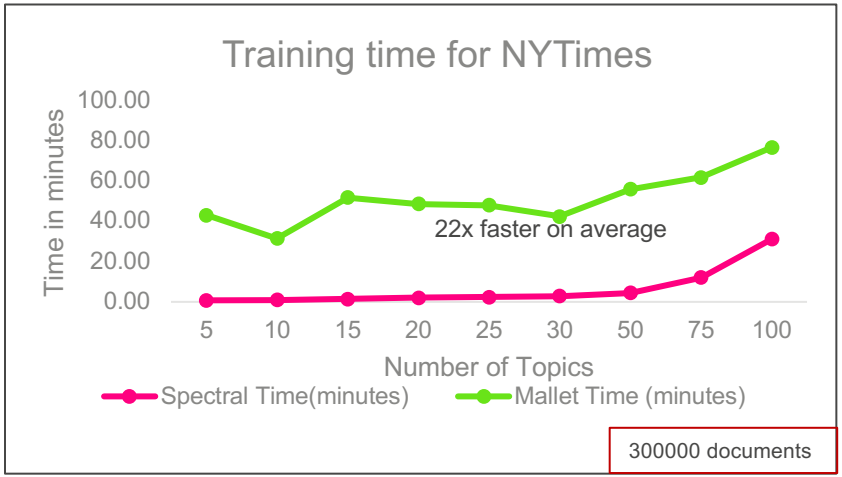
- Tensor algebra is **parallelizable** like linear algebra.
- **Faster** than other algorithms for LDA
- **Flexible:** Training and inference decoupled
- **Guaranteed** in theory to converge to global optimum

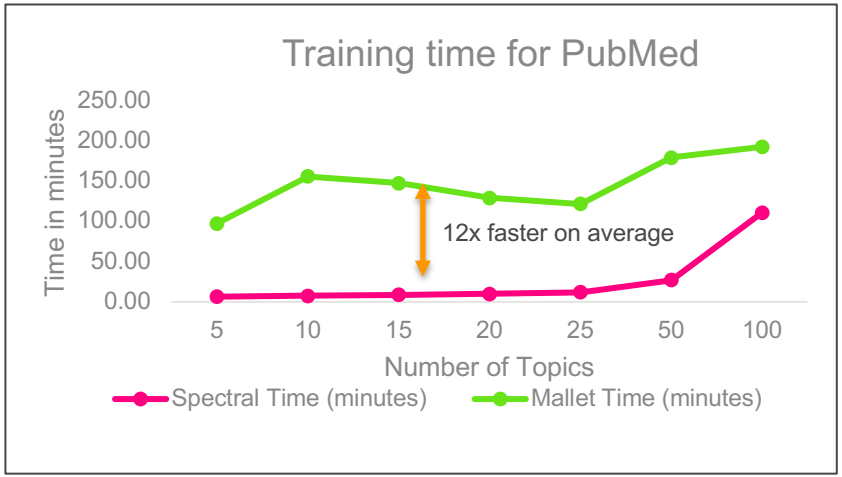
A. Anandkumar etal,Tensor Decompositions for Learning Latent Variable Models, JMLR 2014.

Spectral LDA on AWS SageMaker



SageMaker LDA training is faster

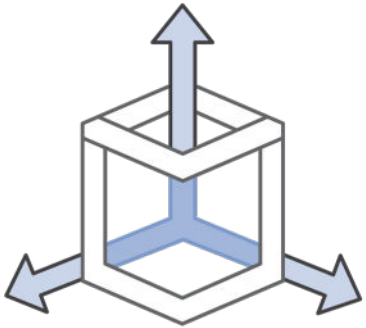




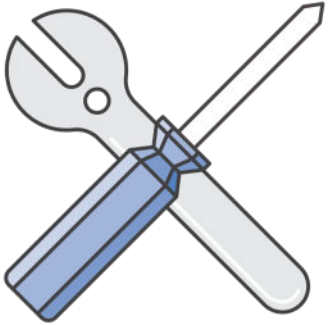
- Mallet is an open-source framework for topic modeling
- Mallet does training and inference together
- Benchmarks on AWS SageMaker Platform

Introducing Amazon SageMaker

The quickest and easiest way to get ML models from idea to production



End-to-end Machine Learning Platform

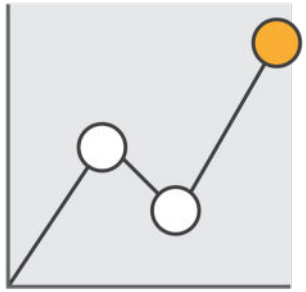


Zero setup

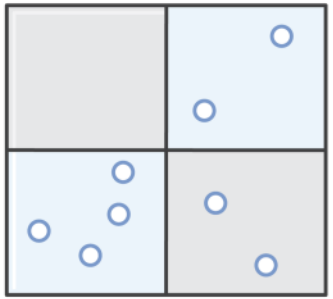
Flexible model training

Pay by the second

Many optimized algorithms on SageMaker



XGBoost, FM, and Linear for classification and regression



Kmeans and PCA for clustering and dimensionality reduction

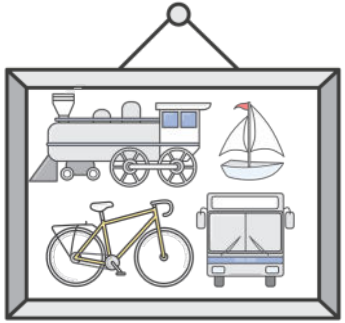
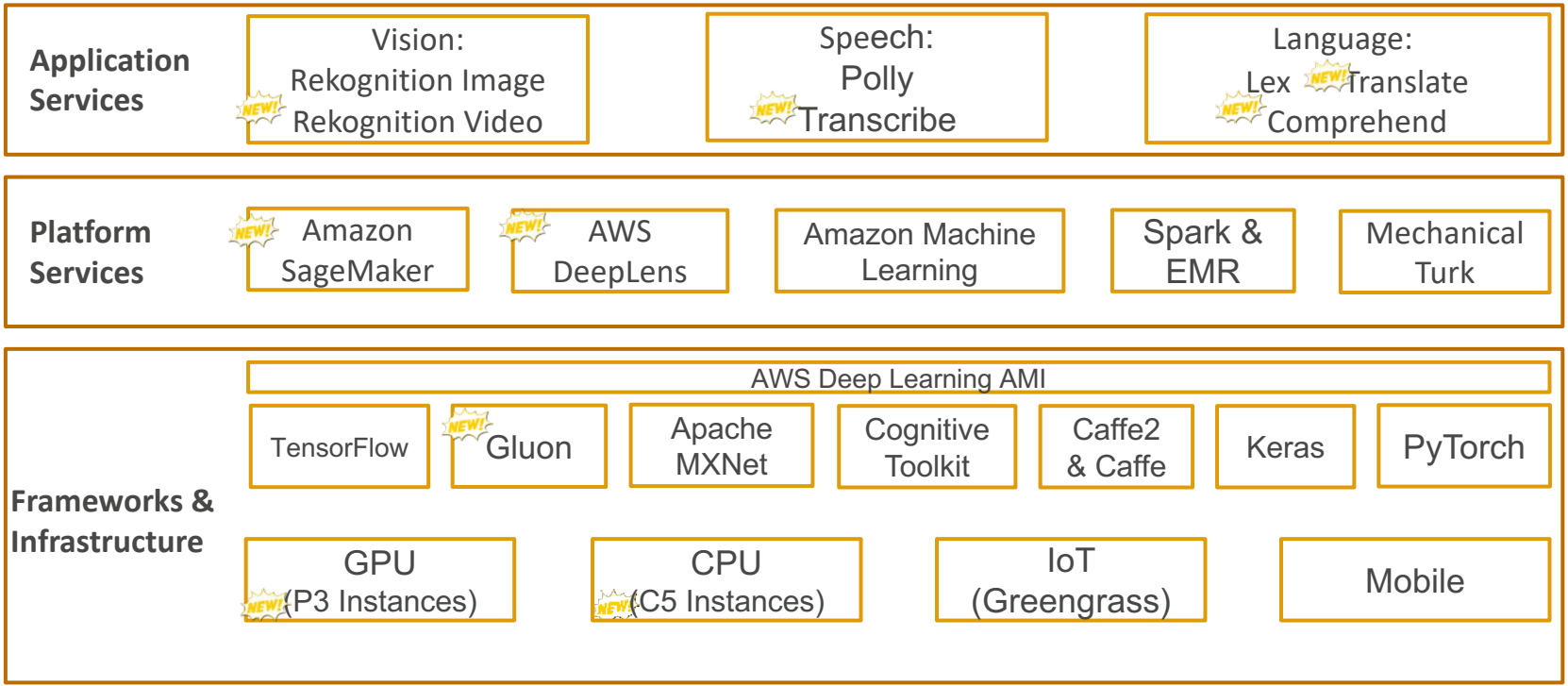


Image classification with convolutional neural networks

LDA and NTM for topic modeling, seq2seq for translation

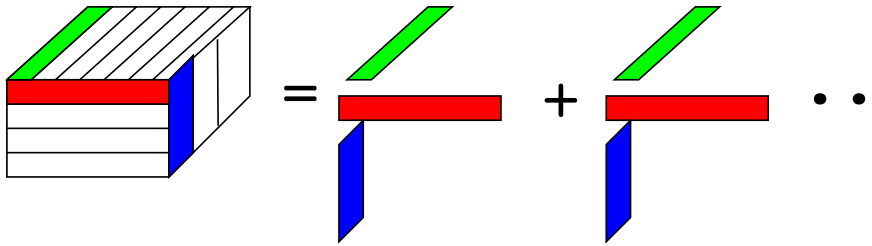
AWS ML Stack


Application Services
Vision:
Rekognition Image
Rekognition Video
Speech:
Polly
Transcribe
Language:
Lex
Translate
Comprehend
Platform Services
Amazon SageMaker
AWS DeepLens
Amazon Machine Learning
Spark & EMR
Mechanical Turk
Frameworks & Infrastructure
AWS Deep Learning AMI
TensorFlow
Gluon
Apache MXNet
Cognitive Toolkit
Caffe2 & Caffe
Keras
PyTorch
GPU
(P3 Instances)
CPU
(C5 Instances)
IoT
(Greengrass)
Mobile

CONCLUSION

Conclusion

- **Tensors** are higher extensions of matrices.
- Encode data dimensions, modalities and higher order relationships.
- Tensor algebra is richer than matrix algebra. Richer neural network architectures. More compact networks/better accuracy.
- Tensor operations are embarrassingly parallel. New primitives for tensor operations.



THANK YOU