# Course Summary

Joseph E. Gonzalez
Co-director of the RISE Lab
jegonzal@cs.berkeley.edu

# Recap:
# Reflecting on the Class

# Covered Many Topics

Problem Formulation

Neural Architectures

Deep Learning Frameworks

Reinforcement Learning

Distributed Learning

DNN Compilers

Hyperparameter Search

Secure Learning

Learned Data Structures

Autonomous Vehicles

ML for Networking

Dynamic Networks

Model Compression

Prediction Serving

Natural Language Processing

Explainability

Scheduling for Training

# Covered Many Topics

## Learning

Neural Architectures

Reinforcement Learning

Hyperparameter Search

Dynamic Networks

Model Compression

Natural Language Processing

Explainability

Problem Formulation

Deep Learning Frameworks

Distributed Learning

DNN Compilers

Secure Learning

Learned Data Structures

Autonomous Vehicles

ML for Networking

Prediction Serving

Scheduling for Training

## Systems

?

Distributed Dataflow Systems

Actors

Real-time Systems

Architecture of a Compiler

Vector Accelerators

# Big Ideas and Results

➢ Emerging trends in AI research

➢ Tradeoff between statistical and computation efficiency

➢ Role of systems (simplification, scale) in "AI Revolution"

➢ New applications of AI techniques to systems problems

➢ New problem domains (e.g., autonomous driving)

# Big Ideas in ML Research

➤ Generalization
  ➤ What is being "learned"?

➤ Inductive Biases and Representations
  ➤ What assumptions about domain enable efficient learning?

➤ Efficiency (Data and Computation)
  ➤ How much data and time are needed to learn?

➤ Details: Objectives/Models/Algorithms

# Big Ideas in Systems Research

➤ Problem Framing
  ➤ Identifying the right problem and solution requirements

➤ Abstraction & Managing Complexity
  ➤ Reducing complex problems into smaller parts

➤ Tradeoffs
  ➤ Understanding fundamental constraints

➤ Details: System design and Implementation

# What is AI-Systems Research?

➤ Should be both good AI and Systems research
  ➤ Provides insights to both communities

➤ Leverages understanding of both domains
  ➤ Examples:
    ➤ Studies tradeoff between statistical and computational efficiency
    ➤ Identify essential abstractions in DNN design
    ➤ Leverages framing of indexes to exploit overfitting

➤ Do we need another venue?
  ➤ ICML, NIPS, ICLR, UAI, AAAI, ICRA, CVPR, ICCV, ECCV, SOSP, OSDI, NSDI, EuroSys, SIGMOD, VLDB, ASPLOS, SOCC ➔ SysML?

# Logistics

➢ **Poster Session** next Wednesday 5/8/19 from 9:00 to 11:00

➢ Final Reports due: Monday 5/13/19 at 11:59PM
  ➢ 8 pages in Google Docs
  ➢ Email link to jegonzal@berkeley.edu and istoica@berkeley.edu