Security and Fairness of Deep Learning

Course Overview

Anupam Datta
CMU

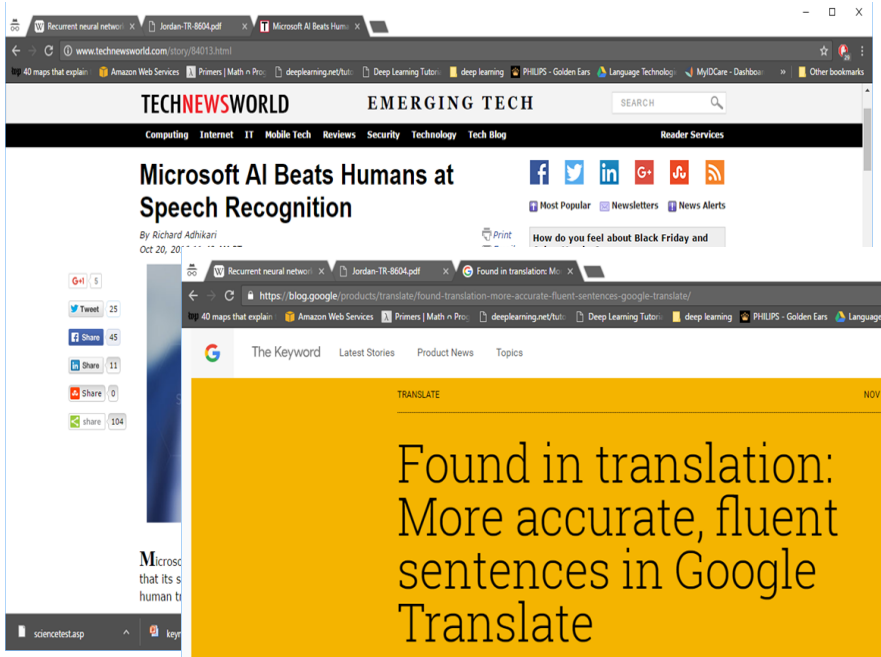
Spring 2018

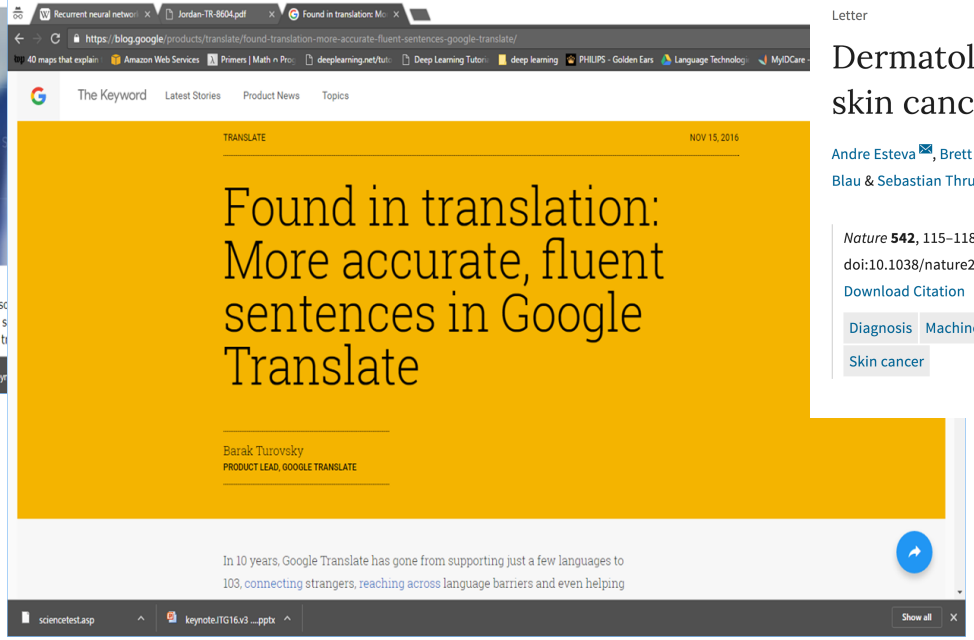
Course staff

- Instructor: Anupam Datta
  - Office: B23 221
  - Email: danupam@cmu.edu
  - Office hours: Thur 12-1pm Pacific
  - Google hangouts: link on Piazza

- TA: Caleb Lu
  - Office: Bldg 19 1031
  - Email: caleb.lu@sv.cmu.edu
  - Office hours: Mon 10am-12pm Pacific
  - Google hangouts: link on Piazza

Recent successes of deep learning





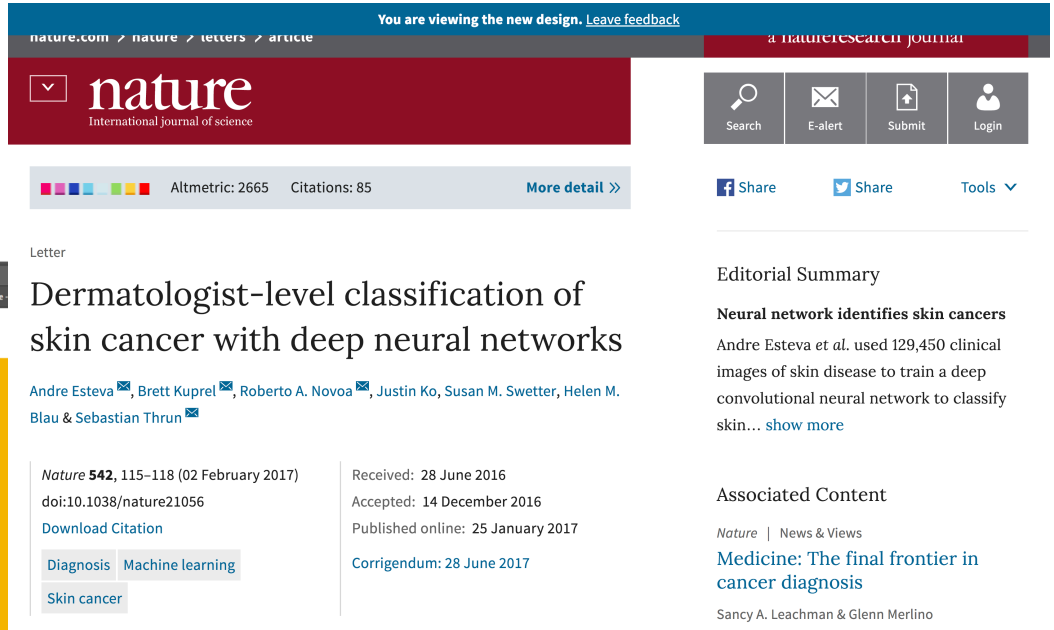
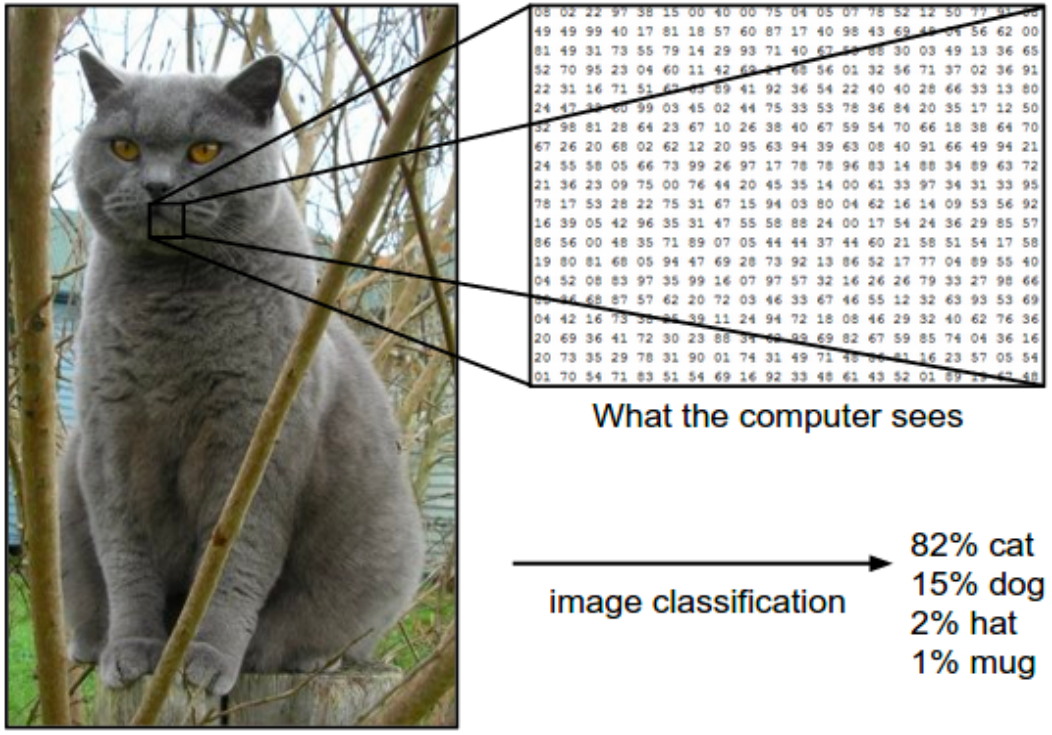
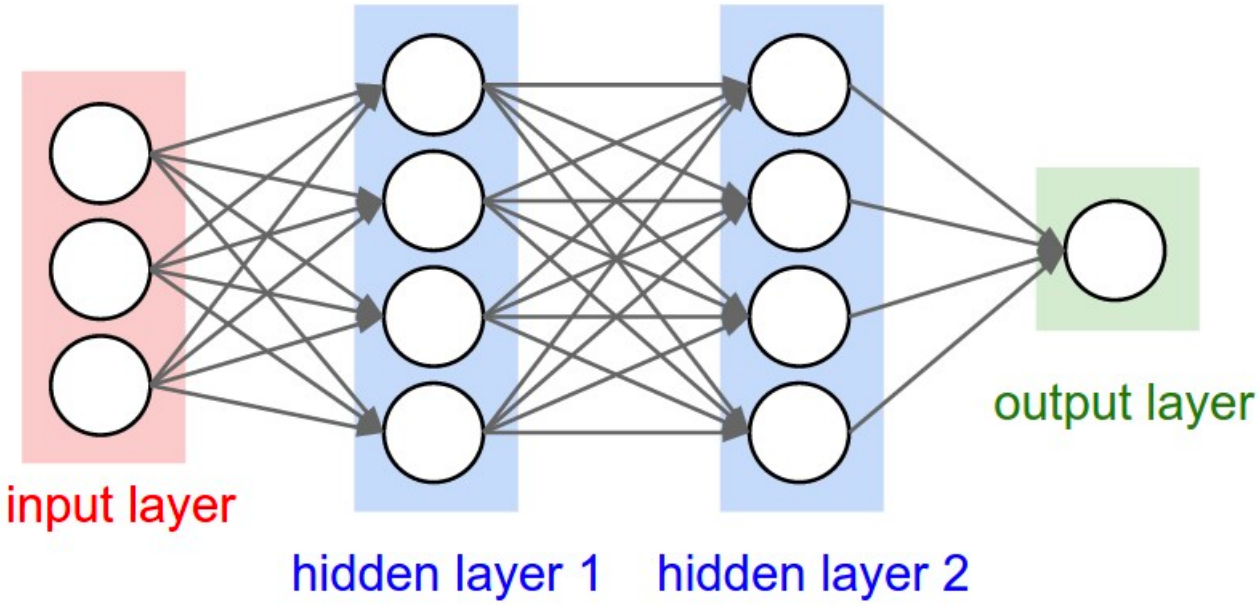


Image classification



Deep neural networks learn representations



Deeper layers learn progressively more abstract representations:
pixels, edges, motifs, parts of objects, objects

Enabling trends

- Large volumes of training data
- Computation power
  - GPUs,…

Course objective

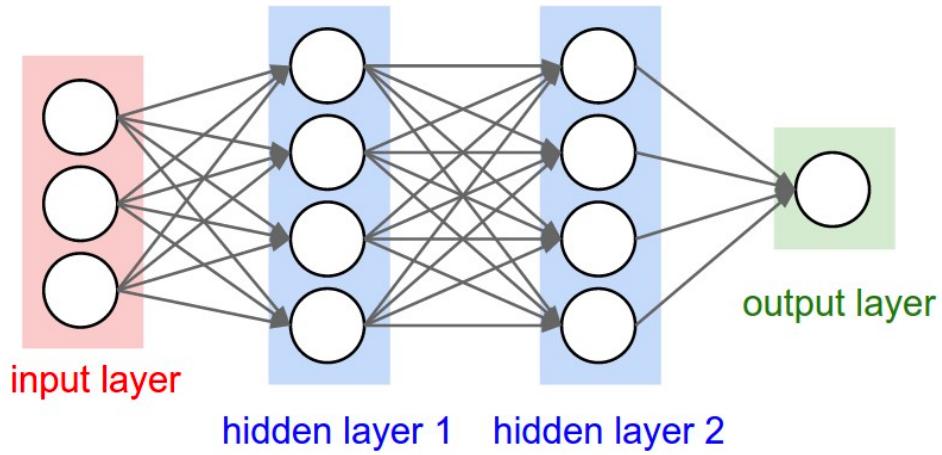
Understand deeply how and why deep networks work and their weaknesses

Course modules

1. Fundamentals of deep networks
2. Unlocking the black box
3. Security of deep learning models
4. Fairness of deep learning

Course modules

1. Fundamentals of deep networks
   - Background on machine learning
   - Architectures, training, platforms
   - Focus on convolutional and recurrent neural networks



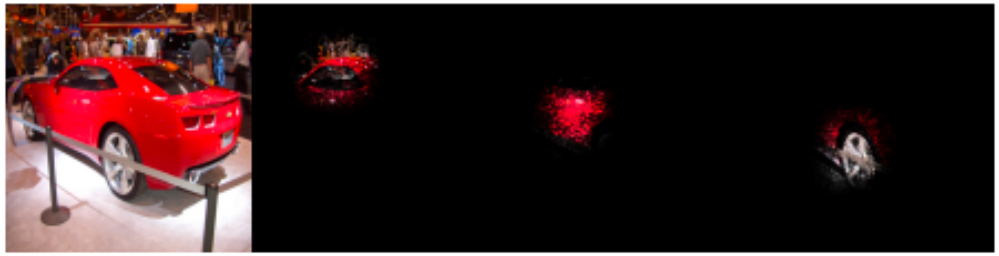
Keras

theano

NumPy

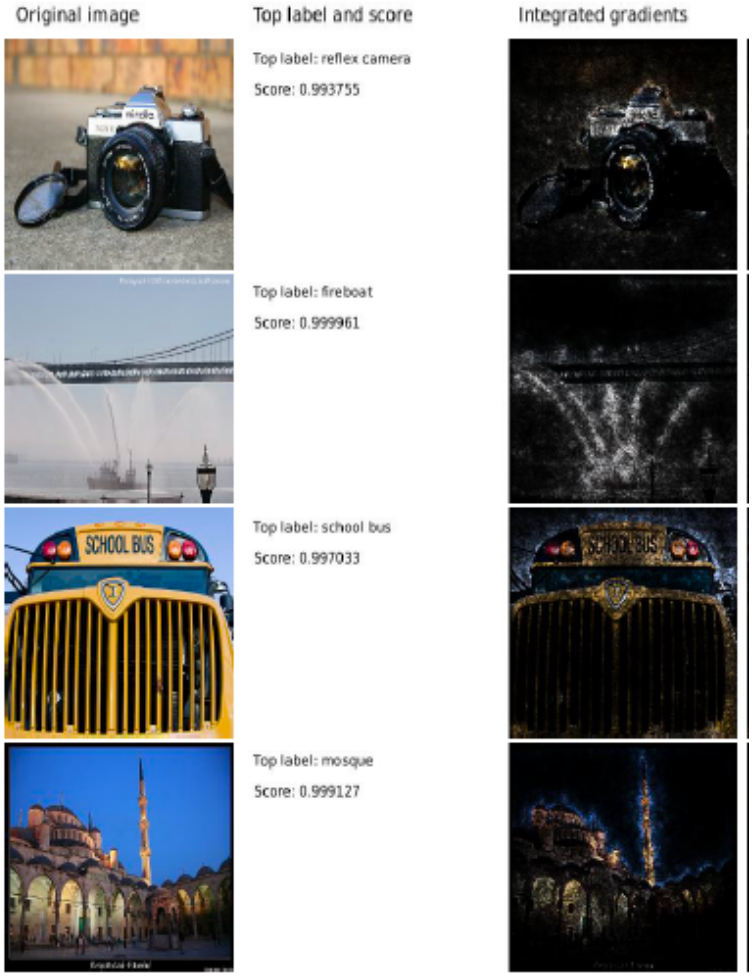
Course modules

2. Unlocking the black box
   - Explaining behavior of deep neural networks



Keras

theano



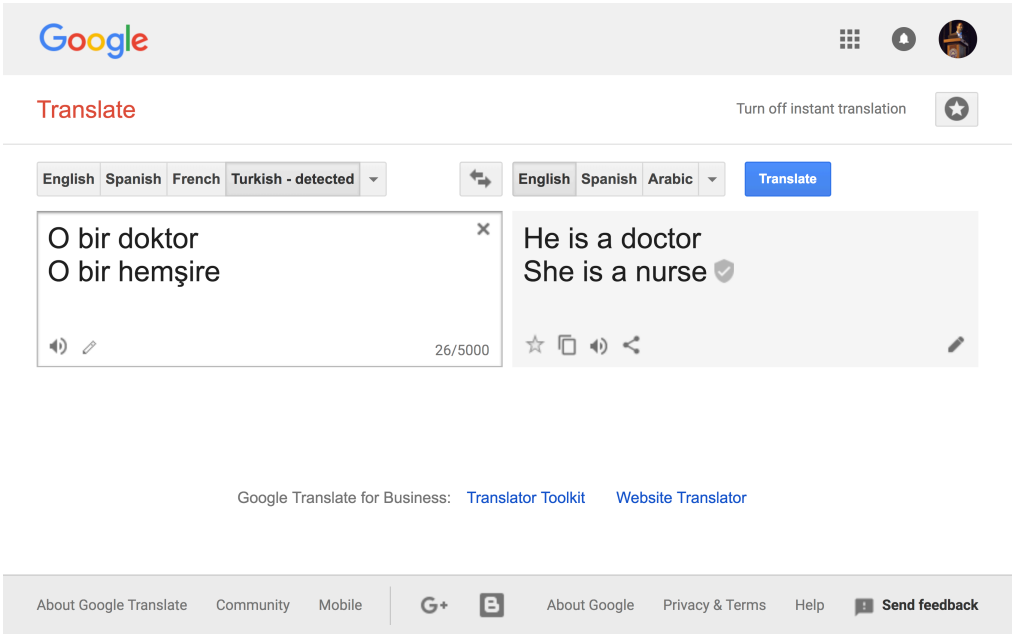
Course modules

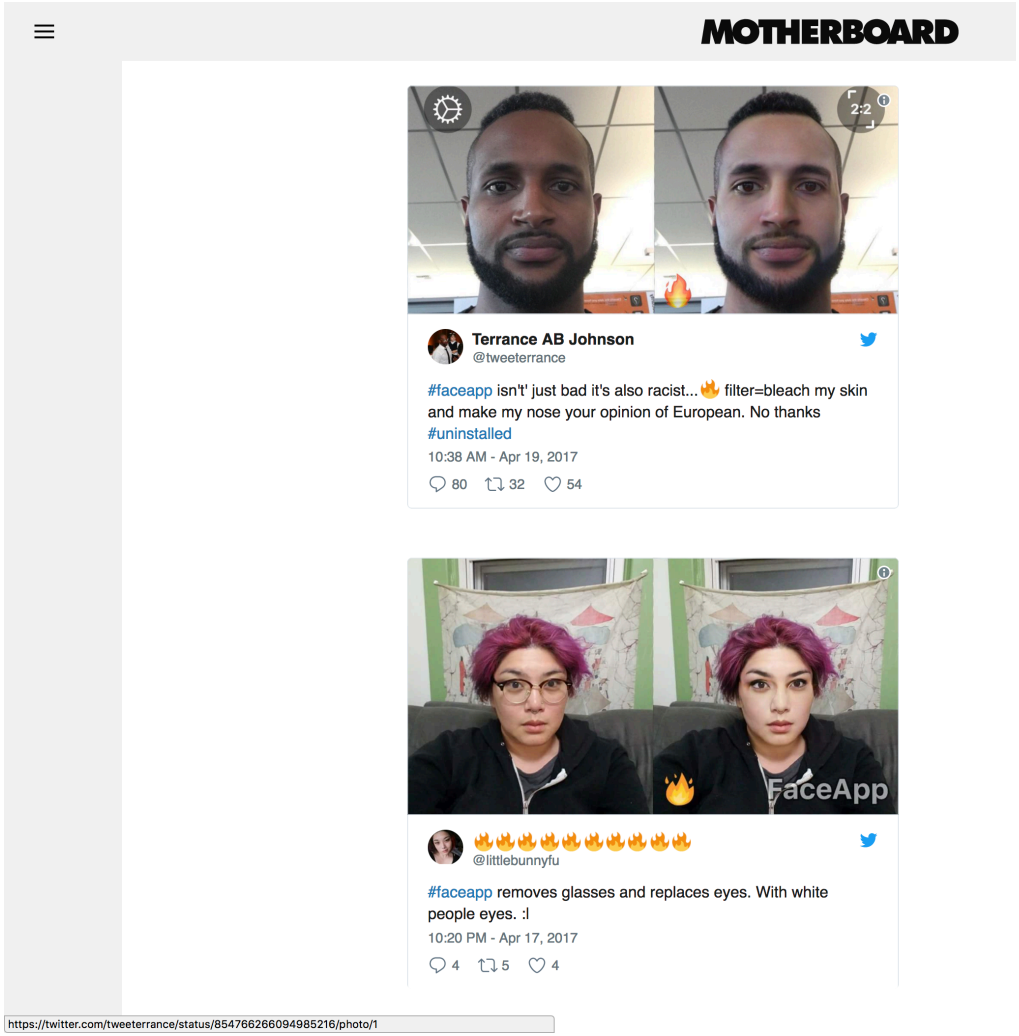
3. Security of deep learning models
   - Attacks on classifiers and defenses

TensorFlow

Course modules

4. Fairness of deep learning
   - Bias and de-biasing





Prerequisites

- No formal prerequisites
- Basics of linear algebra, probability, multivariate calculus
  - Will review briefly in class and provide resources to learn on your own
  - Roughly Chapters 1-5 of Deep Learning textbook by Goodfellow et al.
- Familiarity with Python
  - Necessary for programming homework
- Quick class poll

Logistics

- Lectures: Tue & Thur, 10:30-11:50am Pacific
- Web page: http://www.ece.cmu.edu/~ece739/
- Temp web page: http://www.andrew.cmu.edu/user/kaijil/class-18739/

- Canvas (for grades, homework)
- Piazza (for all other communication)
  - Please enroll; you should have received invitation

- Textbook
  - Deep Learning textbook by Goodfellow, Bengio, Courville

Grading

- Homework: 80%
  - 4 x 20%
- Paper summaries: 10%
  - 5 x 2%
- Class participation: 10%
  - Be present and engaged in class and piazza

Collaboration policy on homework

- You are allowed to discuss homework problems with other students in the class, but are required to write out solutions independently and to acknowledge any collaboration or other source.

CMU Computing Policy

CMU Policy on Cheating