CS395T - DEEP LEARNING SEMINAR

Philipp Krähenbühl

OVERVIEW

[Philipp Krähenbühl](#)
office [GDC 4.824](#)
office hours by appointment (send email)

TA Huihuang Zheng
office [GDC 6.802](#)
office hours We 10-11am

Try piazza first!

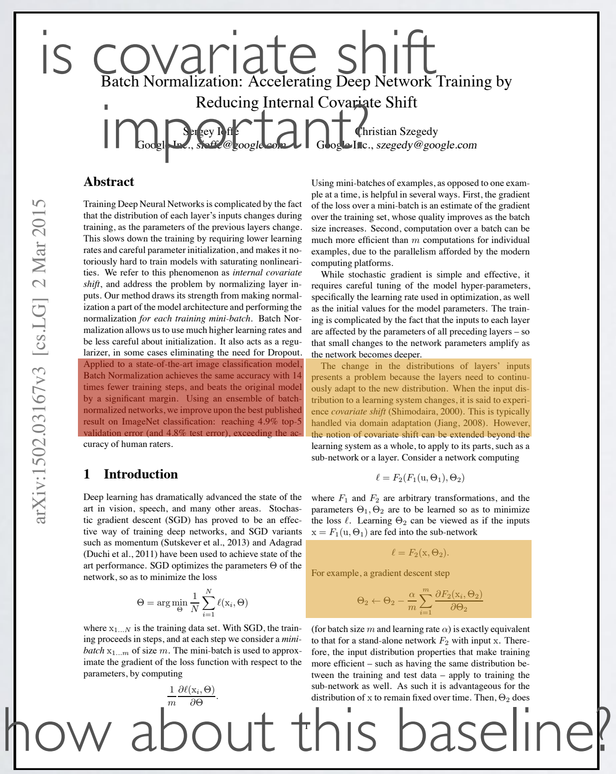
OVERVIEW

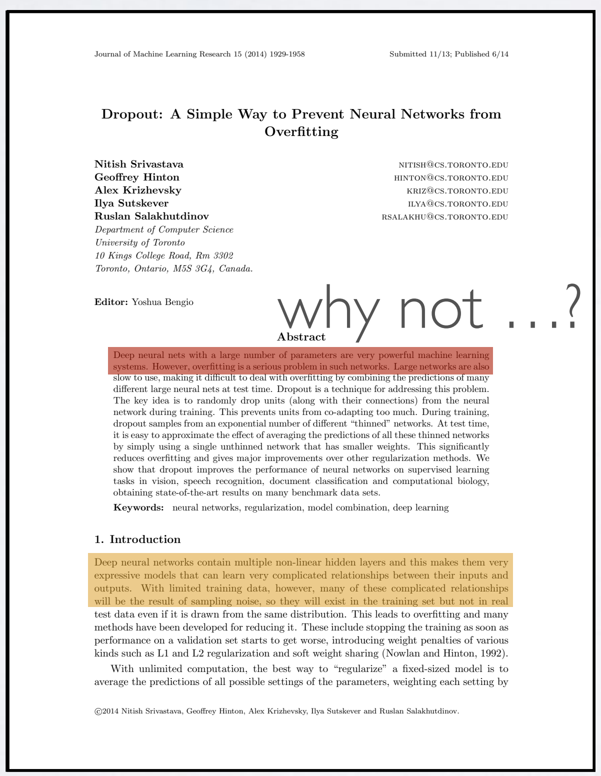
https://philkr.github.io/CS395T/

OVERVIEW

before class

Read and review two papers





in class

10 min intro (Philipp)

10-15 min paper 1 (one of you)

25 min paper 1 (all of you)

10-15 min paper 2 (one of you)

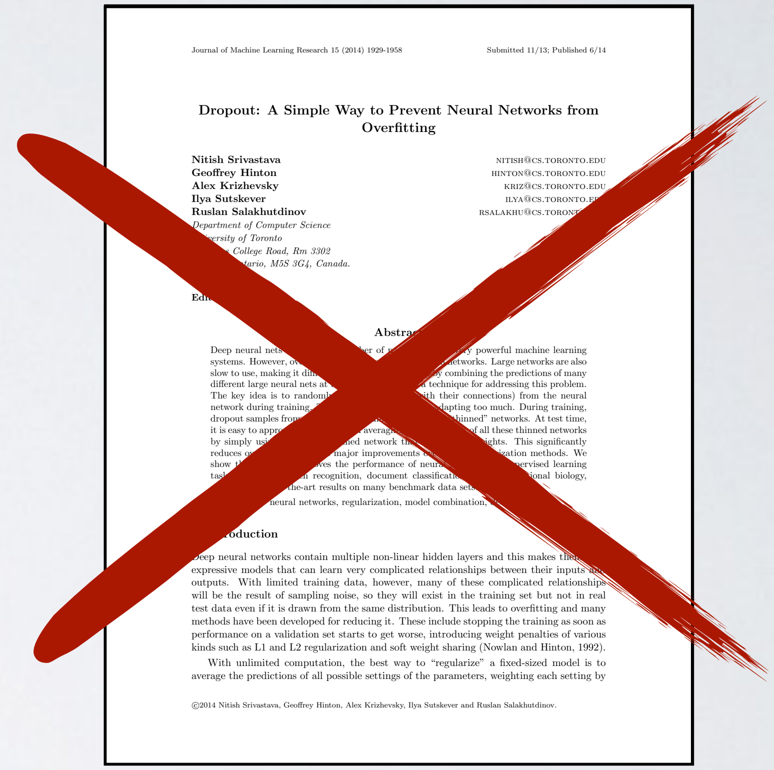
25 min paper 2 (all of you)

PRESENTATIONS

visual

~~no incredibly long walls of text or math than nobody can understand, follow or otherwise parse.~~

show presentation to Philipp 1 week ahead



Journal of Machine Learning Research 15 (2014) 1929-1958 Submitted 11/13; Published 6/14

Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Nitish Srivastava NITISH@CS.TORONTO.EDU
Geoffrey Hinton HINTON@CS.TORONTO.EDU
Alex Krizhevsky KRIZ@CS.TORONTO.EDU
Ilya Sutskever ILYA@CS.TORONTO.ED
Ruslan Salakhutdinov RSALAKHU@CS.TORONT
Department of Computer Science
…versity of Toronto
… College Road, Rm 3302
…tario, M5S 3G4, Canada.

Edi…

Abstrac…

Deep neural nets … ber of … y powerful machine learning systems. However, ov… networks. Large networks are also slow to use, making it di… by combining the predictions of many different large neural nets at … a technique for addressing this problem. The key idea is to random… with their connections) from the neural network during training. … dapting too much. During training, dropout samples from … thinned" networks. At test time, it is easy to appr… averagi… of all these thinned networks by simply usi… ned network th… ights. This significantly reduces o… major improvements … ization methods. We show t… ves the performance of neura… pervised learning tas… n recognition, document classificatio… ional biology, … the-art results on many benchmark data sets…

… neural networks, regularization, model combination, …

…roduction

Deep neural networks contain multiple non-linear hidden layers and this makes the… expressive models that can learn very complicated relationships between their inputs an… outputs. With limited training data, however, many of these complicated relationships will be the result of sampling noise, so they will exist in the training set but not in real test data even if it is drawn from the same distribution. This leads to overfitting and many methods have been developed for reducing it. These include stopping the training as soon as performance on a validation set starts to get worse, introducing weight penalties of various kinds such as L1 and L2 regularization and soft weight sharing (Nowlan and Hinton, 1992).

With unlimited computation, the best way to "regularize" a fixed-sized model is to average the predictions of all possible settings of the parameters, weighting each setting by

©2014 Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov.

no scrolling through paper

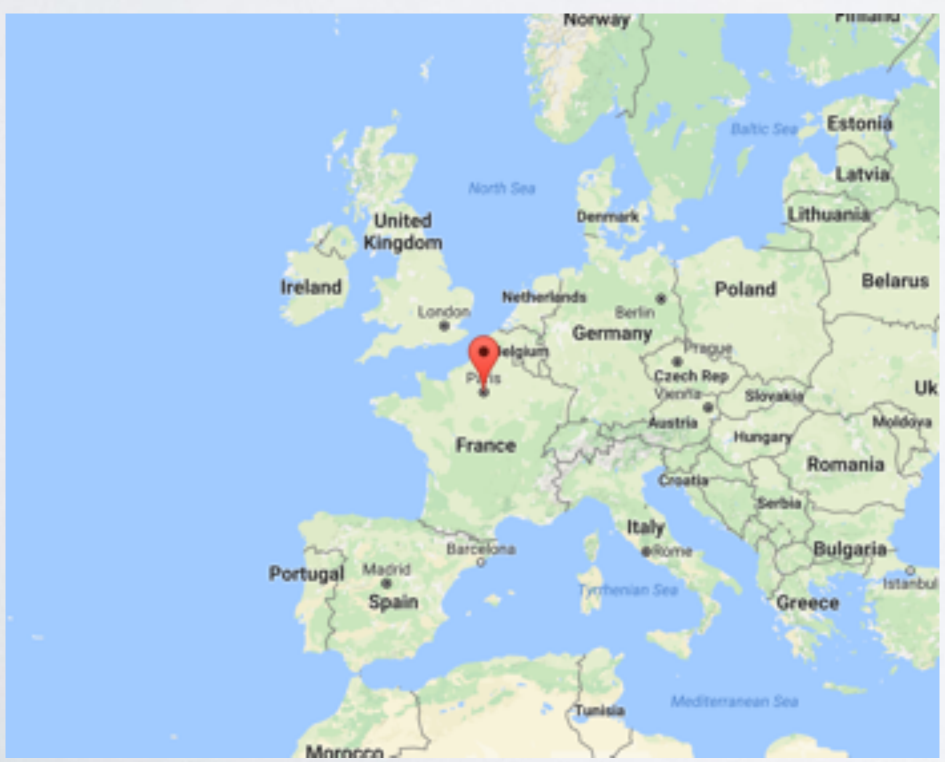
OVERVIEW

- two deep learning projects
  - train a deep network
  - write a latex report
  - present your work

PROJECTS

- two deep learning projects
  - Teams up to 3
  - **GPU access**
  - Python preferred
    - Caffe, TF or Theano

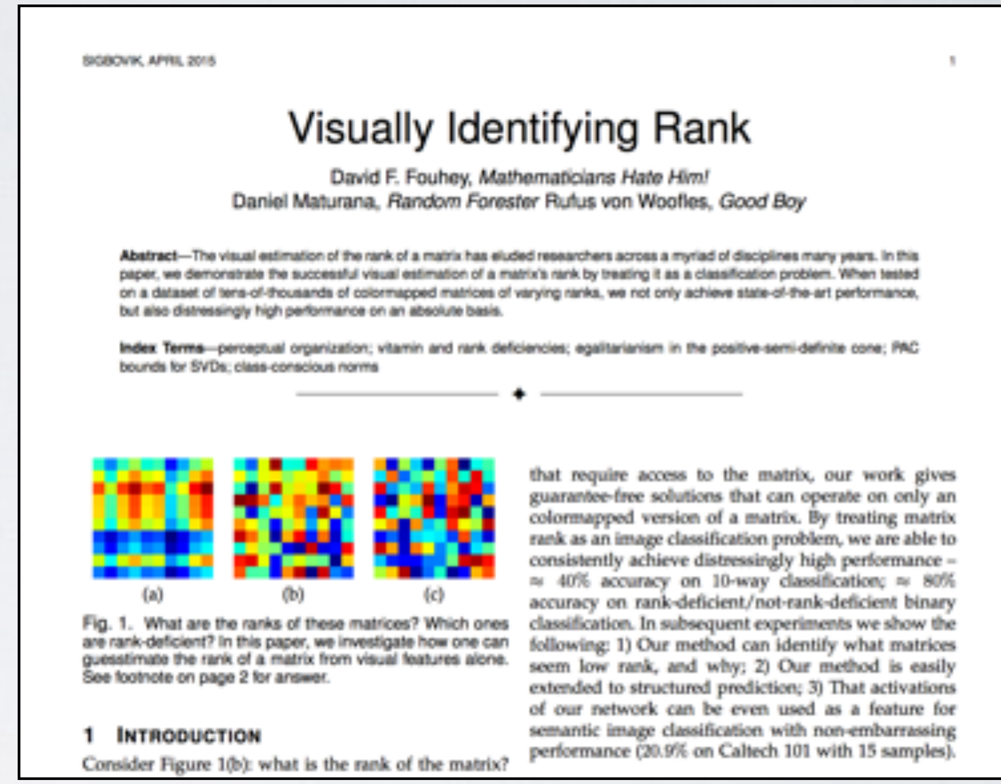
PROJECT 1



1960s

PROJECT 2

- Open ended
  - can be your research
  - **not** published by Dec 31
  - level of a top tier workshop publication
    - CVPR, ICCV, ICML, NIPS, ACL, SIGGRAPH
    - SIGBOVIK


SIGBOVIK, APRIL 2015 1

Visually Identifying Rank

David F. Fouhey, *Mathematicians Hate Him!*
Daniel Maturana, *Random Forester* Rufus von Woofles, *Good Boy*

Abstract—The visual estimation of the rank of a matrix has eluded researchers across a myriad of disciplines many years. In this paper, we demonstrate the successful visual estimation of a matrix's rank by treating it as a classification problem. When tested on a dataset of tens-of-thousands of colormapped matrices of varying ranks, we not only achieve state-of-the-art performance, but also distressingly high performance on an absolute basis.

Index Terms—perceptual organization; vitamin and rank deficiencies; egalitarianism in the positive-semi-definite cone; PAC bounds for SVDs; class-conscious norms

(a) (b) (c)

Fig. 1. What are the ranks of these matrices? Which ones are rank-deficient? In this paper, we investigate how one can guesstimate the rank of a matrix from visual features alone. See footnote on page 2 for answer.

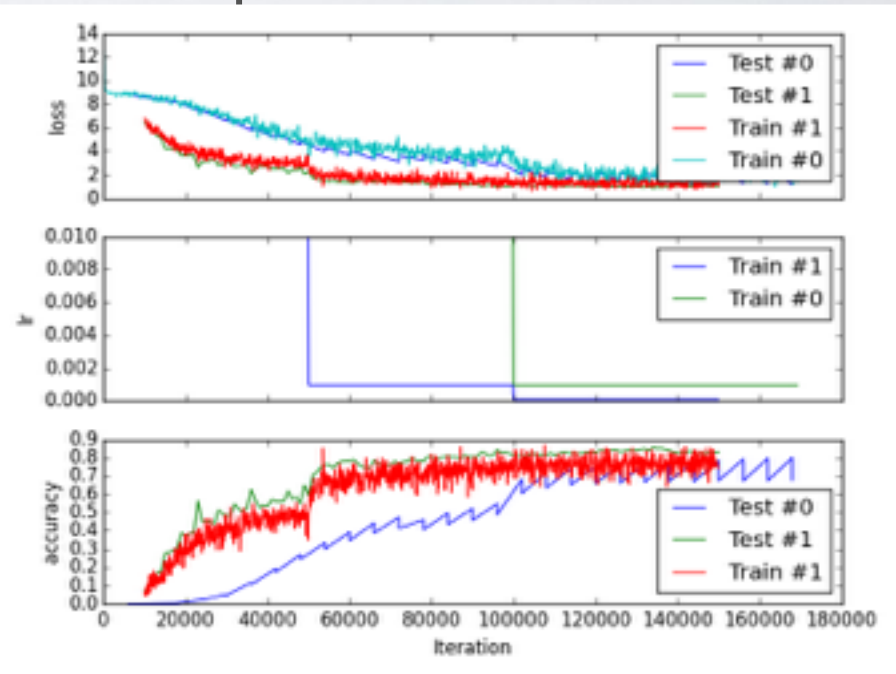
1 INTRODUCTION

Consider Figure 1(b): what is the rank of the matrix?

that require access to the matrix, our work gives guarantee-free solutions that can operate on only an colormapped version of a matrix. By treating matrix rank as an image classification problem, we are able to consistently achieve distressingly high performance – ≈ 40% accuracy on 10-way classification; ≈ 80% accuracy on rank-deficient/not-rank-deficient binary classification. In subsequent experiments we show the following: 1) Our method can identify what matrices seem low rank, and why; 2) Our method is easily extended to structured prediction; 3) That activations of our network can be even used as a feature for semantic image classification with non-embarrassing performance (20.9% on Caltech 101 with 15 samples).

PREREQUISITES

- 391L - Intro Machine learning (or equivalent)
- 311 or 311H - Discrete math for computer science (or equivalent)
- Proficiency in Python
- Basic **deep learning background**



GOALS

- Review a deep learning paper
- Give an interesting DL presentation
- Devise and execute a DL project

GRADES

- 30% paper presentation
- 30% project 1 (10% presentation, 20% project)
- 40% project 2 (10% presentation, 30% project)
- (optional) 12.5% volunteering for second presentation

GRADES

```python
def grade(p):
  from math import floor
  if p < 50: return 'F'
  v = (100-p) * 4 / (50 + 1e-5)
  return chr(ord('A')+floor(v)) +
           ['+','','','-'][floor((v-floor(v))*4)]
```

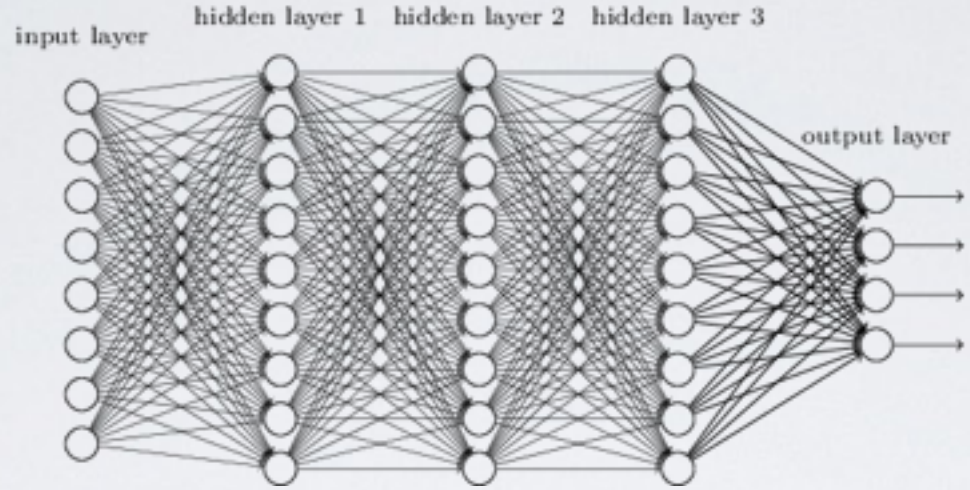
we might train a deep network to grade instead

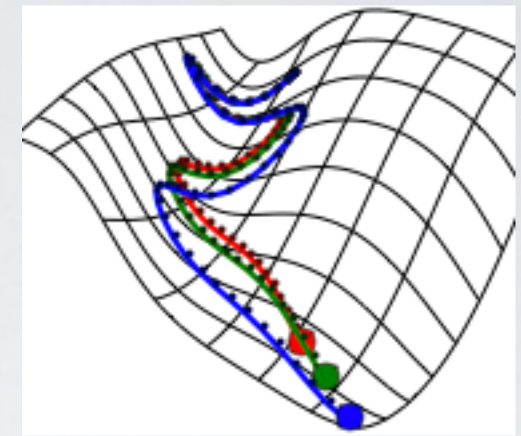
AUDITING

- Allowed
  - No homework or presentation required
  - Paper review and discussion required

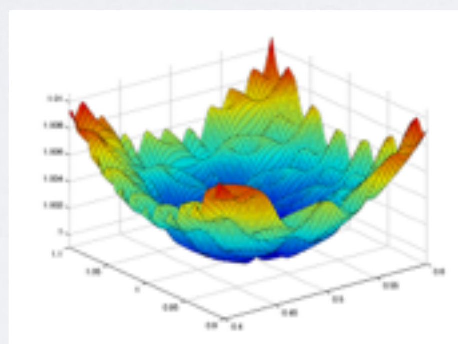
TOPICS

week 2

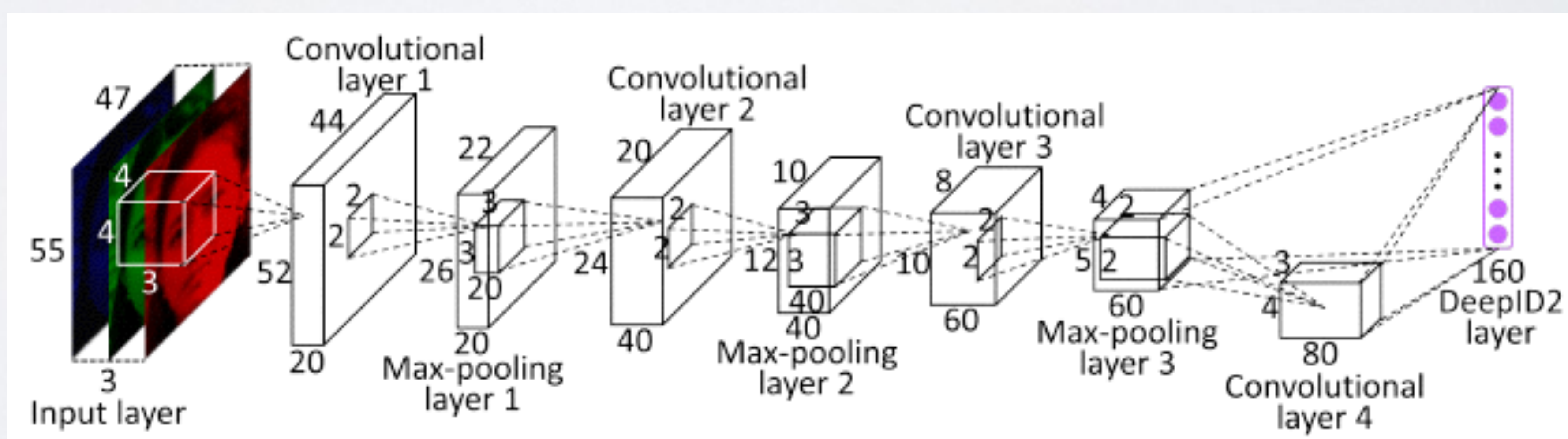




week 3

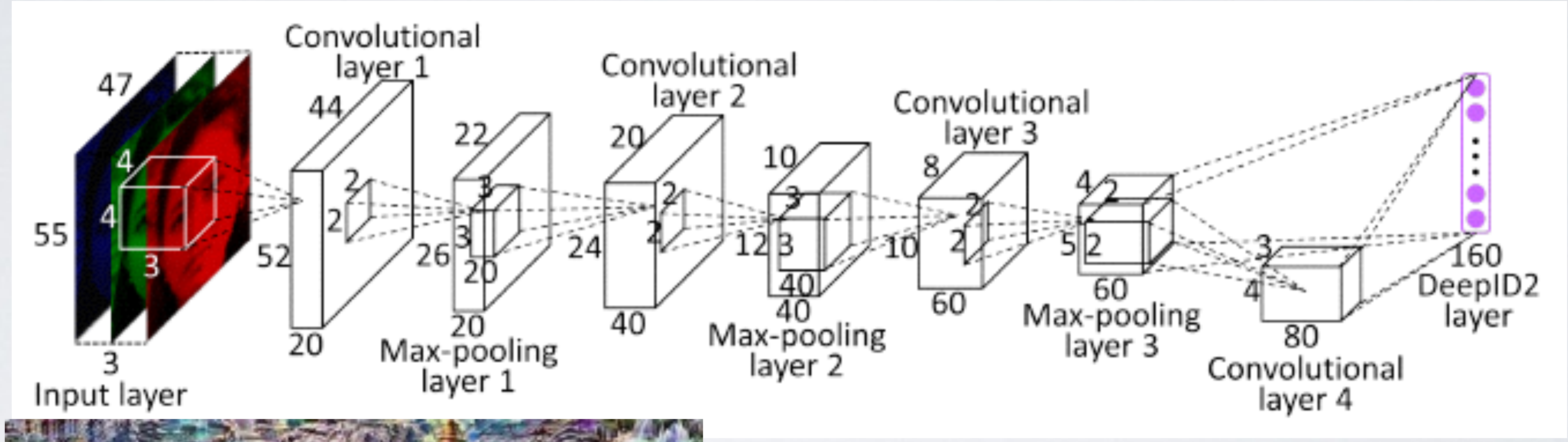


week 4



TOPICS

week 5

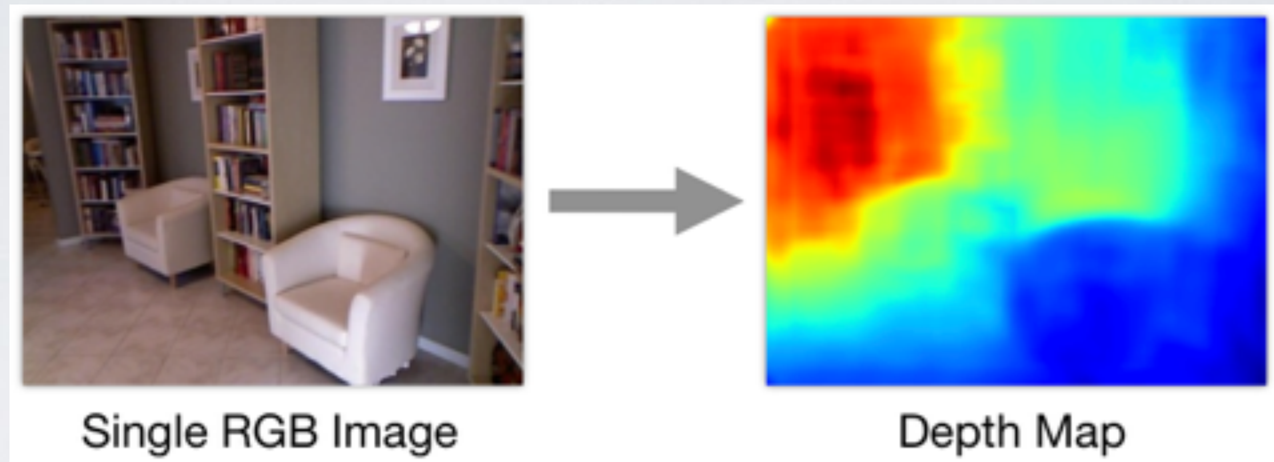


week 6

week 7 Project 1 presentations

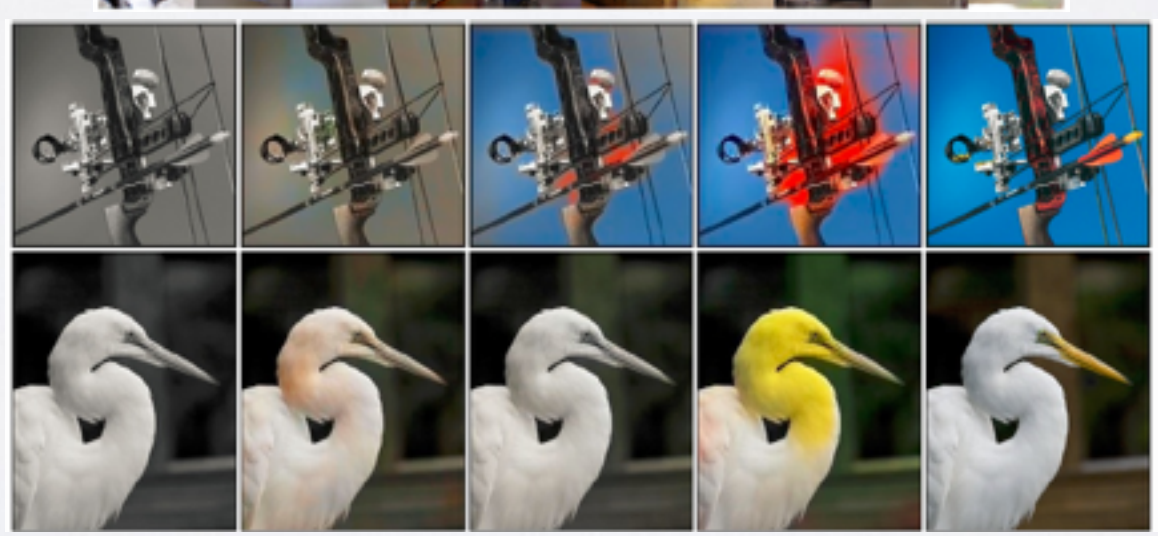
TOPICS

week 8



week 9

week 10

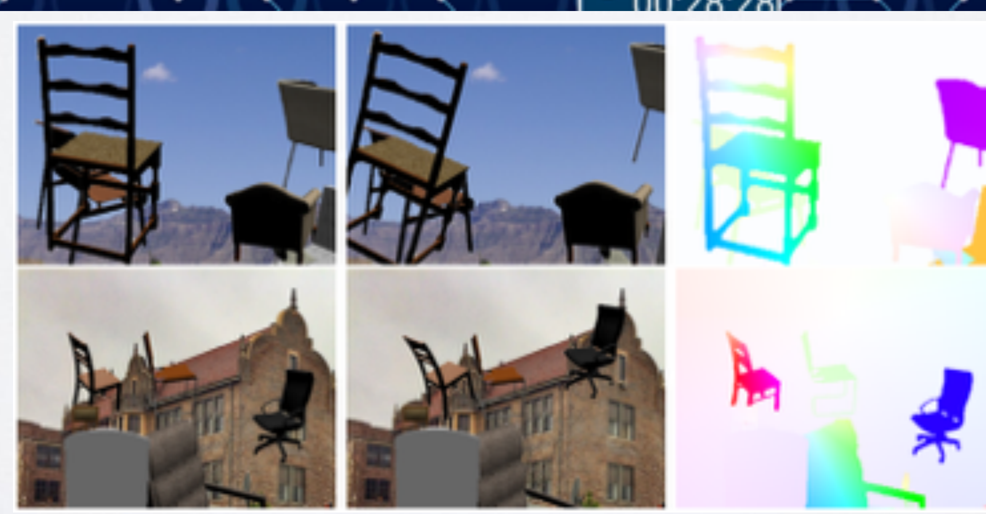


TOPICS

week 11

week 12

week 13



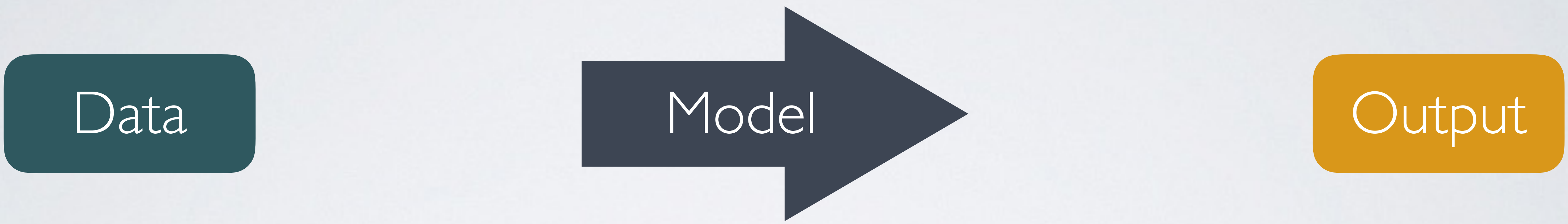
TOPICS

week 14	TPD
week 15	Project 2 presentations

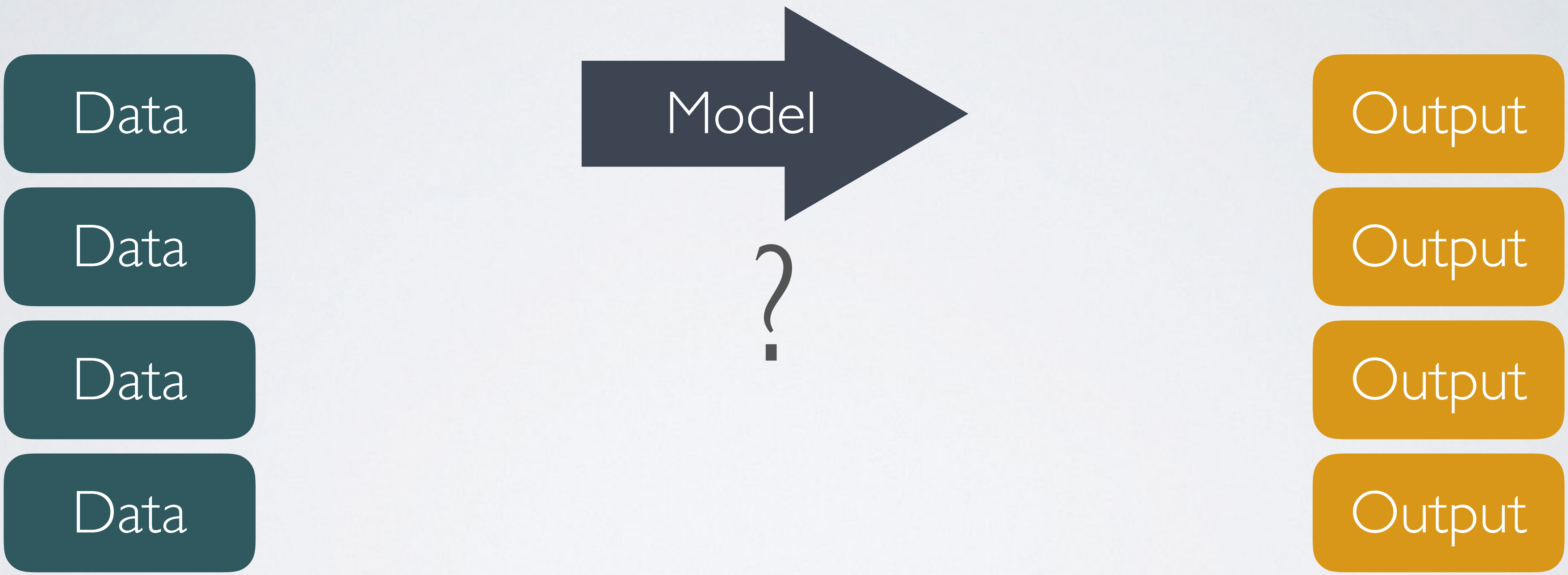
THE N-WORD

- Neural
  - try to keep Neuroscience out of this class
  - try to motivate through optimization and ML
    - instead of biology

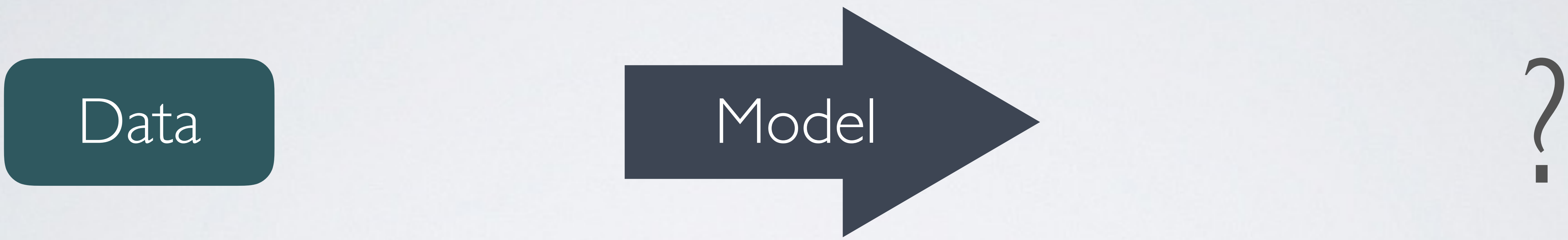
REVIEW



TRAINING



TESTING / INFERENCE



DATA

feature

Data

$f_1 = \{a,b,c,\ldots\}$

$f_2 = \{d,e,f,\ldots\}$

…

EXAMPLE: GRADING

Data

feature

	score Project 1	score Project 2	age	height
f_1	51	44	23	182
f_2	25	80	26	172

EXAMPLE: GRADING

f_1	51	44	23	182
f_2	25	80	26	172

Model

A+

A-

continuous and (differentiable)

EXAMPLE: GRADING

f_1 51 44 23 182 Model

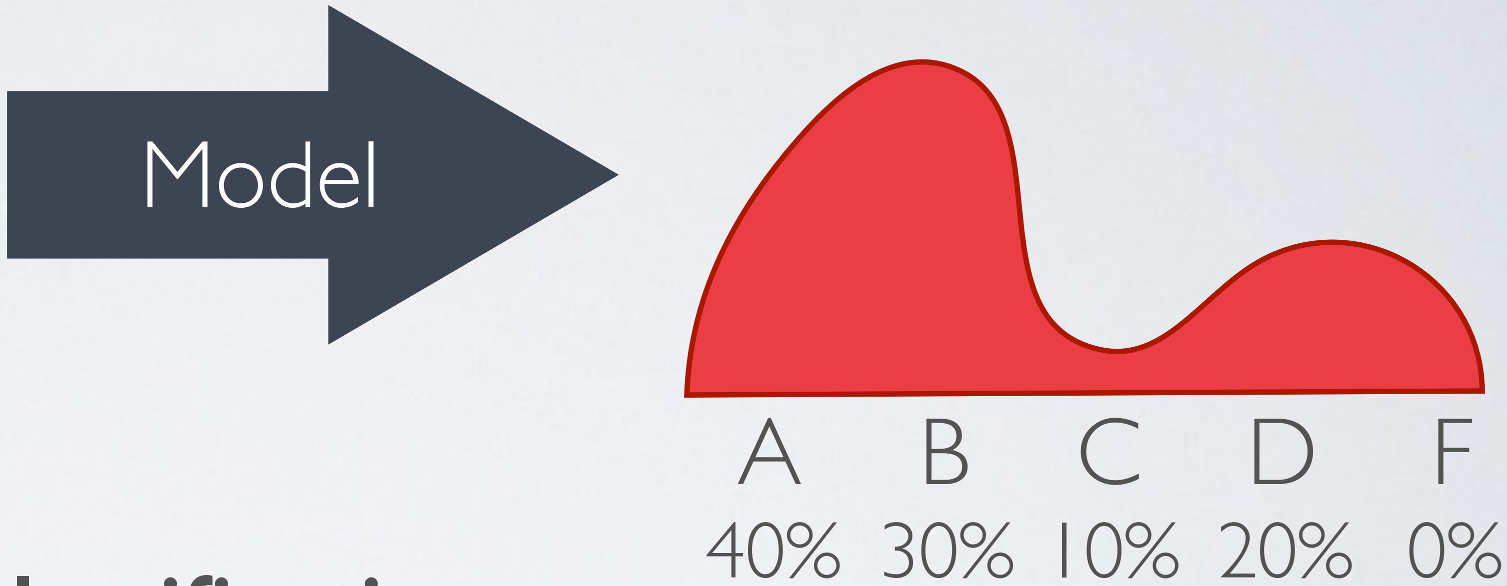
A B C D F

regression:
predict continuous value and round

linear regression
$g = A f + b$

EXAMPLE: GRADING

f_1 51 44 23 182



classification:
predict continuous distribution over grades

logistic regression
$P_g = \text{softmax}(A\, f + b)$

$\text{softmax}(x) = \exp(x) / (\boldsymbol{\Sigma} \exp(x))$

TRAINING

f_1	51	44	23	182	Model	A+
f_2	25	80	26	172		A-

?

EXAMPLE: GRADING

regression

$$\underset{A,b}{\text{minimize}} \ \Sigma_i (g_i - A\, f_i - b)^2$$

classification

$$\underset{A,b}{\text{maximize}} \ \Sigma_i \log P(g_i)$$

FAIRNESS

$f_1 > f_2 \quad \Rightarrow \quad g_1 \geq g_2$

EXAMPLE: GRADING

regression

$$\min_{A,b} \Sigma_i (g_i - A f_i - b)^2 + \Sigma_{i,j:f_i<f_j} \max(Af_i - Af_j, 0)$$

classification

$$\max_{A,b} \Sigma_i \log P(g_i) + \Sigma_{i,j:f_i<f_j} \max(Af_i - Af_j, 0)$$

EXAMPLE: GRADING

regression

$$\underset{A,b}{\text{minimize}}\ (g_i - A\, f_i - b)^2$$

$$A \geq 0$$

classification

$$\underset{A,b}{\text{maximize}}\ \log P(g_i)$$

$$A_0 \geq A_1$$

NEXT CLASS

- Look at list of papers
  - Send Huihuang your top picks per email
    - instructions on Piazza