Deep Learning

Advanced Machine Learning for NLP
Jordan Boyd-Graber
INTRODUCTION
Deep Learning was once known as “Neural Networks”
But it came back . . .

• More data
• Better tricks
  (regularization)
• Faster computers
And companies are investing …

Google Hires Brains that Helped Supercharge Machine Learning

BY ROBERT MCMILLAN  03.13.13  |  6:30 AM  |  PERMALINK
And companies are investing ...
And companies are investing ...
Map inputs to output

$$h_{w,b}(x) = \sum_i W_i x_i + b$$

Activation function:
$$f(z) = 1 + \exp(-z)$$
Map inputs to output

Input

Vector $x_1 \ldots x_d$

inputs encoded as
real numbers
Map inputs to output

Input
Vector $x_1 \ldots x_d$

Output
 multiply inputs by weights

$$f\left(\sum_i W_i x_i + b\right)$$
Map inputs to output

\[ h_{w,b}(x) = f \left( \sum_i W_i x_i + b \right) \]

Input
Vector \( x_1 \ldots x_d \)

Output

add bias
Map inputs to output

Input
Vector $x_1 \ldots x_d$

Output
$$f\left(\sum_i W_i x_i + b\right)$$

Activation
$$f(z) \equiv \frac{1}{1 + \exp(-z)}$$

pass through nonlinear sigmoid
Why is it called activation?
In the shallow end

- This is still logistic regression
- Engineering features $x$ is difficult (and requires expertise)
- Can we learn how to represent inputs into final decision?