Multivariate Regression

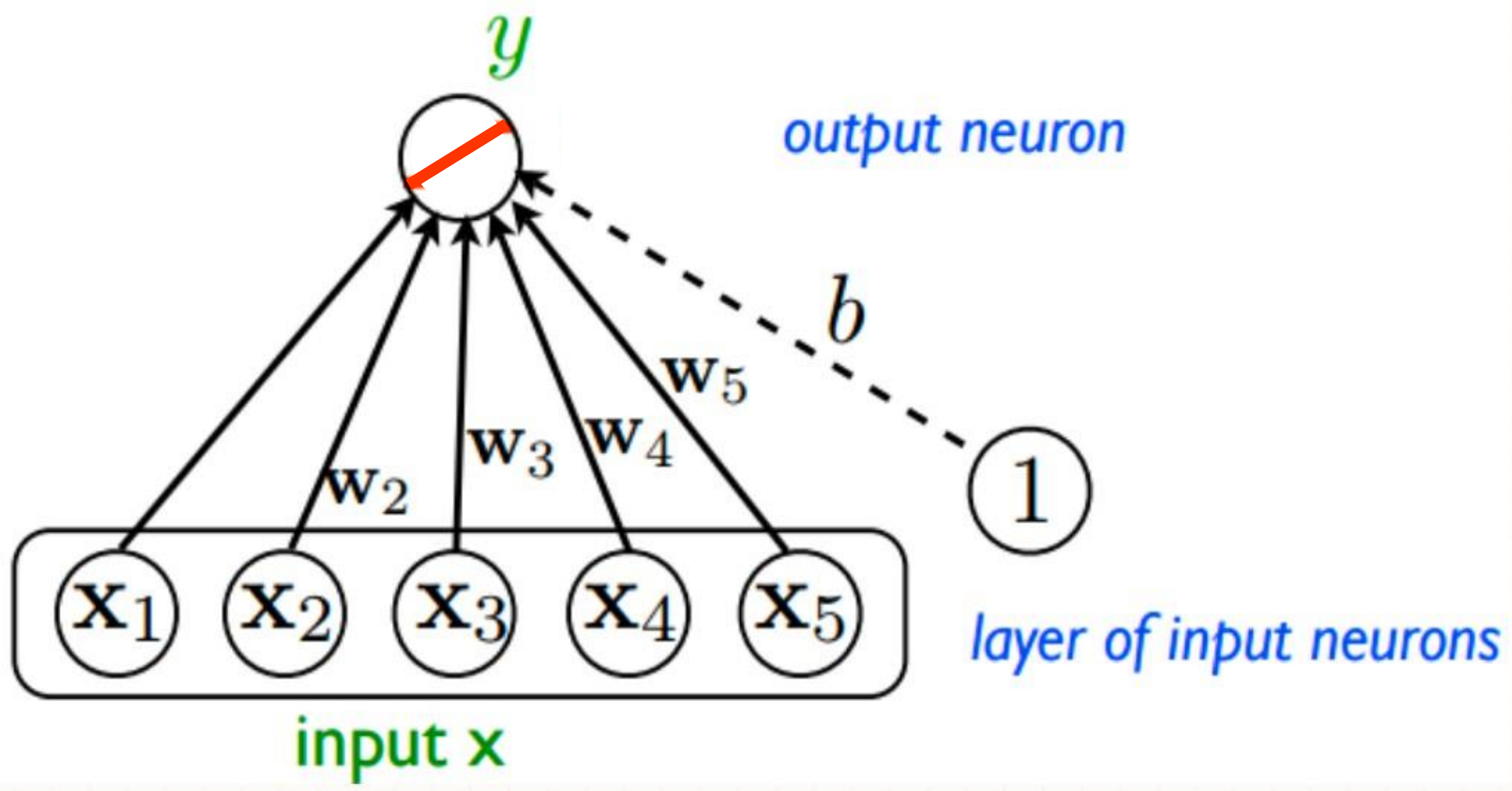
Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-635

9/29/2023

Special Thanks: Sargur N. Srihari, Andrew Ng, Carlos Guestrin



y
output neuron
b
w5
w2
w3
w4
1
x1
x2
x3
x4
x5
layer of input neurons
input x

Multiple Features (Input Variables)

Size in feet^2 (x_1)	Number of bedrooms (x_2)	Number of floors (x_3)	Age of home (years) (x_4)	Price ($) in 1000's (y)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…				…

Notation:

- n = Number of features
- $x^{(i)}$= Input features of i^{th} training example
- $x_j^{(i)}$= Value of feature j in i^{th} training example

$x_3^{(2)} = ?$

$x_3^{(4)} = ?$

Hypothesis: Multivariate Form

Previously: $h_\theta(x) = \theta_0 + \theta_1 x$

Now: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

- For convenience of notation, define $x_0 = 1$ ($x_0^{(i)} = 1$ for all examples)
- $\boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in R^{n+1}$ $\quad\quad$ $\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in R^{n+1}$
- $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$
  $= \boldsymbol{\theta}^\top \boldsymbol{x}$

Gradient Descent

- Previously ($n = 1$)

Repeat until convergence{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) x^{(i)}$$

}

- New algorithm ($n \geq 1$)

Repeat until convergence{

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) x_j^{(i)}$$

}

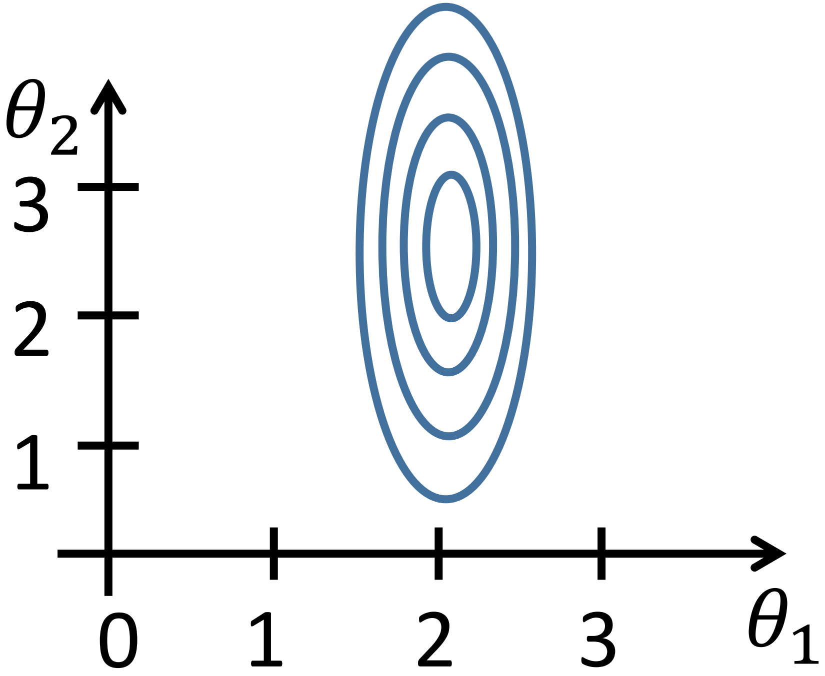
Simultaneously update
θ_j, for $j = 0, 1, \cdots, n$

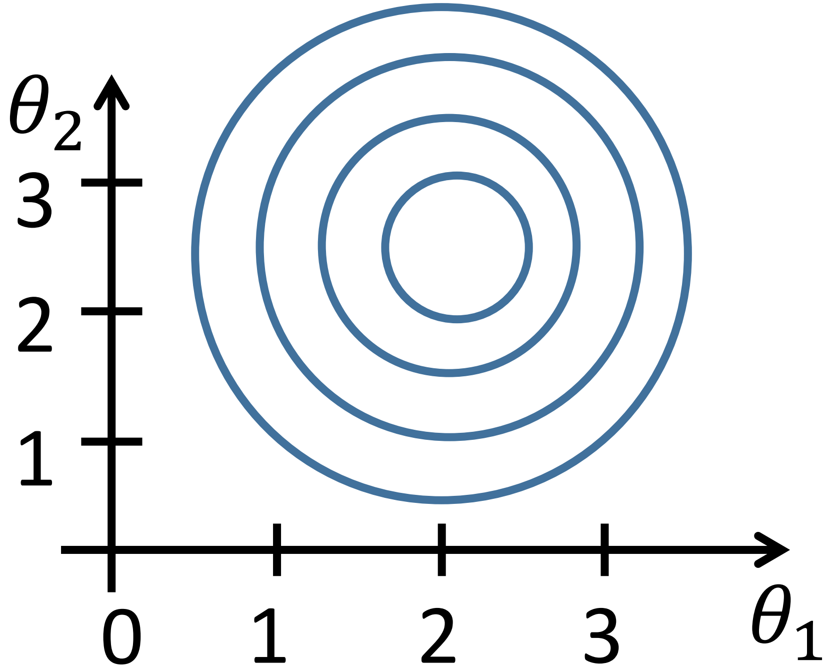
Gradient Descent in Practice: Feature Scaling

- Idea: Make sure features are on a similar scale (e.g,. $-1 \leq x_i \leq 1$)
- E.g. $x_1 =$ size (0-2000 feat^2)

  $x_2 =$ number of bedrooms (1-5)

$$a_i = \frac{v_i - \min v_i}{\max v_i - \min v_i}$$





Gradient Descent in Practice: Learning Rate

- Automatic convergence test
- α too small: slow convergence
- α too large: may not converge

- To choose α, try

0.001, … 0.01, …, 0.1, … , 1

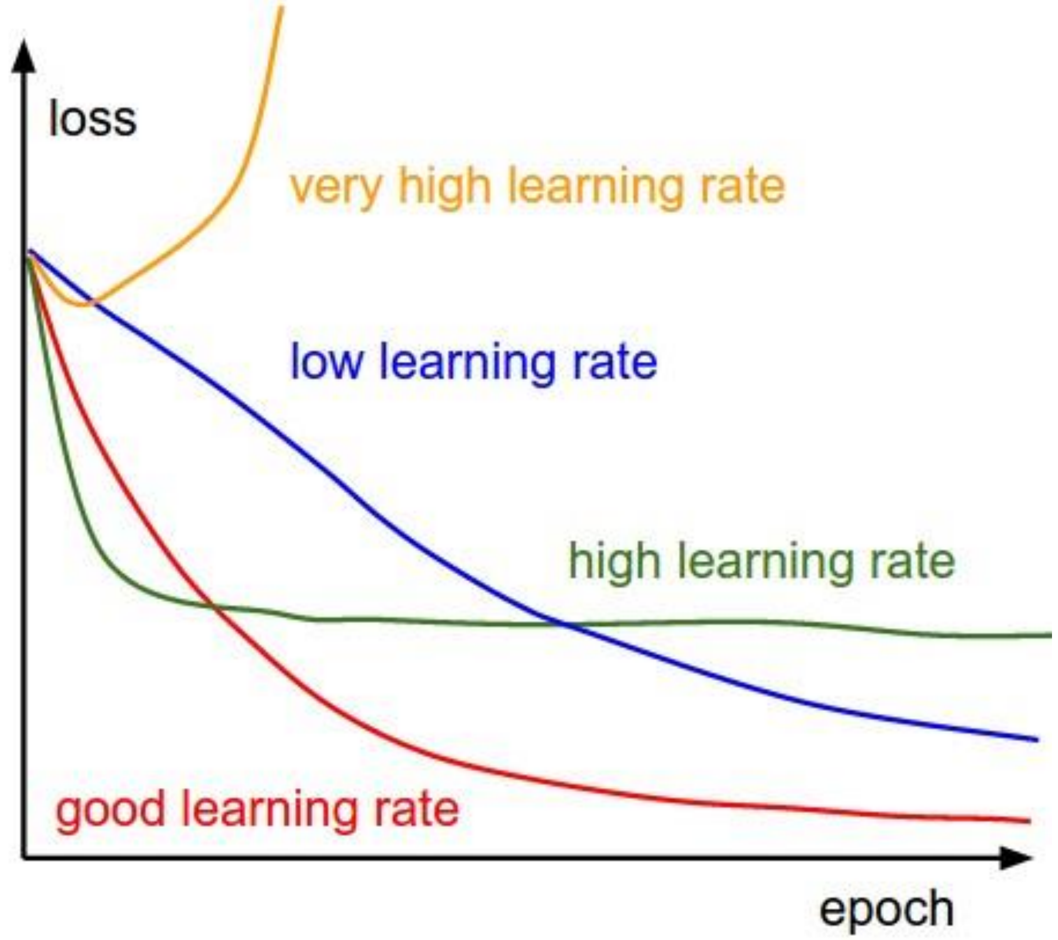


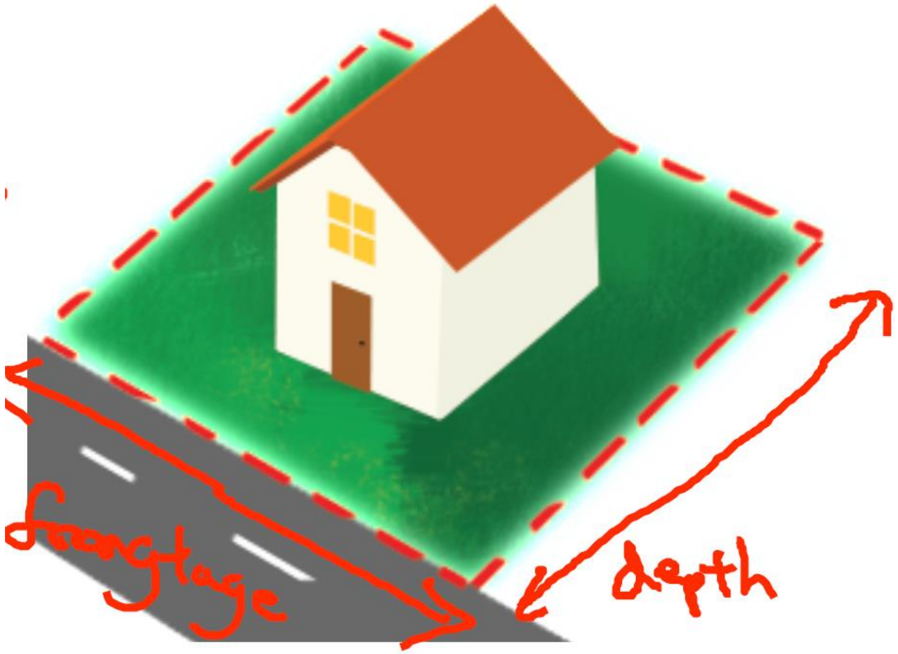
Image credit: CS231n@Stanford

House Price(s) Prediction

- $h_\theta(x) = \theta_0 + \theta_1 \times \text{frontage} + \theta_2 \times \text{depth}$

- Area: $xy = \text{frontage (x)} \times \text{depth (y)}$

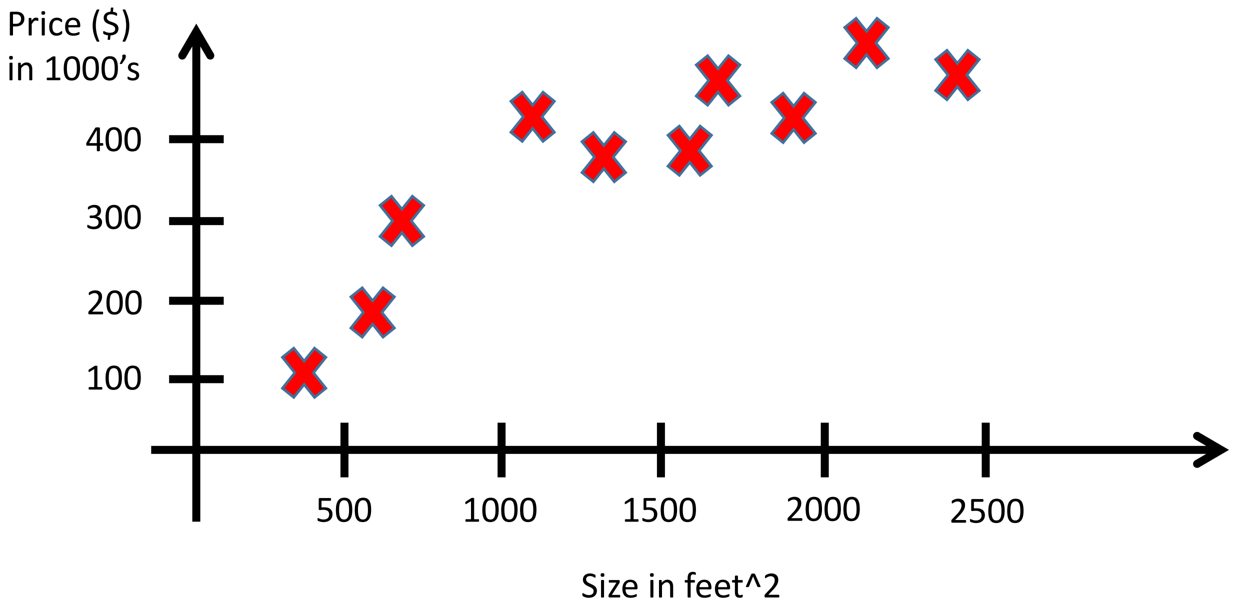
- Housing model: $h_\theta(x) = \theta_0 + \theta_1 xy$
  (a multiplicative feature interaction)



Slide credit: Andrew Ng

A Polynomial Hypothesis

Apply same principles of linear regression derivation for each parameter (just change index j > 1)



x_1 = (size)

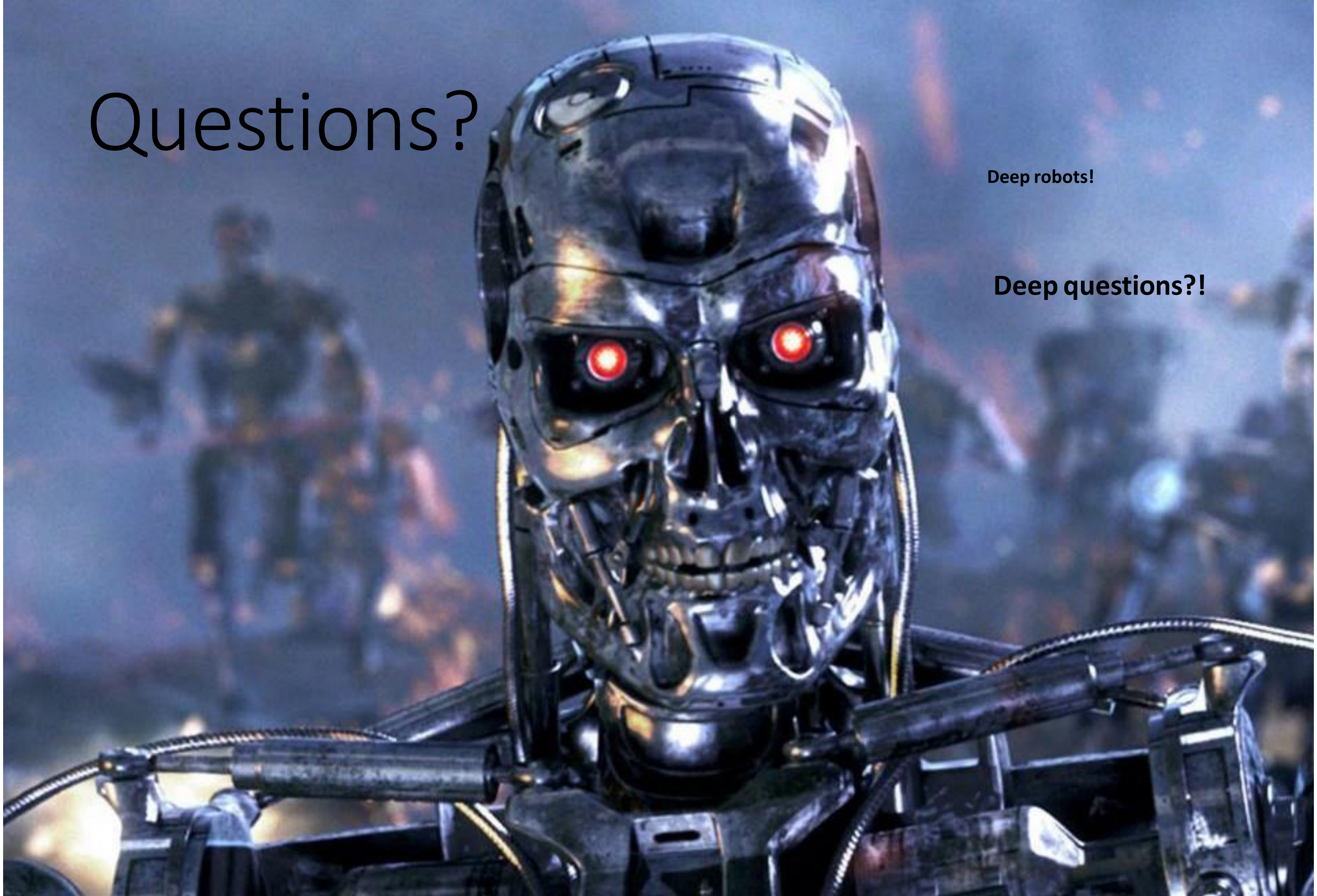
x_2 = (size)^2

x_3 = (size)^3

- $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

$$= \theta_0 + \theta_1(size) + \theta_2(size)^2 + \theta_3(size)^3$$

Questions?



Deep robots!

Deep questions?!