MACHINE LEARNING

LESSON 02:

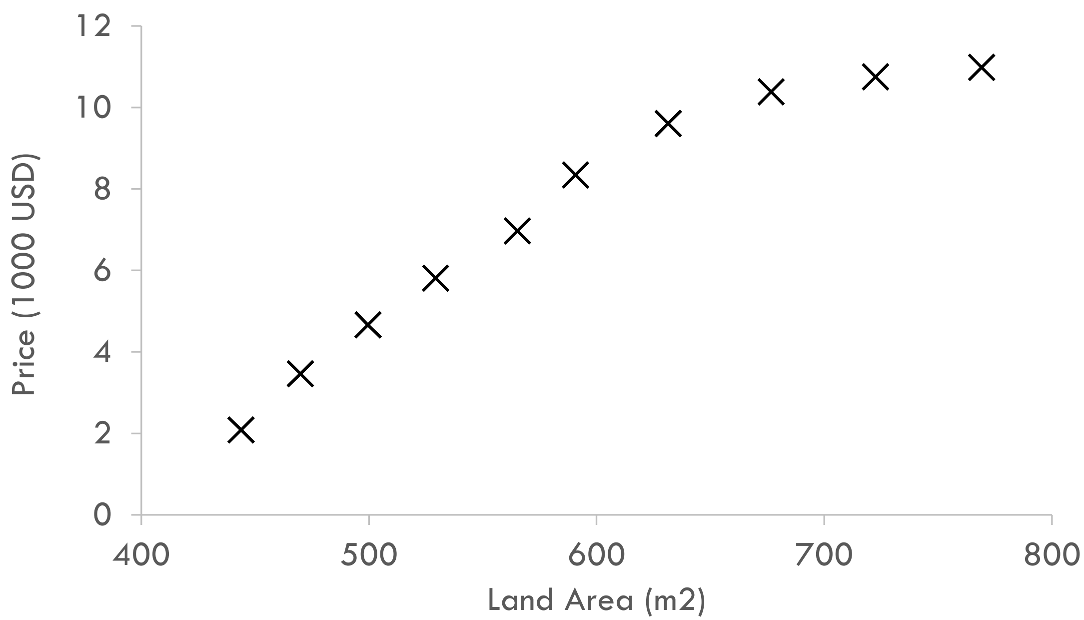
Linear Regression

Prepared by Dr. Dona Valy

Inspired by and adapted from the Coursera Machine Learning course by Andrew Ng
(coursera.org/learn/machine-learning)

EXAMPLE

Land price prediction



Supervised Learning

Regression Problem:
predict continuous valued output

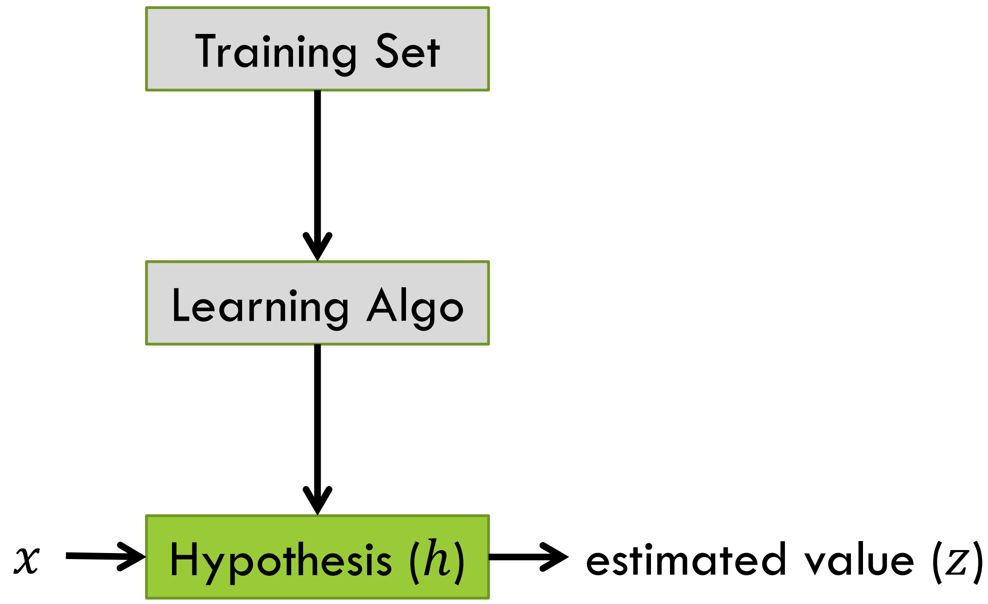
TRAINING SET

- Notations:
  - m = number of training examples
  - x = input variable / feature
  - y = output/target variable
  - (x, y) = one training example
  - $(x^{(i)}, y^{(i)})$ = the i^{th} training example

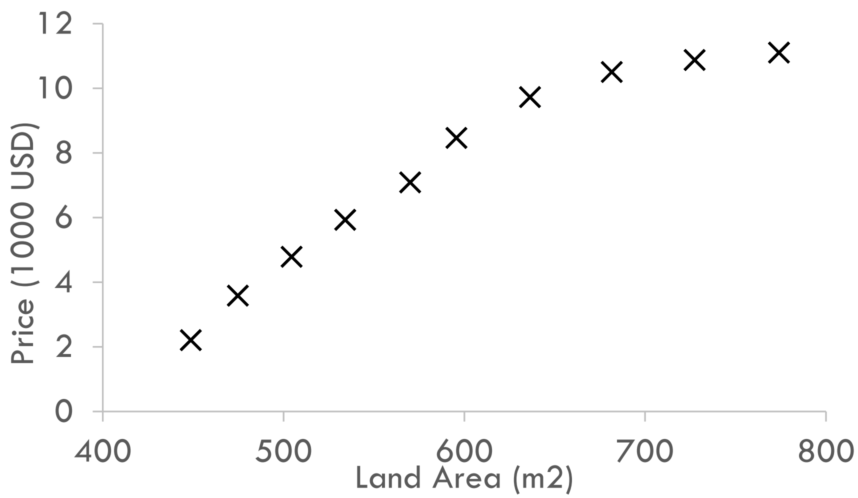
Example: training set of land price

Size in m^2 (x)	Price in 1000 USD (y)
90	2.1
130	3.9
210	4.5
…	…

MODEL REPRESENTATION



COST FUNCTION



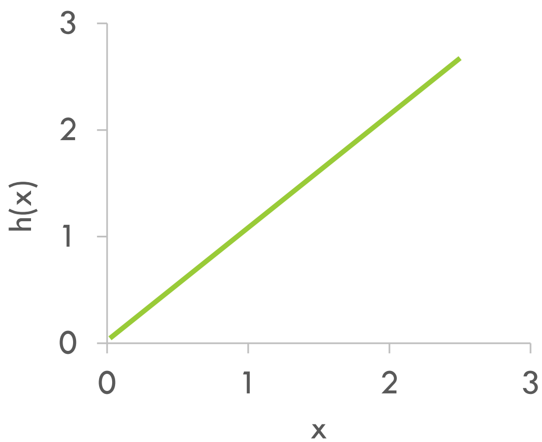
Size in m^2 (x)	Price in 1000 USD (y)
90	2.1
130	3.9
210	4.5
…	…

- Hypothesis: $h(x) = ax + b$ where a and b are called **parameters**

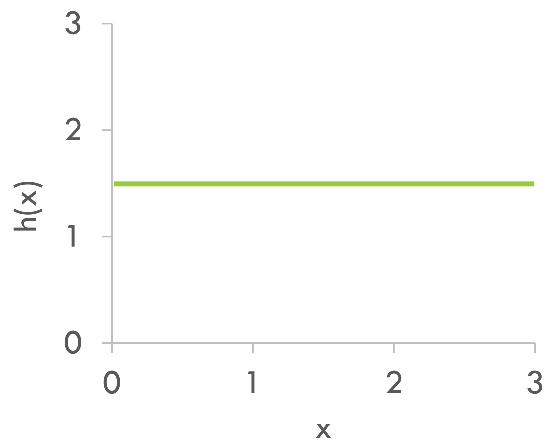
How to choose $\boldsymbol{a}$ and $\boldsymbol{b}$?

COST FUNCTION (CONT.)

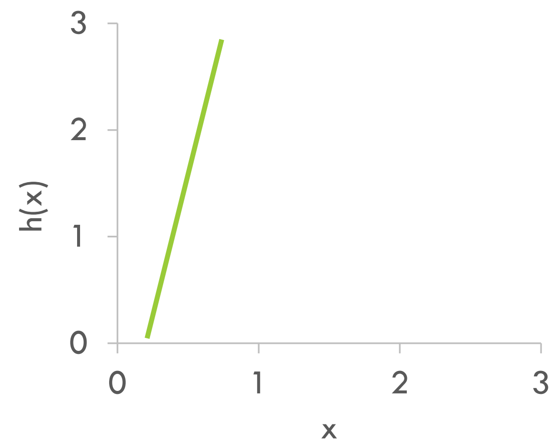
- Hypothesis: $h(x) = ax + b$



$a = 1$
$b = 0$

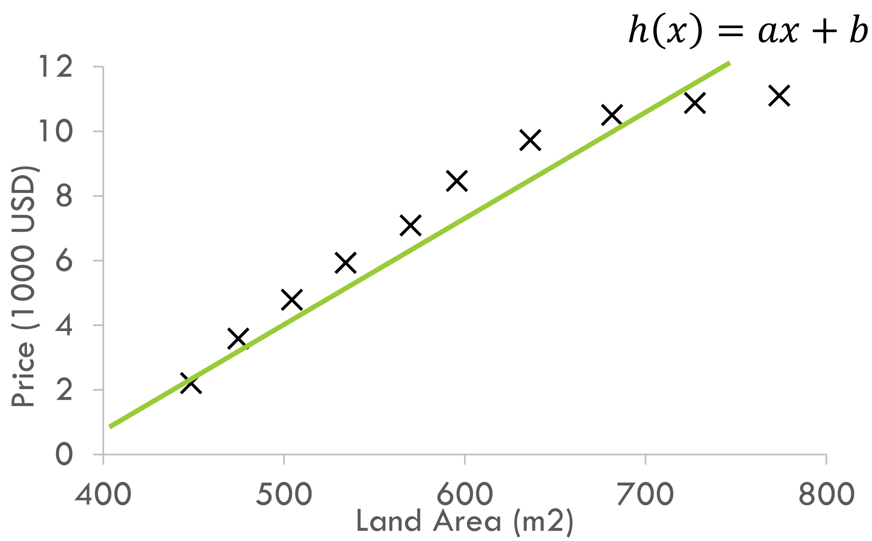


$a = 0$
$b = 1.5$



$a = 2.5$
$b = -0.5$

COST FUNCTION (CONT.)



- Choose a and b so that $h(x)$ is close to y for the training examples (x, y)

For each $(x^{(i)}, y^{(i)})$:
Minimize $|h(x^{(i)}) - y^{(i)}|$

$\Rightarrow$ Minimize $\frac{1}{\mathrm{m}} \sum_{i=1}^{m} |h(x^{(i)}) - y^{(i)}|$

Squared Error Function:

$$J(a, b) = \frac{1}{\mathrm{m}} \sum_{1=1}^{m} \left(h(x^{(i)}) - y^{(i)}\right)^2$$

$$\Rightarrow \min_{a,b} J(a, b)$$

GRADIENT DESCENT

Have some function $J(a, b)$

Want $\min_{a,b} J(a, b)$

Algo Outline:

- Start with some value of a and b
- Keep changing a and b to reduce $J(a, b)$ until hopefully we end up at a minimum

GRADIENT DESCENT (CONT.)

repeat until convergence {

$$grad_a = \frac{\partial}{\partial a} J(a, b)$$

$$grad_b = \frac{\partial}{\partial b} J(a, b)$$

Partial Derivative

$$a \mathrel{+}= -\alpha . grad_a$$

$$b \mathrel{+}= -\alpha . grad_b$$

}

Learning Rate

$$J(a, b) = \frac{1}{\mathrm{m}} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$\Rightarrow \quad \frac{\partial}{\partial a} J(a, b) = \frac{2}{\mathrm{m}} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right) . x^{(i)}$$

$$\frac{\partial}{\partial b} J(a, b) = \frac{2}{\mathrm{m}} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right)$$

GRADIENT DESCENT (CONT.)

- Things to consider:
  - Choosing the learning rate α
  - Global minimum vs local minimum

EXERCISE 2.1

Given a dataset of land price as illustrated in the table on the right, find a linear regression model which fits the data. Train the model using the gradient descent algorithm.

Size in m^2 (x)	**Price in 1000 USD (y)**
90	7.1
130	10.9
210	19.2
300	28
350	32.8
420	39.9
480	46.1
530	51
640	62.2
710	68.9

LR WITH MULTIPLE VARIABLES

Example: training set of land price

Size in m^2 **(x_1)**	**Distance from the city center in Km** **(x_2)**	**Price in 1000 USD (y)**
90	5	2.1
130	1	3.9
210	2.6	4.5
…	…	…

- Notation:
  - n = number of features
  - $x^{(i)}$ = input features of the i^{th} example
  - $x_j^{(i)}$ = value of feature j of the i^{th} example

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

LR WITH MULTIPLE VARIABLES

- Hypothesis: $h(x) = ax_1 + bx_2 + c$

  or $h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

  or $h(x) = \sum_{j=0}^{n} \theta_j x_j$ where $x_0 = 1$ and $n = 2$
- Cost function: $J(\theta_0, \theta_1, \dots, \theta_n) = J(\theta) = \frac{1}{\mathrm{m}} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)} \right)^2$
- Gradient descent:

repeat until convergence {

$$grad_{\theta_j} = \frac{\partial}{\partial \theta_j} J(\theta) = \frac{2}{\mathrm{m}} \sum_{1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)} \right) . x_j^{(i)}$$

$$\theta_j \mathrel{+}= -\alpha . grad_{\theta_j}$$

}

Do this for all $j = 0..n$

FEATURE SCALING

- Make sure all features are on a similar scale
- Solution:
  - Replace x_j with $\frac{x_j - \mu_j}{s_j}$

- mean $\mu_j = \frac{1}{m}\sum_{i=1}^{m} \mu_j^{(i)}$
- s_j standard deviation or simply $\max - \min$

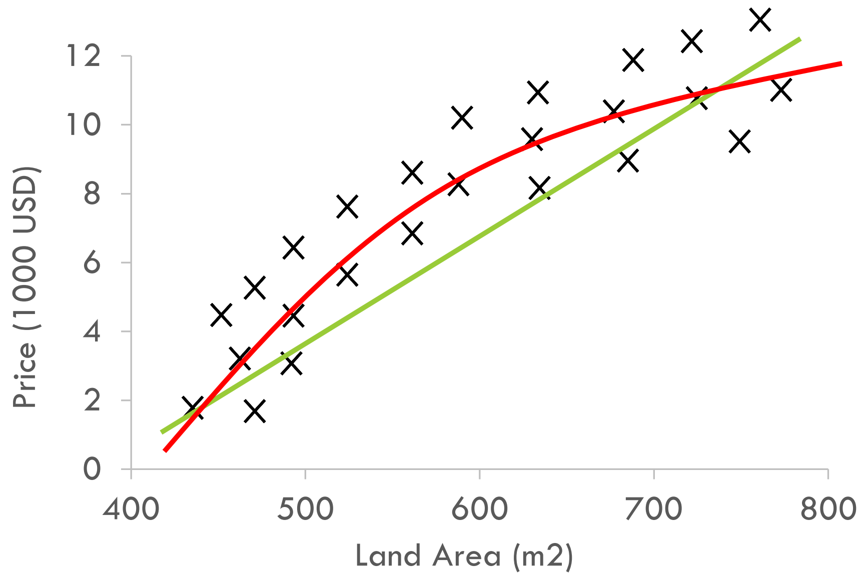
POLYNOMIAL REGRESSION

- Linear: $h(\theta) = \theta_0 + \theta_1 x$
- Polynomial:

$$h(\theta) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Let's denote $x_1 = x, x_2 = x^2$

$\Rightarrow h(\theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$



- Other polynomial examples:
  - $h(\theta) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$
  - $h(\theta) = \theta_0 + \theta_1 x + \theta_2 \sqrt{x}$

EXERCISE 2.2

Given a dataset of land price:

1. Build a linear regression which predicts the land price using both the land_area and the distance_to_city feature. (See the dataset in 'land_price_1.csv')
2. Using only the distance feature, build a model with hypothesis $h(x) = \theta_0 + \theta_1 x + \theta_2 \sqrt{x}$ to predict the land price. (See the dataset in 'land_price_2.csv')