Week 1 Lecture 2

Review of linear algebra & Introduction to optimization

GEOL 4397: Data analytics and machine learning for geoscientists

Jiajia Sun, Ph.D.

Jan. 17th, 2019

UNIVERSITY of HOUSTON

YOU ARE THE PRIDE

EARTH AND ATMOSPHERIC SCIENCES

Week	Date	Topics	Comments
		Overview of syllabus	
1	01/15 Tues	Lecture: Introduction to Machine learning: applications	
	01/17 Thur	Lecture: Review of linear algebra	
2	01/22 Tues	Lab: Linear algebra in Python	Not graded
	01/24 Thur	Lecture: Introduction to optimization	
3	01/29 Tues	Lab: Gradient descent + Linear regression	Report due on 02/05 at 5:30 pm
	01/31 Thur	Lecture: Introduction to machine learning: concepts	
4	02/05 Tues	Lecture: Logistic regression	
	02/07 Thur	Lab: Logistic regression	Report due on 02/14 at 5:30 pm
5	02/12 Tues	Lecture: Support vector machine	
	02/14 Thur	Lab: Support vector machine	Report due on 02/21 at 5:30 pm
6	02/19 Tues	Lecture: Decision trees	
	02/21 Thur	Lab: Decision trees	Report due on 02/28 at 5:30 pm
7	02/26 Tues	Lecture: Random Forest	
	02/28 Thur	Lab: Random forest	Report due on 03/07 at 5:30 pm
8	03/05 Tues	Lecture: Ensemble learning	
	03/07 Thur	Lab: Ensemble learning	Reprot due on 03/19 at 5:30 pm
9	03/12 Tues	No class due to spring break	
	03/14 Thur	No class due to spring break	
10	03/19 Tues	Review & Recap	
	03/21 Thur	Exam	
11	03/26 Tues	Lecture: Clustering	
	03/28 Thur	Lab: Clustering	Report due on 04/04 at 5:30 pm
12	04/02 Tues	Lecture: Introduction to TensorFlow	
	04/04 Thur	Lab: TensorFlow	Not graded
13	04/09 Tues	Lecture: Introduction to neural networks 1	
	04/11 Thur	Lecture: Introduction to neural networks 2	
14	04/16 Tues	Lab: Deep learning	Report due on 04/23 at 5:30pm
	04/18 Thur	Lecture: Convolutional neural networks 1	
15	04/23 Tues	Guest lecture: Convolutional neural networks 2	
	04/25 Thur	Lab: CNN (optional)	Report due on 05/02 at 5:30 pm
16	04/30 Tues	final presentation??	
	05/02 Thur	final presentation??	
Note	28 class meetings		04/29 last day of class

Today’s agenda

- Linear regression model

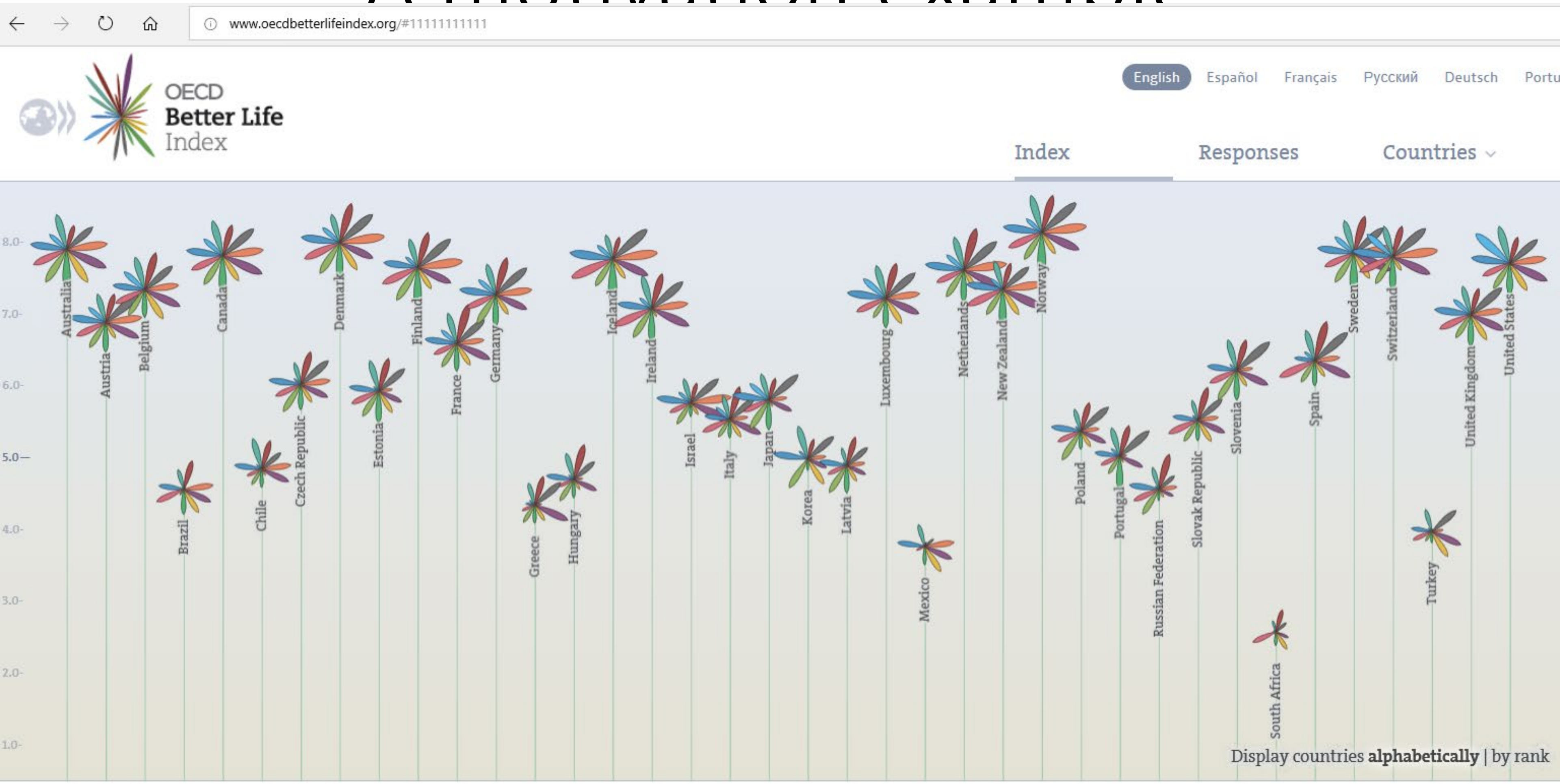
- Review of linear algebra

A motivation example

- Does money makes people happy?

- Download OECD's Better Life Index data (http://www.oecdbetterlifeindex.org/)

A motivation example



A motivation example

- Does money makes people happy?

- Download OECD's Better Life Index data (http://www.oecdbetterlifeindex.org/)

- Download GPD per capita from IMF's website (https://goo.gl/doxvTP) https://www.imf.org/external/pubs/ft/weo/2016/01/weodata/weorept.aspx?pr.x=32&pr.y=8&sy=2015&ey=2015&scsm=1&ssd=1&sort=country&ds=.&br=1&c=512%2C668%2C914%2C672%2C612%2C946%2C614%2C137%2C311%

World Economic Outlook Database, April 2016

Step 5 of 5 — 1 » 2 » 3 » 4 » 5 — New Query

5. Report for Selected Countries and Subjects

You will find notes on the data and options to download the table below your results.

Shaded cells indicate IMF staff estimates

Country	Subject Descriptor	Units	Scale	Country/Series-specific Notes	2015
Afghanistan	Gross domestic product per capita, current prices	U.S. dollars	Units		599.994
Albania	Gross domestic product per capita, current prices	U.S. dollars	Units		3,995.383
Algeria	Gross domestic product per capita, current prices	U.S. dollars	Units		4,318.135
Angola	Gross domestic product per capita, current prices	U.S. dollars	Units		4,100.315
Antigua and Barbuda	Gross domestic product per capita, current prices	U.S. dollars	Units		14,414.302
Argentina	Gross domestic product per capita, current prices	U.S. dollars	Units		13,588.846
Armenia	Gross domestic product per capita, current prices	U.S. dollars	Units		3,534.860
Australia	Gross domestic product per capita, current prices	U.S. dollars	Units		50,961.865
Austria	Gross domestic product per capita, current prices	U.S. dollars	Units		43,724.031
Azerbaijan	Gross domestic product per capita, current prices	U.S. dollars	Units		5,739.433
The Bahamas	Gross domestic product per capita, current prices	U.S. dollars	Units		23,902.805
Bahrain	Gross domestic product per capita, current prices	U.S. dollars	Units		23,509.981
Bangladesh	Gross domestic product per capita, current prices	U.S. dollars	Units		1,286.868
Barbados	Gross domestic product per capita, current prices	U.S. dollars	Units		15,773.555
Belarus	Gross domestic product per capita, current prices	U.S. dollars	Units		5,749.119
Belgium	Gross domestic product per capita, current prices	U.S. dollars	Units		40,106.632
Belize	Gross domestic product per capita, current prices	U.S. dollars	Units		4,841.735
Benin	Gross domestic product per capita, current prices	U.S. dollars	Units		780.063

A motivation example

- Does money makes people happy?

- Download OECD’s Better Life Index data (http://www.oecdbetterlifeindex.org/)

- Download GPD per capita from IMF’s website (https://goo.gl/doxvTP)

A motivation example

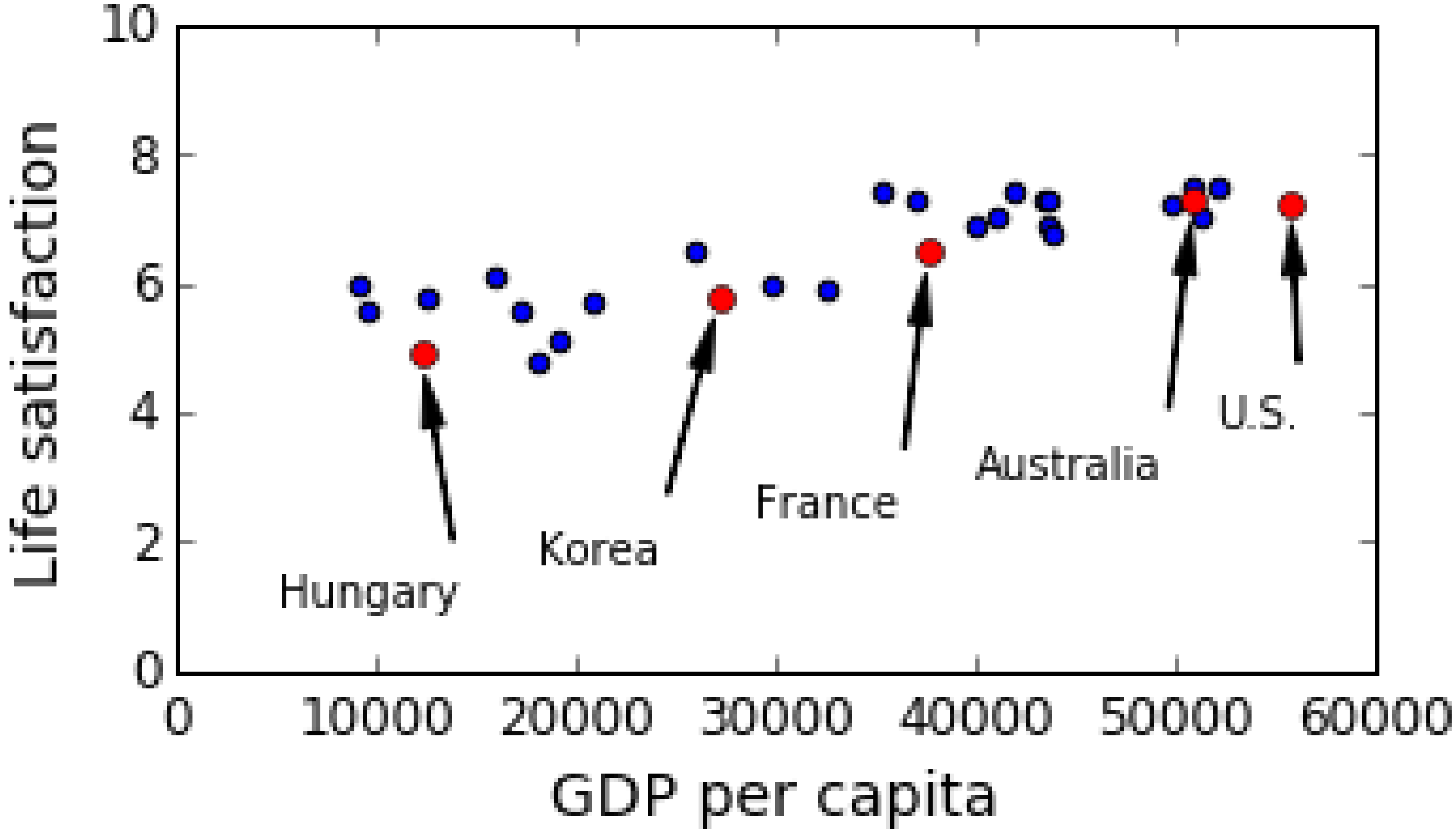


Figure from Aurelien Geron's ML book, page 19

Goal

- Learn a model from the data

Goal

- Build a model from the data

Goal

- Train a model from the data

Goal

- Train a model from the training data

Goal

- Train a model from the training data
- Make predictions

Goal

- Learn a model from the training data
- Make predictions

What is machine learning?

- My definition: the field of study that gives computers the ability to learn from data (e.g., discovering patterns and relations among input data), and make predictions.

Training data

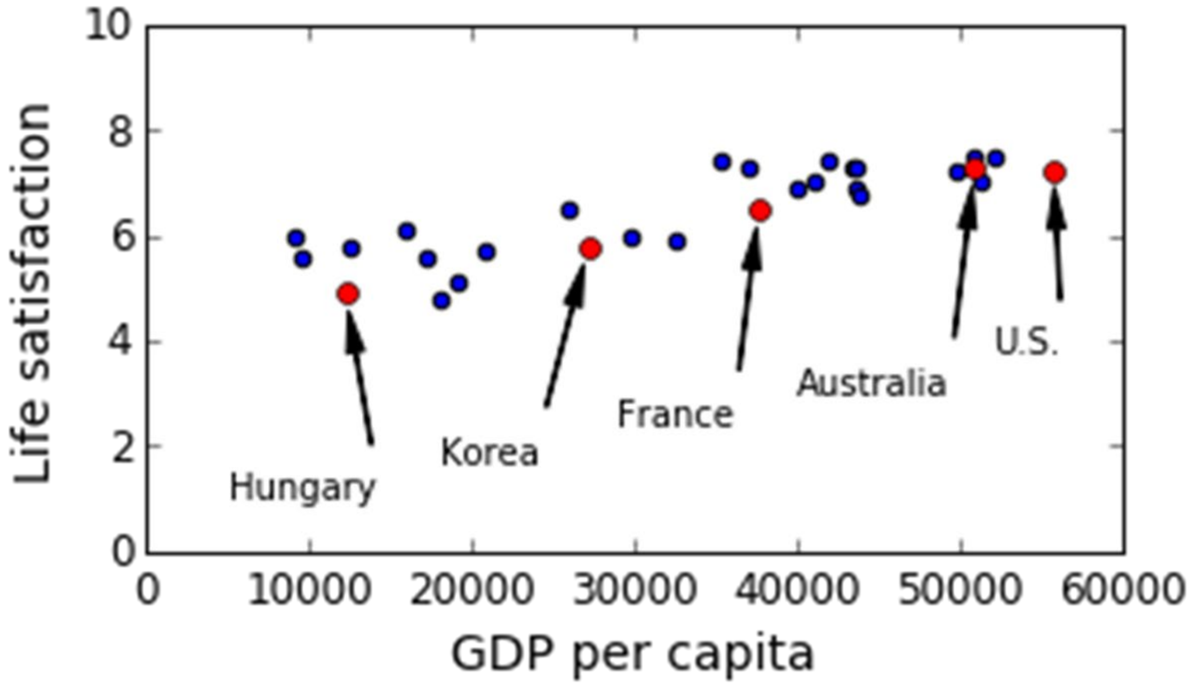


Figure from Aurelien Geron's ML book, page 19

Training data

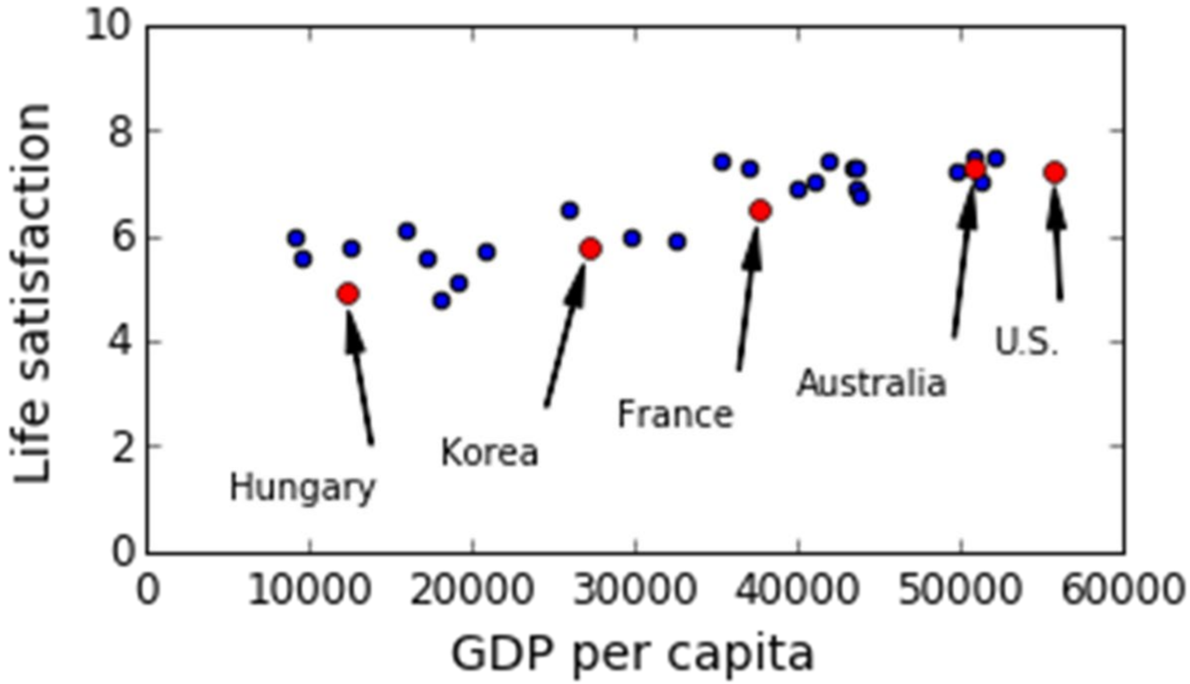


Figure from Aurelien Geron's ML book, page 19

- There does seem to be a trend!
- Looks like life satisfaction goes up more or less linearly with the country's GDP per capita.

Hypothesis

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

- $h_\theta(x)$: life satisfaction
- x : GDP per capita
- θ_0 , θ_1: model parameters (to be learned from training data)

Linear relationship

- By tweaking the model parameters, we can use the linear regression model to represent any linear function.

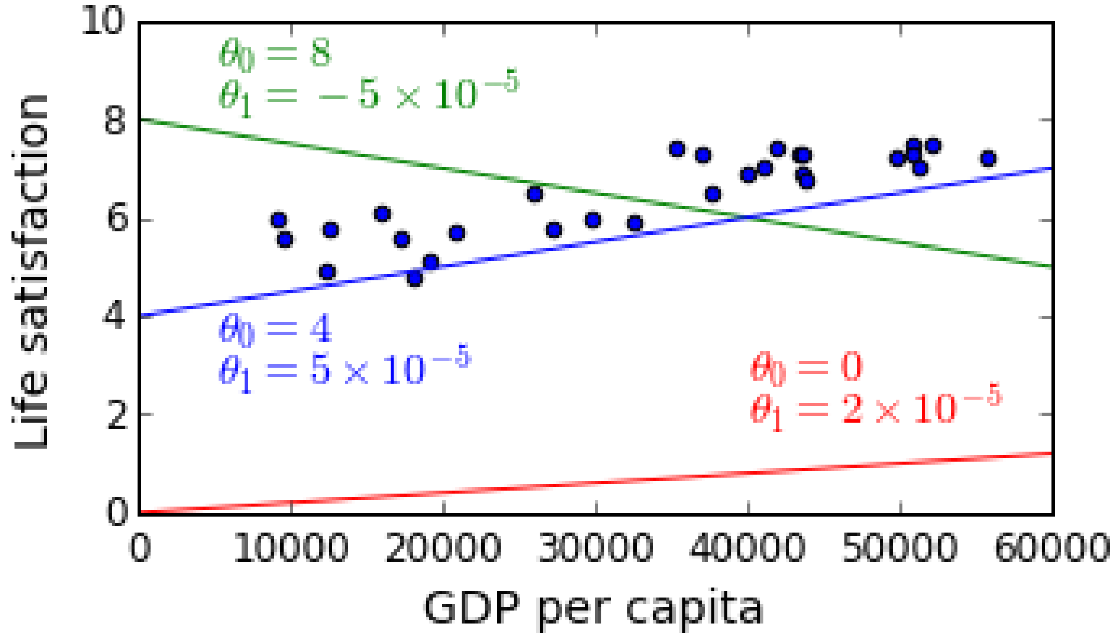


Figure from Aurelien Geron’s ML book, page 19

Training/Learning

- Which one to choose?

Training/Learning

- Which one to choose?
- Or, a more general question is, How to train/learn these model parameters??

Training/Learning

- Which one to choose?
- Or, a more general question is, How to train/learn these model parameters??
- The answer is by

defining a cost function

&

minimizing it

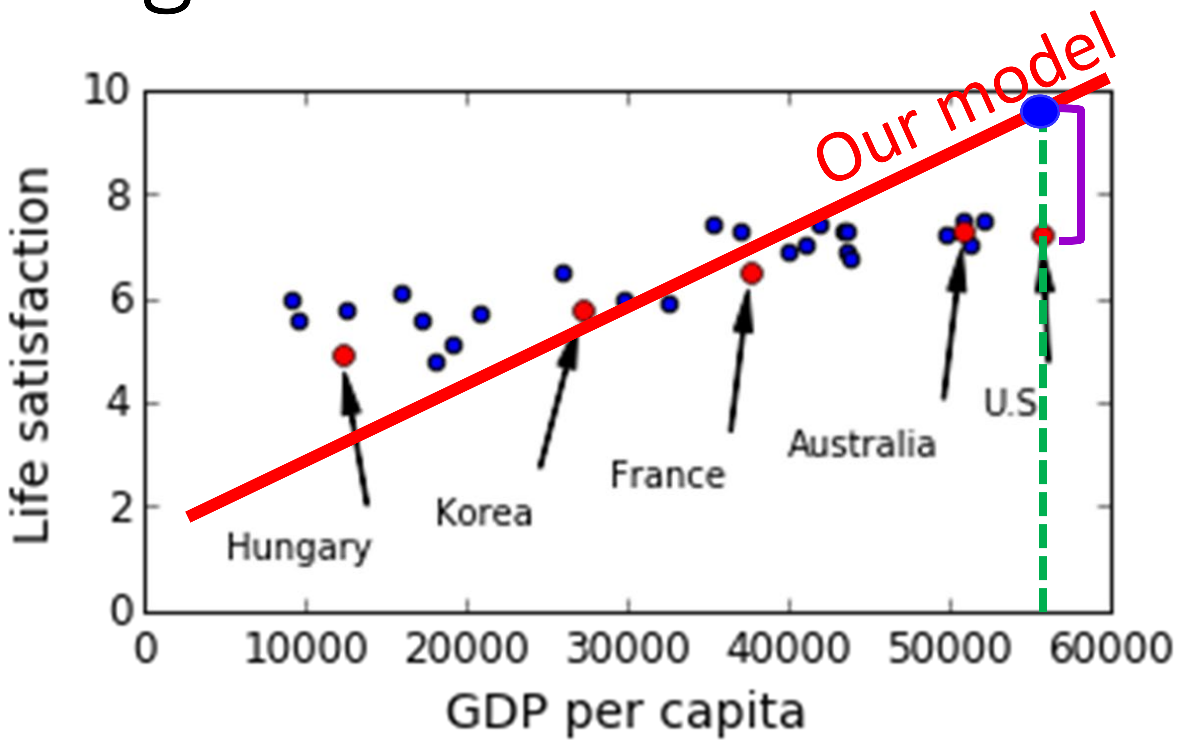
Cost function: basic idea

- Measures how bad (or good) a candidate model is
- Specifically, measure the difference between predictions from our model and the training data

Cost function: basic idea

- Measures how bad (or good) a candidate model is
- Specifically, measure the difference between predictions from our model and the training data
- The objective is to minimize the cost function so as to minimize the difference between predictions and observations

Building a cost function



For simplicity, let us focus on one country, say, U.S.

- What is the predicted value for life satisfaction?
- What is the value from training data?

Difference between predicted and true values

- In our training data, we have M countries. Suppose U.S. is the i^{th} country.
- Predicted value: $h_\theta(x^{(i)})$
- True value: $y^{(i)}$

Difference between predicted and true values

- In our training data, we have M countries. Suppose U.S. is the i^{th} country.
- Predicted value: $h_\theta(x^{(i)})$
- True value: $y^{(i)}$
- Difference:

$$(h_\theta(x^{(i)}) - y^{(i)})^2$$

Difference between predicted and true values

- In our training data, we have M countries. Suppose U.S. is the i^{th} country.
- Predicted value: $h_\theta(x^{(i)})$
- True value: $y^{(i)}$
- Difference:

$$(h_\theta(x^{(i)}) - y^{(i)})^2$$

- This is only for U.S.

Difference between predicted and true values

- In our training data, we have M countries. Suppose U.S. is the i^{th} country.
- Predicted value: $h_\theta(x^{(i)})$
- True value: $y^{(i)}$
- Difference:

$$(h_\theta(x^{(i)}) - y^{(i)})^2$$

- This is only for U.S.
- Remember that we want our model to fit all of our training data, not just one of them. Therefore, we sum the differences over all countries

Difference between predicted and true values

- In our training data, we have M countries. Suppose U.S. is the i^{th} country.
- Predicted value: $h_\theta(x^{(i)})$
- True value: $y^{(i)}$
- Difference:

$$\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

- This is only for U.S.
- Remember that we want our model to fit all of our training data, not just one of them. Therefore, we sum the differences over all countries

$$\sum_{i=1}^{M}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Cost function

$$\sum_{i=1}^{M}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Remember

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

Cost function

$$\sum_{i=1}^{M} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Remember

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

Therefore,

$$\sum_{i=1}^{M} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

Cost function

$$\sum_{i=1}^{M} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Remember

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

Therefore,

$$J(\theta_0, \theta_1) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

Minimization

- Cost function measures the difference between predicted and true values
- Remember that, we want to minimize this difference, i.e.,

$$\min\ J(\theta_0, \theta_1) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

- Learning/training = Minimizing a cost function

- The process of learning a model from training data is essentially the process of minimizing a cost function.

Optimization

- Cost function measures the difference between predicted and true values
- Remember that, we want to minimize this difference, i.e.,

$$\min \; J(\theta_0, \theta_1) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

- Learning/training = Minimizing a cost function

- The process of learning a model from training data is essentially the process of minimizing a cost function.
- Optimization: finding optimal parameter values that minimize a cost function

Optimization

- Cost function measures the difference between predicted and true values
- Remember that, we want to minimize this difference, i.e.,

$$\min\ J(\theta_0, \theta_1) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

- $\boxed{\text{Learning/training = Minimizing a cost function}}$

- The process of learning a model from training data is essentially the process of optimization.
- Optimization: finding optimal parameter values that minimize a cost function

Best fit model

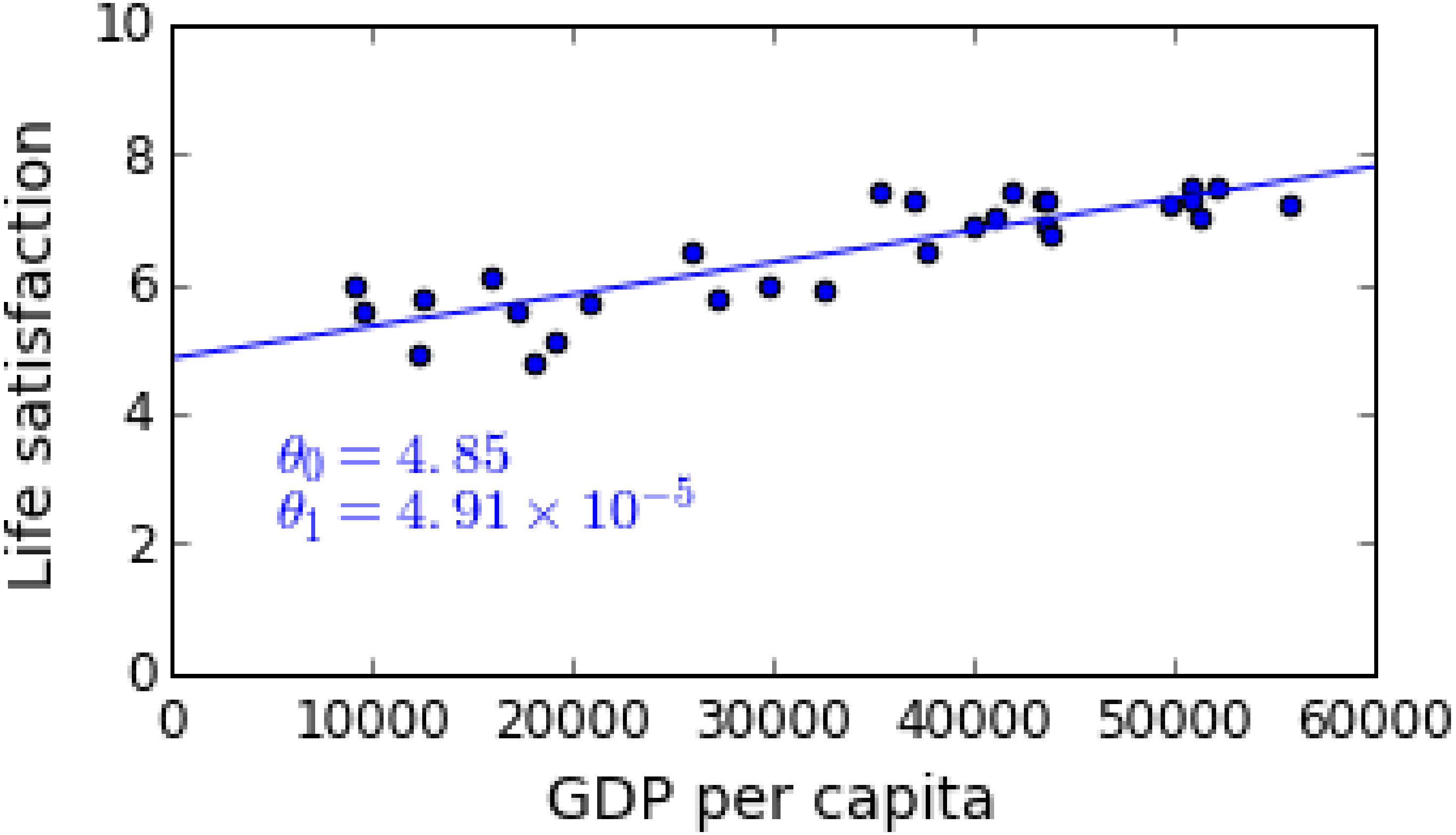


Figure from Aurelien Geron's ML book, page 20

- How does linear algebra come into play?

Hypothesis: linear regression model

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

- $h_\theta(x^{(i)})$: life satisfaction for i^{th} country
- $x^{(i)}$: GDP per capita for i^{th} country
- θ_0 , θ_1: model parameters (to be learned from training data)

Hypothesis: linear regression model

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

- $h_\theta(x^{(i)})$: output variable/target variable for i^{th} country
- $x^{(i)}$: input variable/feature for i^{th} country
- θ_0, θ_1: model parameters (to be learned from training data)

Features

- In previous example, we only have one feature, i.e., GDP per capita, (to predict life satisfaction)
- There are many other relevant features!
  - Employment rate
  - Education
  - Medical care
  - Air quality
  - Crime
  - …

Hypothesis: multiple input variables

$$h_\theta\left(x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)}\right) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)}$$

- $h_\theta(x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)})$: output variable/target variable
- $x_j^{(i)}$: j^{th} input variable/features for i^{th} country
- $\theta_0, \theta_1, \dots, \theta_N$: model parameters (to be learned from training data)
- A straightforward generalization to multiple input variables

Cost function

- Cost function for one input variable

$$J(\theta_0, \theta_1) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

- Cost function for multiple input variables

$$J(\theta) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)} - y^{(i)})^2$$

Let us use linear algebra to simplify the cost function!

Matrix

- Rectangular array of numbers

$$\boldsymbol{A} = \begin{bmatrix} 1402 & 25 \\ 1650 & 46 \\ 2058 & 57 \end{bmatrix}_{3X2}$$

Dimension of a matrix: # of rows X # of columns

Matrix elements

$$\boldsymbol{A} = \begin{bmatrix} 1402 & 25 \\ 1650 & 46 \\ 2058 & 57 \end{bmatrix}_{3X2}$$

A_{ij}: the element/entry at the i^{th} row and j^{th} column

Vector: An N X 1 matrix

$$\boldsymbol{y} = \begin{bmatrix} 604 \\ 731 \\ 172 \\ 495 \end{bmatrix}$$

- N = 4, therefore, $\boldsymbol{y}$ is a 4-dimensional vector
- y_i: i^{th} element

Norm

- Measures the length of a vector

$$a = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 2 \end{bmatrix}$$

$$\|a\| = \sqrt{1^2 + 3^2 + 4^2 + 2^2} = \sqrt{30}$$

$$\|a\|^2 = 1^2 + 3^2 + 4^2 + 2^2 = 30$$

Matrix addition

$$\underset{\text{3X2}}{\begin{bmatrix} 5.2 & 2.0 \\ 4.6 & 0.5 \\ 1.9 & 4.3 \end{bmatrix}} + \underset{\text{3X2}}{\begin{bmatrix} 0.5 & 1.4 \\ 4.1 & 1.1 \\ 5.8 & 3.7 \end{bmatrix}} = \underset{\text{3X2}}{\begin{bmatrix} 5.7 & 3.4 \\ 8.7 & 1.6 \\ 7.7 & 8.0 \end{bmatrix}}$$

$$\underset{\text{3X2}}{\begin{bmatrix} 5.2 & 2.0 \\ 4.6 & 0.5 \\ 1.9 & 4.3 \end{bmatrix}} + \underset{\text{2X2}}{\begin{bmatrix} 0.5 & 1.4 \\ 4.1 & 1.1 \end{bmatrix}} = error$$

Scalar multiplication

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 6 \\ 7 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 6 & 18 \\ 21 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 12 & 16 \\ 8 & 4 \end{bmatrix} \div 4 = \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix}$$

In-class quiz

$$3 \times \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \\ 7 \end{bmatrix} - \begin{bmatrix} 9 \\ 0 \\ 3 \end{bmatrix} \div 3$$

Remember: division and multiplication take precedence over addition and subtraction

Matrix-vector multiplication

$$A \quad \times \quad x \quad = \quad y$$

$$\underset{2X3}{\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}} \times \underset{3X1}{\begin{bmatrix} x \\ y \\ z \end{bmatrix}} = \underset{2X1}{\begin{bmatrix} ax + by + cz \\ dx + ey + fz \end{bmatrix}}$$

- The number of columns of matrix $\boldsymbol{A}$ must be equal to the number of elements in vector $\boldsymbol{x}$

Matrix-vector multiplication

$$A \times x = y$$

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \cdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{bmatrix}$$

m x n n x 1 m x 1

- To get y_i, multiple the i^{th} row of matrix $\boldsymbol{A}$ with the elements of vector $\boldsymbol{x}$, and add them up

In-class quiz

$$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & 4 \end{bmatrix} \times \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

Matrix-matrix multiplication

$$A_{\text{mxn}} \times \; B_{\text{nxp}} = \; C_{\text{mxp}}$$

$$\underbrace{\begin{bmatrix} & & \\ & & \end{bmatrix}}_{\text{mxn}} \times \underbrace{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}_{\text{nxp}} = \underbrace{\begin{bmatrix} & & \\ & & \end{bmatrix}}_{\text{mxp}}$$

- The ith column of matrix $\boldsymbol{C}$ is obtained by multiplying $\boldsymbol{A}$ with the i^{th} column of B (for i = 1, 2, …, p)
- Matrix-matrix multiplication is just a sequence of matrix-vector multiplications

In-class quiz

$$\begin{bmatrix} 5 & 2 \\ 3 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix}$$

Identity matrix

$$I_{3\times3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$I_{4\times4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrix inverse

- Suppose A is a m x m matrix, and suppose it has an inverse, then
- We can write the inverse as A^{-1}

$$AA^{-1} = I$$

$$A^{-1}A = I$$

Example

$$\begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{7} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} -2 & \frac{5}{3} \\ 1 & -\frac{2}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix transpose

- $A = \begin{bmatrix} 1 & 3 & 2 \\ 5 & 1 & 4 \end{bmatrix}$
- $B = A^T = \begin{bmatrix} 1 & 5 \\ 3 & 1 \\ 2 & 4 \end{bmatrix}$
- $B_{ij} = A_{ji}$
- If A is an m x n matrix, then A^T is an n x m matrix

Back to our cost function

$$J(\theta) = \sum_{i=1}^{M} (\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)} - y^{(i)})^2$$

Notation

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \cdots \\ \theta_N \end{bmatrix}$$

$$x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \\ x_2^{(i)} \\ \cdots \\ x_N^{(i)} \end{bmatrix}$$

Simplification

$$J(\theta) = \sum_{i=1}^{M} \left(\boxed{\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)}} - y^{(i)} \right)^2$$

$$\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)} = \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}$$

Simplification

$$J(\theta) = \sum_{i=1}^{M} \left(\boxed{\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)}} - y^{(i)} \right)^2$$

$$\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)} = \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}$$

$$J(\theta) = \sum_{i=1}^{M} \left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)} - y^{(i)} \right)^2$$

Matrix-vector form

$$J(\theta) = \sum_{i=1}^{M} (\boldsymbol{\theta}^T \boldsymbol{x}^{(i)} - y^{(i)})^2$$

$$J(\theta) = \|\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}\|^2$$

$$\boldsymbol{X} = \begin{bmatrix} (\boldsymbol{x}^{(1)})^T \\ \dots \\ (\boldsymbol{x}^{(M)})^T \end{bmatrix} \qquad y = \begin{bmatrix} y^{(1)} \\ \dots \\ y^{(M)} \end{bmatrix}$$

Mx(N+1) Mx1