Table of Contents

[Week 1 Review 1](#_Toc515458760)

[Welcome 1](#_Toc515458761)

[Supervised learning 2](#_Toc515458762)

[Regression 2](#_Toc515458763)

[Classifications 2](#_Toc515458764)

[unsupervised learning 2](#_Toc515458765)

[Linear Algebra Review 4](#_Toc515458766)

[Matrix and Vectors 4](#_Toc515458767)

[Matrix 4](#_Toc515458768)

[Vector: An nx1 Matrix 4](#_Toc515458769)

[Matrix Math 4](#_Toc515458770)

[Addition 4](#_Toc515458771)

[Multiplication (Scalar (real number)) 4](#_Toc515458772)

[Matrix Matrix Multiplication 5](#_Toc515458773)

[Matrix Vector Multiplication 5](#_Toc515458774)

[Housing price trick with Matrix: 5](#_Toc515458775)

[Inverse 5](#_Toc515458776)

[Transpose 6](#_Toc515458777)

# Week 1 Review

## Welcome

What is machine learning?

1. Arthur Samuel = "The field of study that gives computers the ability to learn without being explicitly programmed"

2. Tom Mitchell = "A computer program is said to learn from experience E with respect to some class of tasks T and performance measured P,

If its performance as tasks in T, as measured by P, improves with experience E."

Example: playing checkers

E = the experience of playing many games of checkers

T = the task of playing checkers

P = the probability that the program will win the next game

In general, any ML problem can be assigned to one of the two broad classifications

Supervised learning

Example: "right answers" given

Regression: predict continuous valued output (price) = housing price predict

Classifications: discrete valued output (0 or 1) = breast cancer (malignant, benign)

In Supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between input and output.

Example 1:

Given data about the size of houses on a real estate market, try to predict their price. Price as a function of size is a continuous output, so this is a regression problem.

We could turn this example into a classification problem instead of making our output about whether the house "sells for more or less than the asking price." Here we are classifying the houses based on price into two discrete categories.

Example 2:

(a) Regression - Given a picture of a person, we have to predict their age on basis of the given picture

(b) Classification - Given a patient with a tumor, we have to predict whether the tumor is malignant or benign.

Unsupervised learning

Example: news.google.com --> cluster articles together

Organize computing clusters, social network analysis, market segmentation, astronomical data analysis

Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of variables. We can derive this structure by clustering the data based on relationships among the variables in the data. With unsupervised learning there is no feedback based on the prediction results.

Example:

Clustering: Take a collection of 1m different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, and roles and so on.

Non-clustering: The "Cocktail Party Algorithm”, allows you to find structure in a chaotic environment.

Model Representation

House Prices = Supervised learning, regression (predict real-valued output)

Training set of size in feet squared and price in 1,000s

Notation:

m = Number of training examples

x's = "input" variable/features = size in feet

y's = "output" variable /"target" variable = price in 1,000s

(x,y) - one training example

(xi, yi) - ith training example = i is a index, not power of i

x1, y1 = 2104, 460

Training Set --> Learning Algorithm --> hypothesis

Input size of house and output the price of the house

h maps from x's and y's

Linear regression with one variable

Univariate linear regression

Cost Function

Squared error Function - used a lot in regression for cost function.

Cost function = we can measure the accuracy of our hypothesis function by using a cost function.

This takes an average difference (actually a fancier version of an average) of all results of the hypothesis with inputs from x's and the actual output y's.

Cost Function - Intuition I

y = mx + b

Points:

(1,1), (2,2), (3,3)

J(theta 0.5) = (1/(2m))[(0.5-1)^2 + (0.5-2)^2 + (0.5-3)^2]

= (1 / (2\*3)) (3\*5)

= (3\*5)/6

= 0.58

J(theta 0) = (1/(2m)) (1^2 + 2^2 +3^2)

= (1/6) \* 14

= 2.3

Cost Function - Intuition II

Gradient Descent

Assignment

a := b --> a := a+1

Truth Assertion

a = b --> a = a+1 (wrong)

Gradient Descent Intuition

As we approach a local minimum, gradient descent will automatically take smaller steps.

So, no need to decrease theta over time.

## Linear Algebra Review

### Matrix and Vectors

#### Matrix

4x2 Matrix 2x3 Matrix

Dimension of Matrix: number of rows X number of columns

Matrix Elements (entries of matrix)

A = Aij = “I, j entry” in the ith row, jth column  
 A11 = 1402; A12 = 191

#### Vector: An nx1 Matrix

y = n = 4

4 Dimensional Vector

y1 = 460 y2 = 232

### Matrix Math

#### Addition

+ = =

*Can only add matrix of same size*

#### Multiplication (Scalar (real number))

X = =

*3x2 matrix multi 2x1 matrix = 3x1 matrix*

Subtract

Divide

#### Matrix Matrix Multiplication

= (see below with logic)

= =

= =

#### Matrix Vector Multiplication

= =

#### Housing price trick with Matrix:

House sizes: 2104, 1416, 1534, 852

h Θ (x) = -40 + 0.25x

X = =

Prediction = Data Matrix \* parameters

*This is a lot faster and simpler code than for loop.*

More complex example of this:

House sizes = 2104, 1416, 1534, 852

3 competing hypotheses:

1. h Θ(x) = -40 + 0.25x
2. h Θ(x) = 200 + 0.1x
3. h Θ(x) = -150 + 0.4x

X =

#### Inverse

1 = “identity”

Examples:

3 (3-1) = 1 3-1 = 1/3

If A is an m x m matrix (square matrix), and if it has an inverse,

AA-1 = A-1 A = 1

= = = I2x2

#### Transpose

A = AT =

Let A be an m x n matrix, and let B = AT.

Then B is an n x m matrix, and Bij = Aij : B12 = A21