**What is Machine Learning?**

Is a science of getting computers to learn without being explicitly programmed.

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

**Different types of learning algorithms**

In supervised learning, the idea is we're going to teach the computer how to do something. Whereas in unsupervised learning, we're going to let it learn by itself.

You might also hear other ghost terms such as reinforcement learning and recommender systems.

**Supervised machine learning**

Supervised Learning refers to the fact that we gave the algorithm a data set in which the, called, "right answers" were given. task of the algorithm was to just produce more of these right answers.

By regression problem, I mean we're trying to predict a continuous valued output.

The term classification refers to the fact, that here, we're trying to predict a discrete value output.

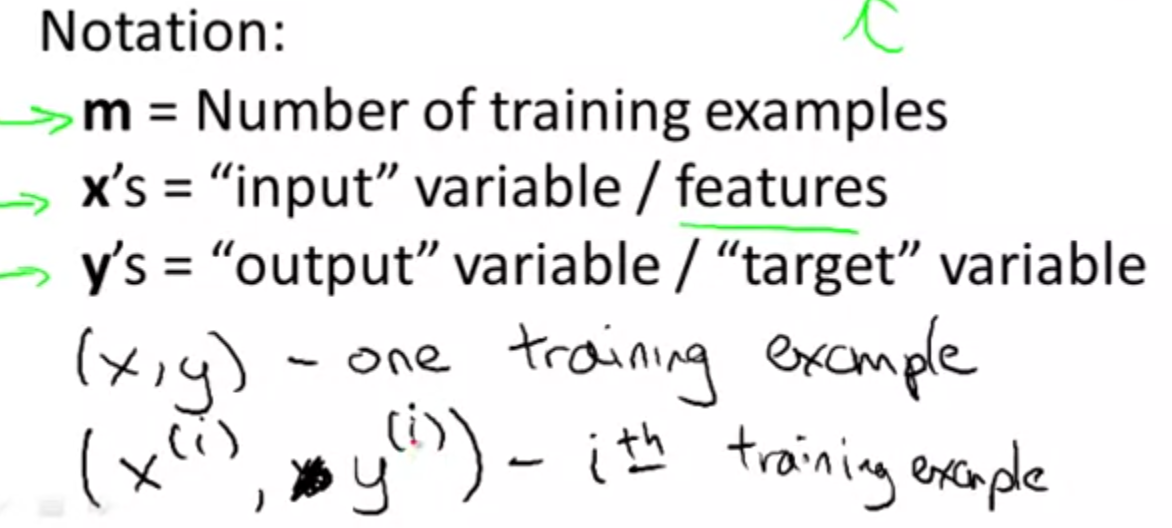
Nota : En la regresión, tratas de encontrar la función que te permite preveer nuevos valores a partir de los valores conocidos. En la clasificación, tratas de estudiar la relación causa-efecto y asocias diferentes causas a diferentes efectos.

**Unsupervised machine learning**

In the clustering algorithm, we're given the data set and we're not told what to do with it and we're not told what each data point is. Instead we're just told, here is a data set. Can you find some structure in the data?

Given this data set, an Unsupervised Learning clustering algorithm might decide that the data lives in different clusters.

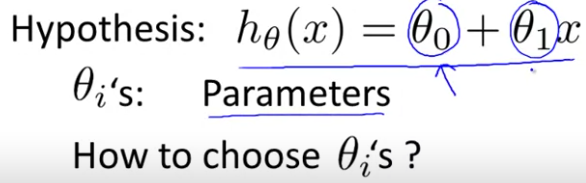
**Linear regression with one variable**



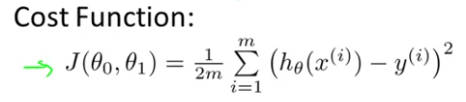
We use the learning algorithm to stablish a hypothesis h from a training set.

We say h maps from x’s to y’s.

**Linear regression with one variable - Cost function**

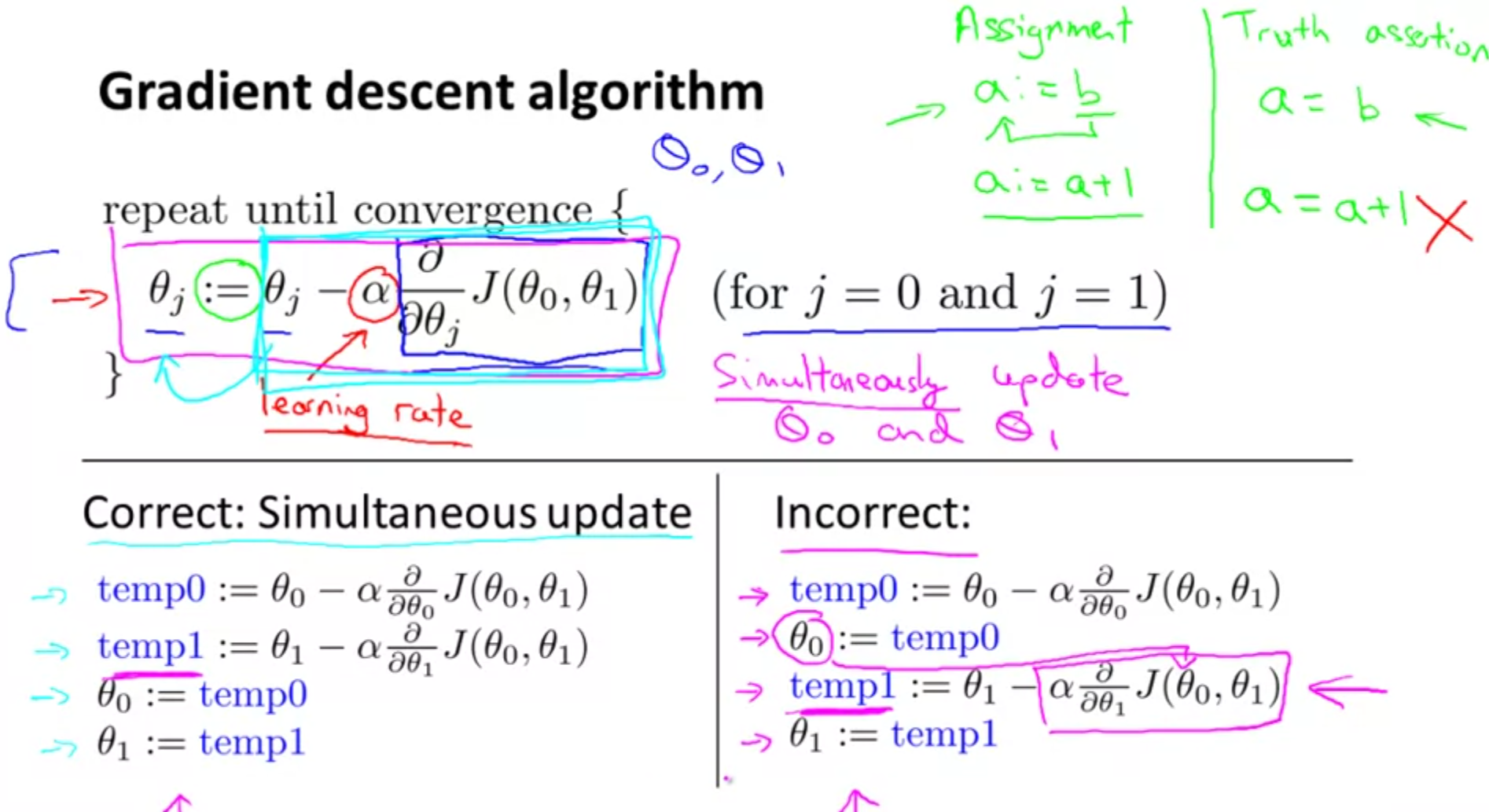


The linear regression consists in obtaining the parameters that minimise the difference (error) between y and h(x) for all the training examples. We express this through a cost function, which indeed consists in the mean value of the square error.

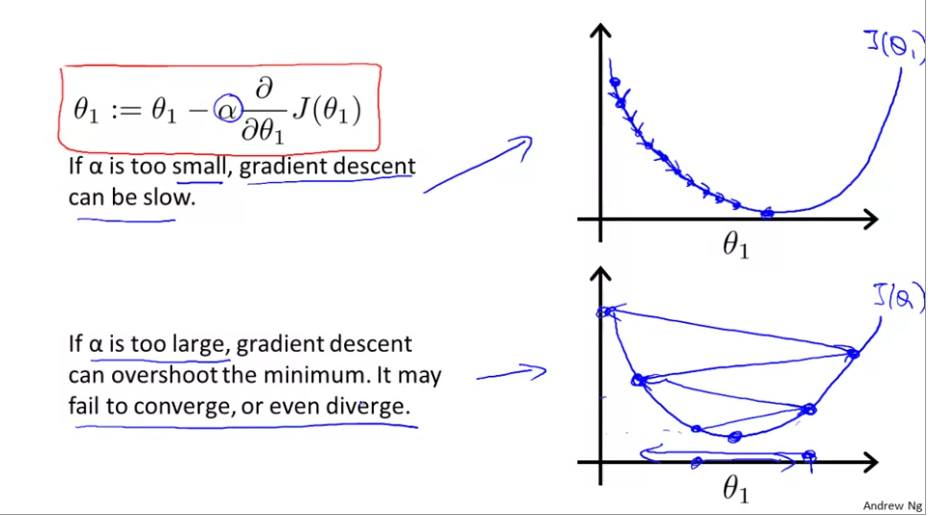


**Finding the parameters that minimise the cost function - Gradient descent**

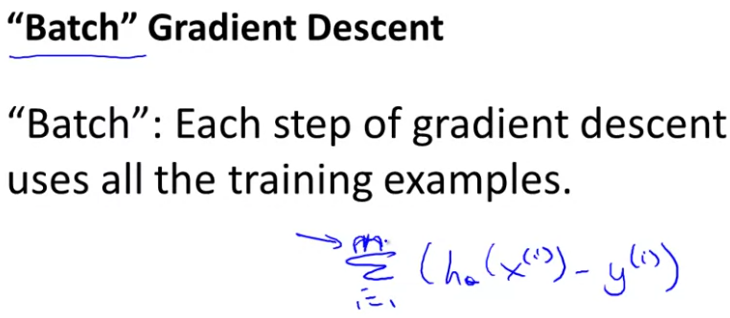
It is an iterative method that starts from some arbitrary parameters and keeps modifying them until we reach a minimum.



The learning rate refers to the length of the steps we take each time.



Note: The gradient descent can have problems if there exist local minimums, as you can end up stuck at a local minimum and not at the maximum minimum. However, the cost function for linear regression always converges, as it only presents one only minimum.



**Linear Algebra Review**

A matrix is a two-dimensional array of dimensions m x n, where m is the number of rows and n is the number of columns. An element of a matrix is indexed as Aij, where i refers to the number of the row and j to the number of the column.

We define a vector as a nx1 matrix, also known as column vectors. An element of a vector is addressed as yi or y[i].

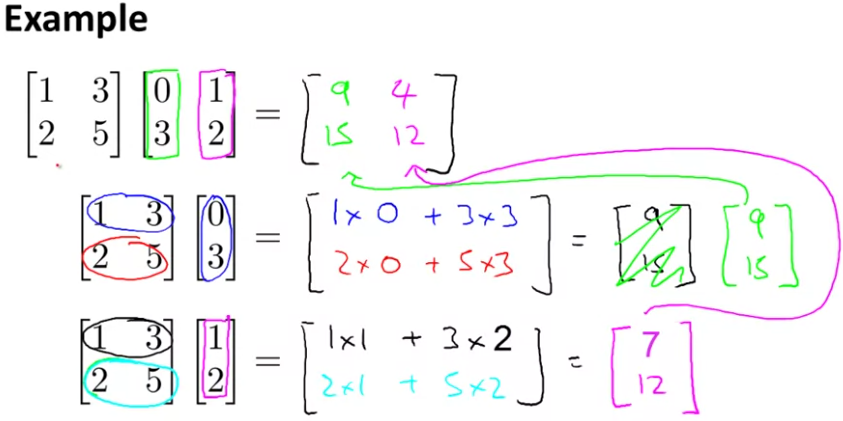
Usually the indexes begin by 1, specially in maths, but in some programming languages, they begin by 0.

We usually use capital letters to refer to a matrix (matrix A,B,C,D…) and normal letters to refer to vector (vector a,b,x,y…).

Matrices are subtracted and added element by element, and we can only do these operations with matrices of the same dimension.

If we divide or multiply a matrix by a scalar, all the elements of the matrix are affected by that operation.

An example of matrix multiplication:

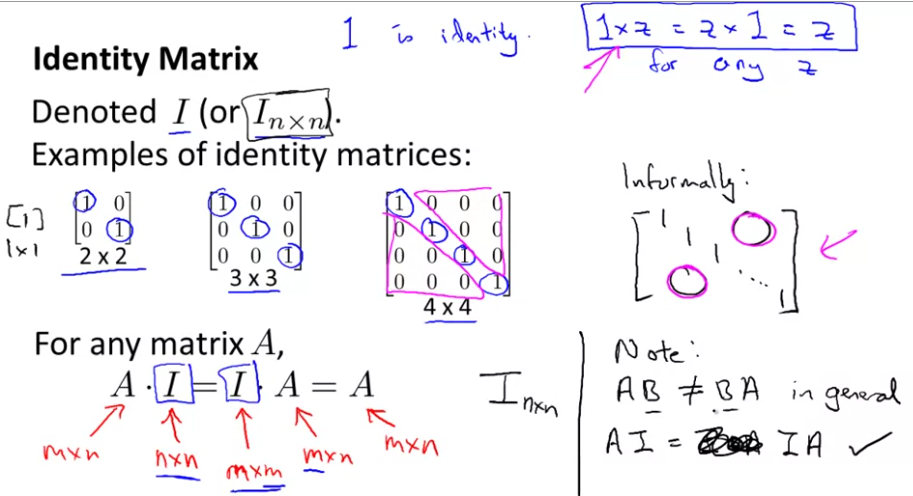


To multiply matrices, the number of columns of the matrix on the left must be equal to the number of rows of the matrix on the right. The resulting matrix will have the same name of row than the first matrix and the same number of columns than the second matrix.

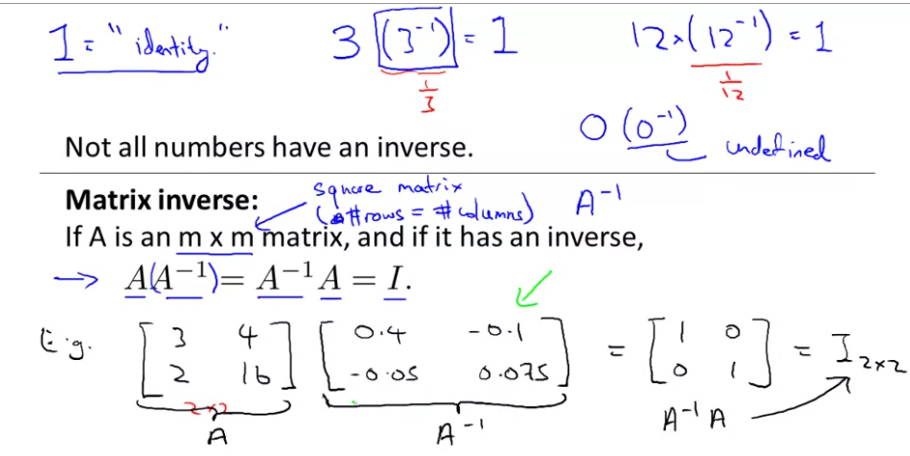
The matrix multiplication is normally not commutative, which means that AxB ≠ BxA.

The matrix multiplication is associative, which means (AxB)xC = Ax(BxC).

The identity matrix is a matrix with its diagonal filled with ones and the rest of elements equals to 0.



The identity matrix is what you obtain when you multiply one matrix by its inverse. IMPORTANT: Only square matrices can have inverse !!



The matrix transpose consists in changing the rows for the columns. An example:

