Abdelrahman Rezk

Web Developer

NLP & ML student

AOU University

What is machine learning
Supervised learning
Unsupervised learning
Cost function
Gradient Descent
Linear Algebra

Machine Learning

ال machine learning هو ازاى تخلى الاله الى عندك تتعلم من نفسها من غير ما تديها برامج او حاجات محددة يعنى بمعنى اصح انت بتدربها على بعض الحاجات وهى المفروض تطور نفسها بنفسها من خلال التجارب التي بتمر عليها وتتعلم من الماضى زى الانسان كده معننا مبنتعلمش بس قشطه.

By Tom Mitchell:

He says, 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

انى الكومبيتر يتعلم من خلال التجارب التي مر عليها زى مثلا لو بنعمل clarify emails as spam or not الاول الكومبيتر بيتعلم انتا ازاى بتصنيف ال emails اما كده او كده بعدها تبتدي تجرب بعض ال tasks الجديدة الى هو مشفهاش ومن خلال ال performance بتاعه الى هو الميلات الى صنفها صح واللى مصنفهاش صح.

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 machine learning problem can be assigned to one of two broad classifications:

Supervised learning and Unsupervised learning.

ال machine learning بيتقسم ل تلت حجات مهمه

1- ال supervised learning وده ببساطة انى بديله داتا وبديله بعض ال labels الخاصة بالداتا ديه عشان يبتدى انه يتعلم منها ويقدر ياخد اكشن بعد كده وبرضه انا بتابع الكلام ده مع ال model

● Linear Regression

- The term Supervised Learning refers to the fact that we gave the algorithm a data set in which the,called, "right answers" were given.

عشان فى الاول انتا بتديله داتا بقيم حقيقية زى اسعار الشقق ك output مقابل مميزات فى الشقق وهو بعدها بيبتدى يتعلم مع نفسه من الحاجات ديه عشان لما تديله داتا جديده يقدر يفيدك ويقولك سعر شقة ما كام.

اولا كده بالنسبه لل machine learning الجزء الاول فيه الا وهو ال supervised learning يعتمد فى التوقع بتاعه على جزئين مهمين الا وهما ال input and output وبيحصل فى ال supervised حاجتين مهمين اما انى محتاج اعمل classification or regression

ال regression الا وهو التوقع وده ببساطة يشتغل على فكره انه بياخد بيانات ليها علاقه ببعض البيانات ديه عباره عن features وهو يبتدى يشتغل على الكلام ده كويس جدا وذاكره ويتعلم منه على سبيل المثال مثلا اسعار الشقق وعندى حوالى 10 الاف raw من ال data الخاصه بالشقق متسجل فيها مكان الشقه والمساحه وغير بقا من الميزات وفى المقابل ال input بتاع ال features ديه بيكون فيه output لل model برضه الى هو اسعار كل شقه منهم اد ايه فالموديل بيشتغل على الكلام ده وفى الاخر بيحصل عملية test انك بتشوف بقا شقه جديده بمواصفات جديدة وتديها للموديل فيبدأ انه يعمل predict ليها انى سعرها اد كذا.

ال regression بيتوقع النتايج ديه بطرق مختلفه ممكن تكون linear او curved وغيره

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

بعض التطبيقات المختلفة لل regression حاجه زى اسعار المنازل والأرصاد الجوية والمشتريات بالنسبة لعميل جديد بناء على مواصفات العملاء الى من النوع ده.

بينما بقا ال classification هو انه بيحاول يقسم الحاجة لجروبات مختلفه زى الصور مثلا وانه يحاول يعمل تصنيف لصور القطط او الكلاب فلما تيجى تديله صوره جديده يقولك لا ديه صوره كلب او صوره قطه وهكذا.

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

يعنى فى الاخر انا بحاول انى اتوقع قيم منفصلة عن بعضها لكن ال regression هو عباره عن انى بحاول اتوقع سعر حاجه ممكن تتغير مع تغير ال data الى داخله ليا انما ال classification هو توقع حاجه معينه مثلا من set او مثلا التوقع ده بيكون binary اما يحصل او لا.

So this is an example of a supervised learning algorithm. And it's supervised learning because we're given the, quotes, "right answer" for each of our examples. Namely we're told what was the actual house, what was the actual price of each of the houses in our data set were sold for and moreover, this is an example of a regression problem where the term regression refers to the fact that we are predicting a real-valued output namely the price.

2- ال unsupervised learning وده انا بديله داتا وهو المفروض يتعلم منها من غير ما اديله اى label او حاجه يتعلم منها هو بيبتدى يقسم الداتا ويتعلم مع نفسه.

انتا بتديله الداتا وهو بيحاول يلاقي structure مختلفة فى الداتا ديه بمعنى بيحاول انه يقسم الداتا لجروبات بناء على الحاجات الى ليها علاقه ببعض.

So this is Unsupervised Learning because we're not telling the algorithm in advance that these are type 1 people, those are type 2 persons, those are type 3 persons and so on and instead what were saying is yeah here's a bunch of data. I don't know what's in this data. I don't know who's and what type. I don't even know what the different types of people are, but can you automatically find structure in the data from the you automatically cluster the individuals into these types that I don't know in advance? Because we're not giving the algorithm the right answer for the examples in my data set, this is Unsupervised Learning.

Unsupervised Learning

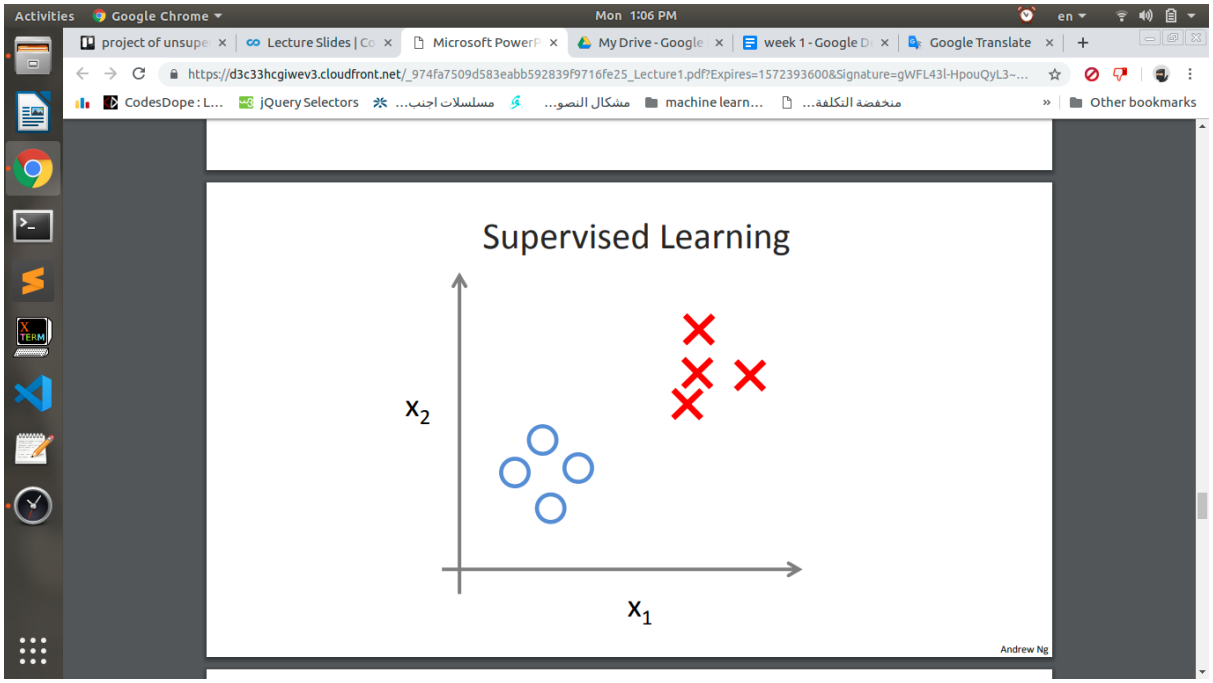
Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.

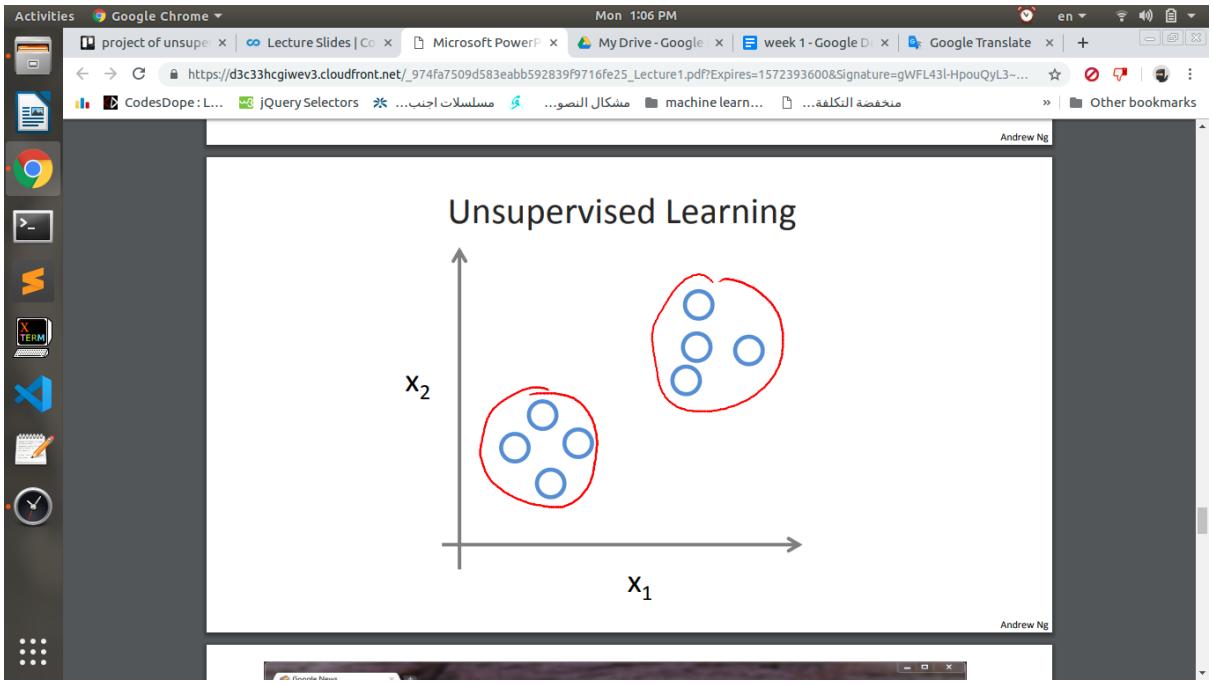
We can derive this structure by clustering the data based on relationships among the variables in the data.

With unsupervised learning there is no feedback based on the prediction results.

Clustering: Take a collection of 1,000,000 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, roles, and so on.

Non-clustering: The "Cocktail Party Algorithm", allows you to find structure in a chaotic environment. (i.e. identifying individual voices and music from a mesh of sounds at a cocktail party).


Supervised Learning
x_2
x_1
Andrew Ng


Unsupervised Learning
x_2
x_1
Andrew Ng

Training set of housing prices (Portland, OR)

Size in feet² (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

m = 47

Notation:

- **m** = Number of training examples
- **x**'s = "input" variable / features
- **y**'s = "output" variable / "target" variable

(x, y) - one training example

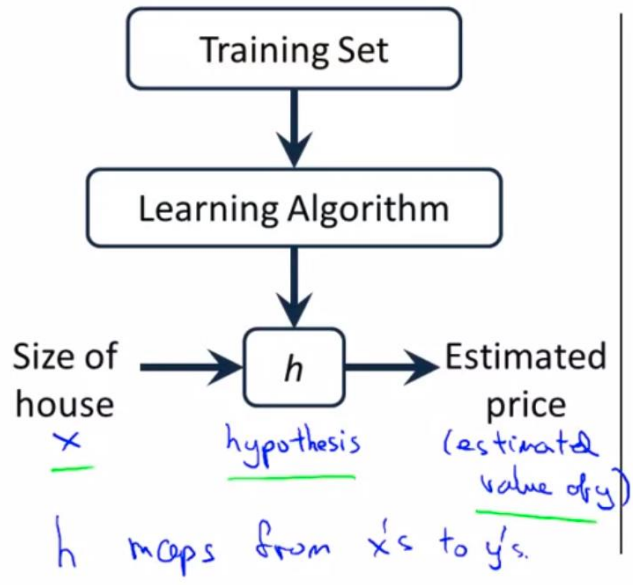
$(x^{(i)}, y^{(i)})$ - i^{th} training example

$x^{(1)} = 2104$

$x^{(2)} = 1416$

$y^{(1)} = 460$

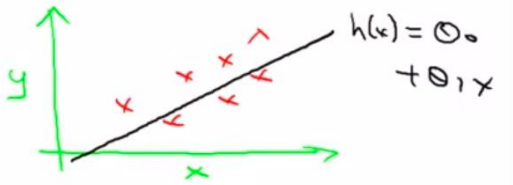
Andrew Ng



How do we represent *h* ?

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

Shorthand: $h(x)$

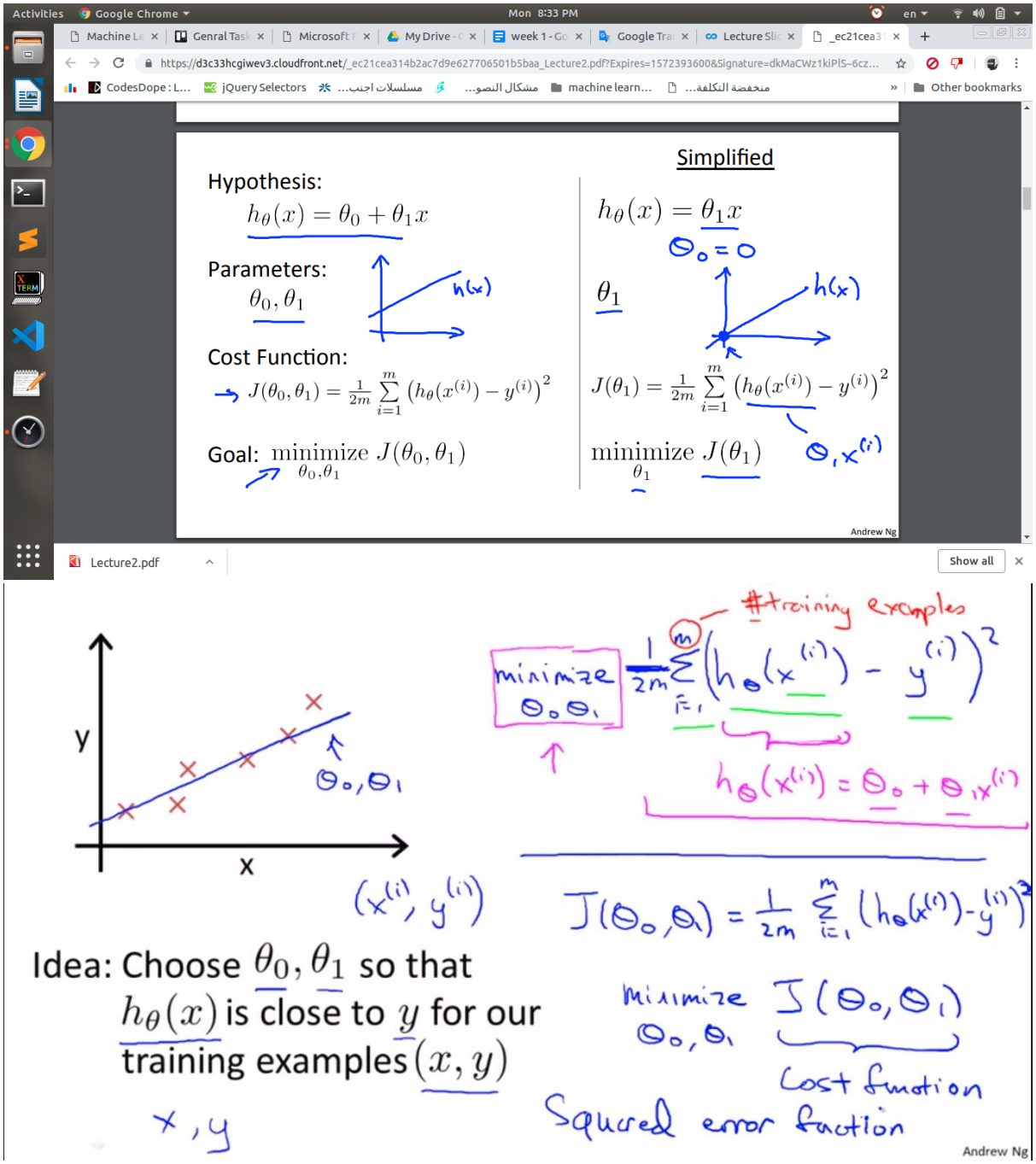


Linear regression with one variable. (x)

Univariate linear regression.

↳ one variable

Andrew Ng

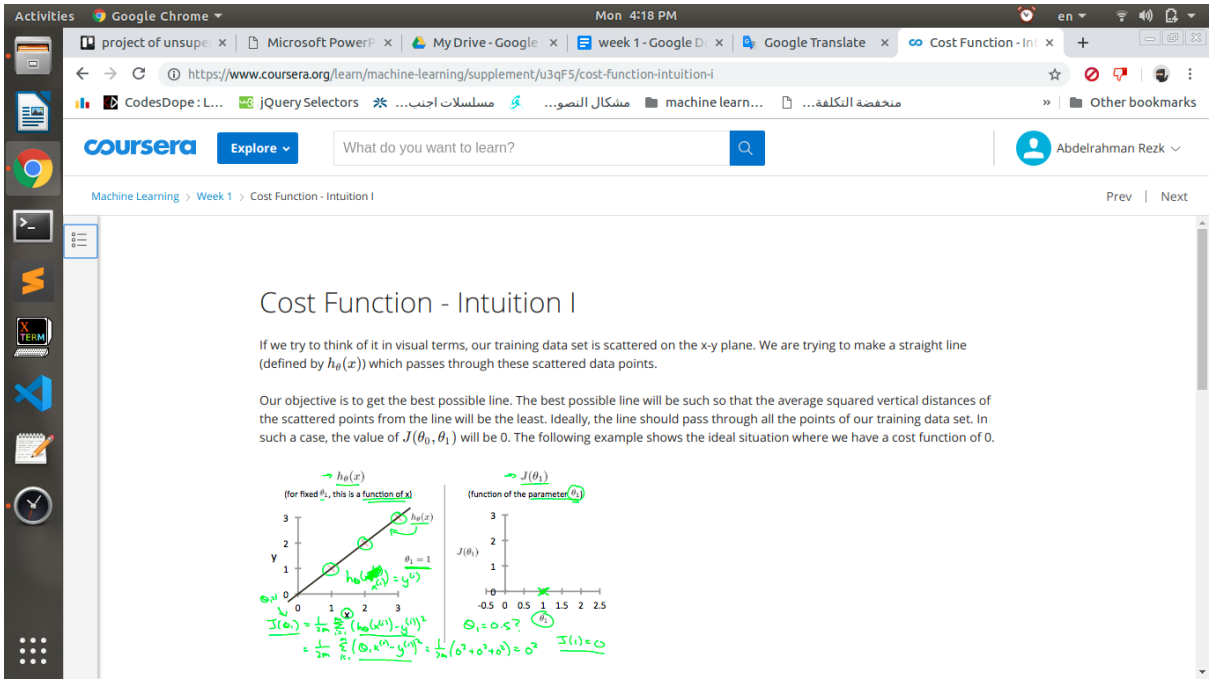


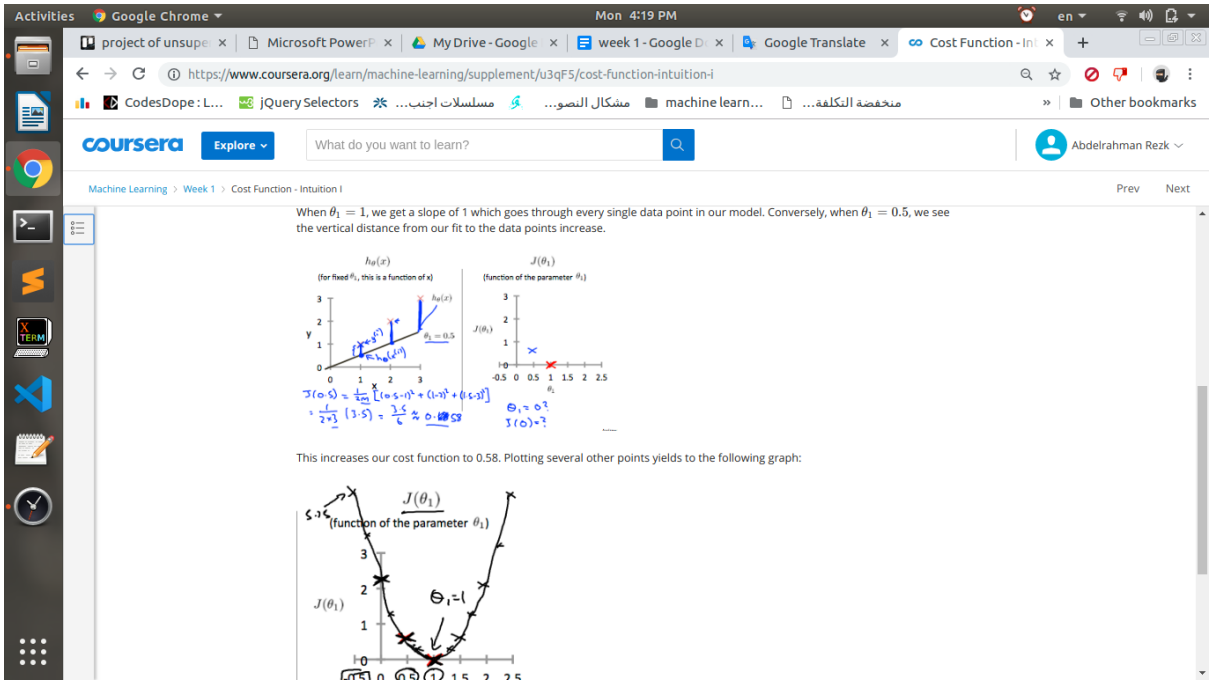
ازاى بقا بقدر احسب ال cost function الا وهى نسبة ال error الى عندى والى هى بتكون الفرق بين ال line الى انا برسمه والنقط ال actual بمعنى الفرق بين ال actual and predictive الى هو y ال h(x)s

وده بيخليني اعرف نسبة ال error عندى اد ايه ومن خلال ده بحاول انى قللها على قدر ما فقدر.

h(x) is consider about theta 0 + theta1*x(i) and i is th-ith raw in our data.

طبعا فيه فرق بين حاجتين مهمين الا وهما ال h(theta)d الى هو الرقم الى انا بختاره لل theta وبعدها بحسب بقا الحاجه التانيه وهى J(theta)f الى هى ال cost function من اختيارى لل theta بالرقم ده طلع بكذا.

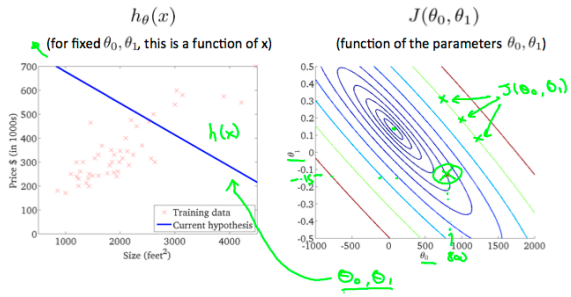




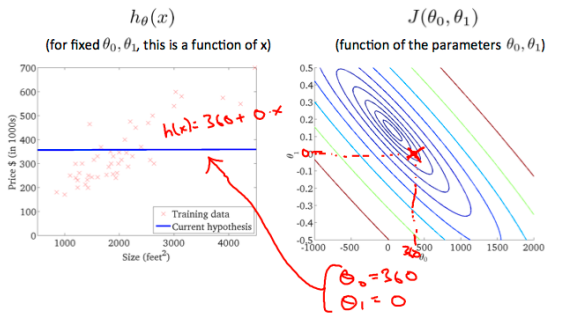
هنا انا مخترتش الا 1 theta فقط وخليت 0 theta ب 0

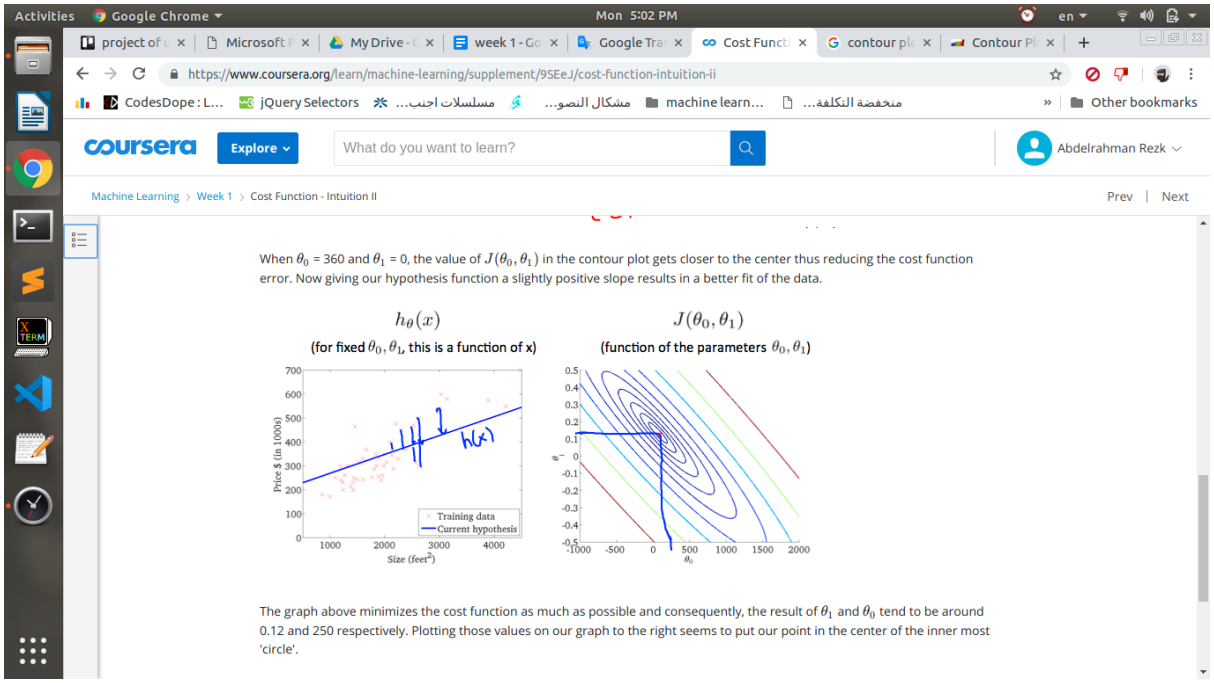
Cost Function - Intuition II

A contour plot is a graph that contains many contour lines. A contour line of a two variable function has a constant value at all points of the same line. An example of such a graph is the one to the right below.



Taking any color and going along the 'circle', one would expect to get the same value of the cost function. For example, the three green points found on the green line above have the same value for $J(\theta_0, \theta_1)$ and as a result, they are found along the same line. The circled x displays the value of the cost function for the graph on the left when θ_0 = 800 and θ_1 = -0.15. Taking another h(x) and plotting its contour plot, one gets the following graphs:

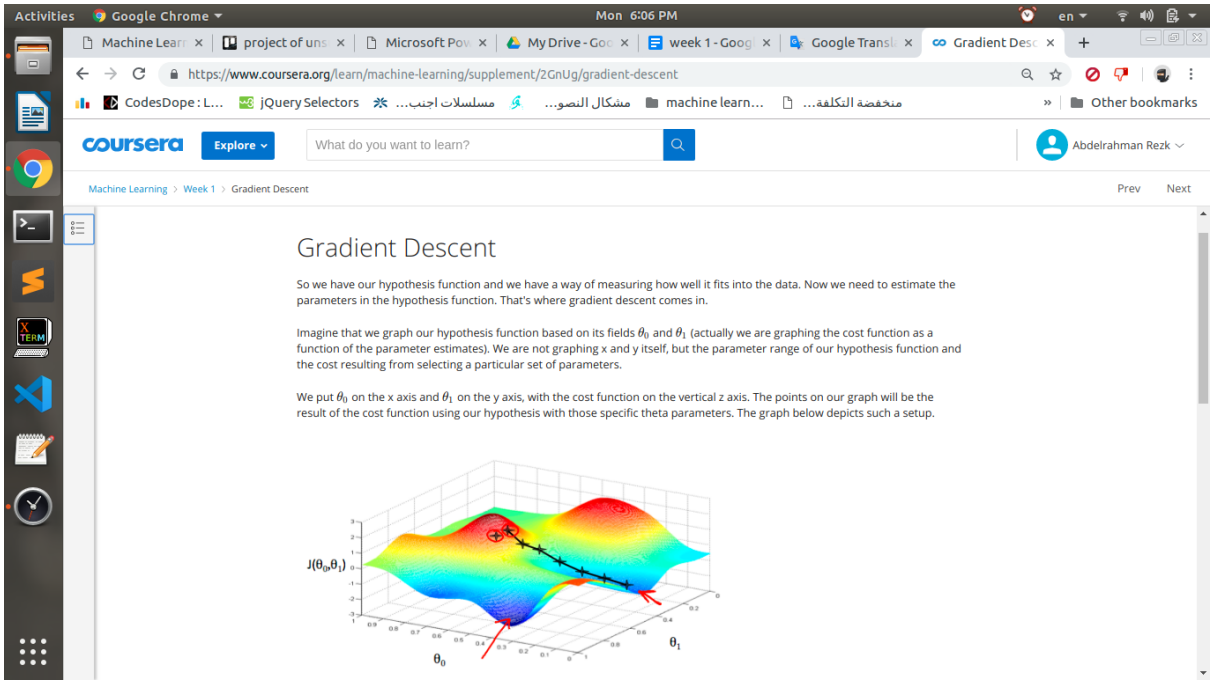


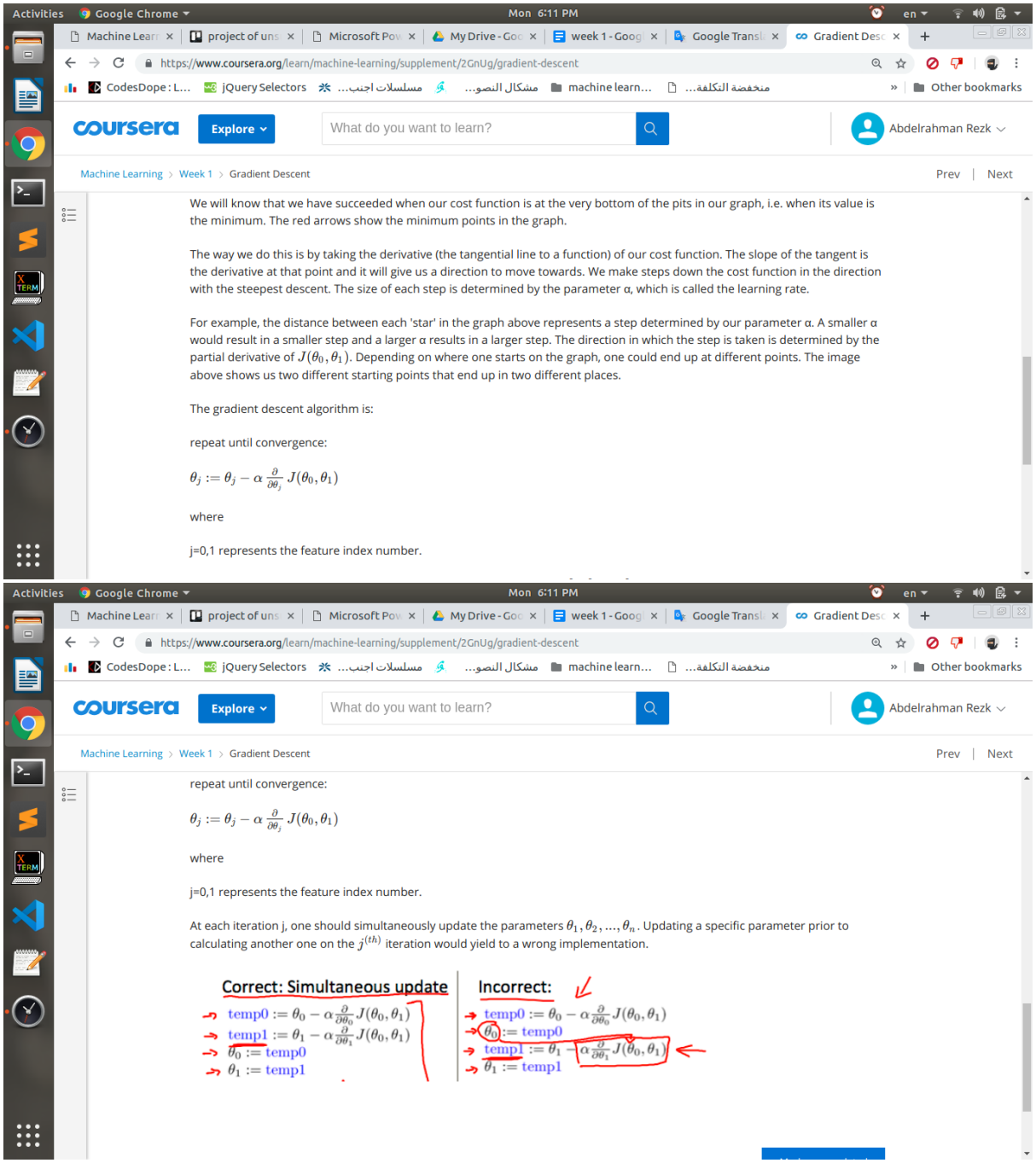


طيب دلوقتي انا عرفت ازاى بحسب ال cost function بعد اما اختار قيم theta 1 and theta 0 واشتغل على المعادله بتاعت ال J(theta0, theta1)d ولكن انا لو عندى داتا كتيره مش هقعد اشتغل manually انى اجرب كل نقطه و اشوفها بتعمل ايه معايا بناء على قيم ال theta إلى اختارتها وهنا بيجى دور

Algorithm called gradient descent

وهو عبارة عن خوارزمية بتحاول انها تقلل نسبة الخطأ على قدر ما تقدر.

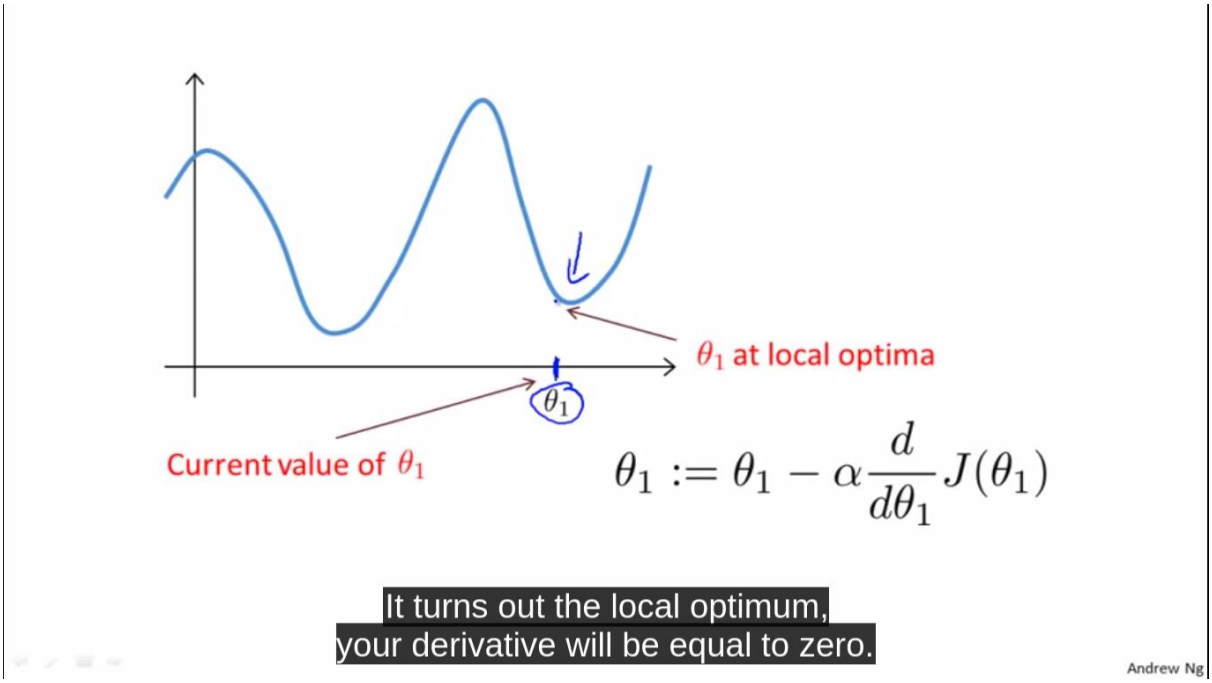


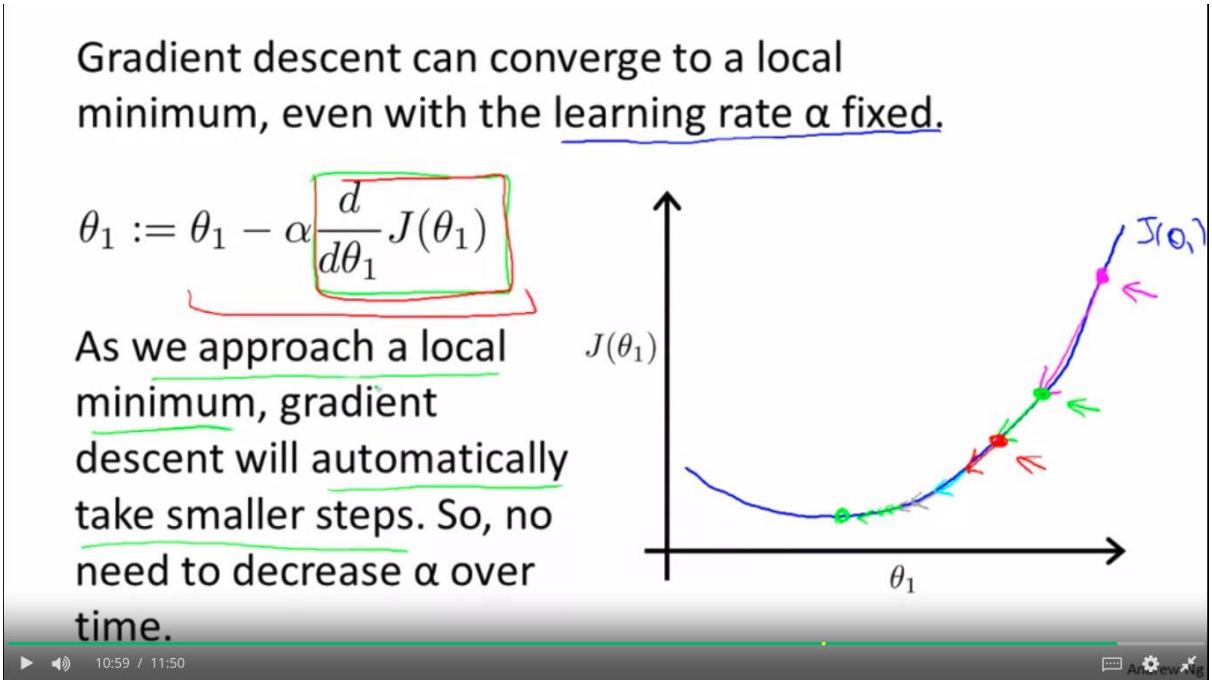


ال derivative part الى فى المعادلة ده هو اللى يساعدنى وبيخلينى اعرف انا هامشى ازاى عشان اوصل لل minimum error وده من خلال انى لما بختار نقطة ما بروح اشوف هى بتمس الخط ازاى واشوف ال slope بتاعها اما positive or negative ومن خلاله بعرف انا هامشى ازاى.

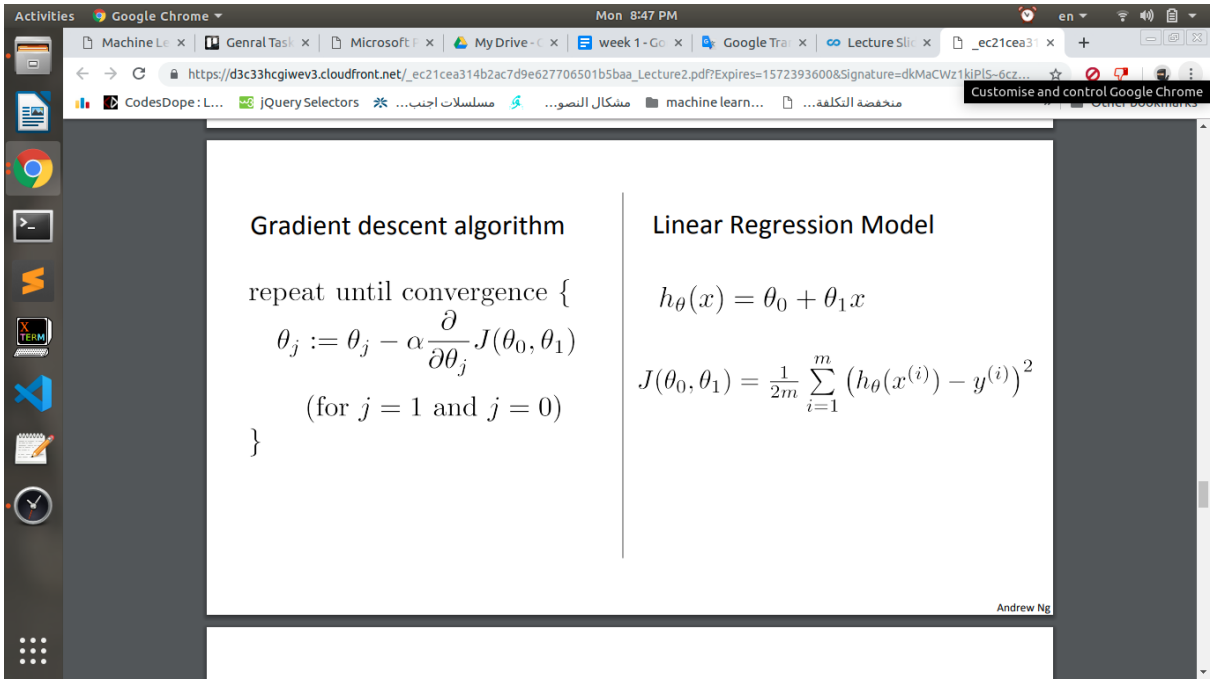
لما ال gradient descent و ال theta بتاعتى بتوصل لل minupmum optimal bottom الى هو بتقلل ال error لأكثر حاجه ممكنه ساعتها ال slope الناتج من ال derivative بيكون ب 0 وبعد كده قيمه ال theta مش هتتغير.

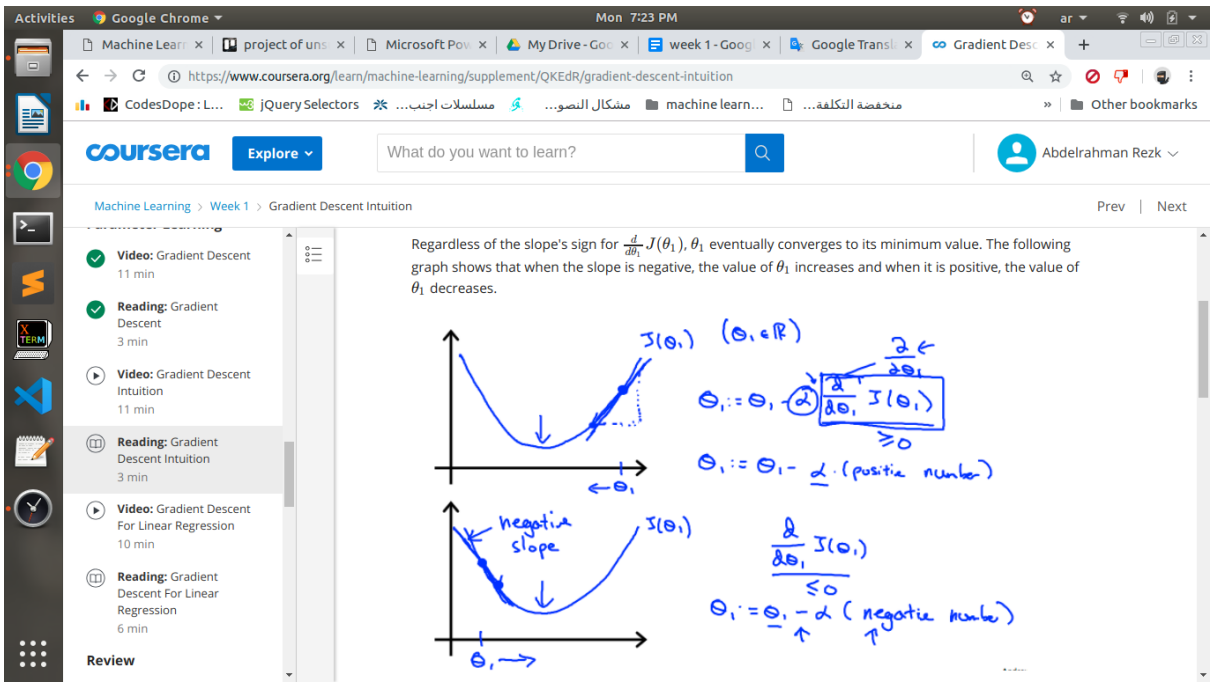
فى ال gradient descent انتا كل اما بتاخد خطوة فى اتجاه انك تقلل ال error الخطوه نفسها بتقل لانك فى كل مرة بتضرب ال learning rate فى ال derivative الجديدة الناتج عن ال slope الجديد وهكذا.

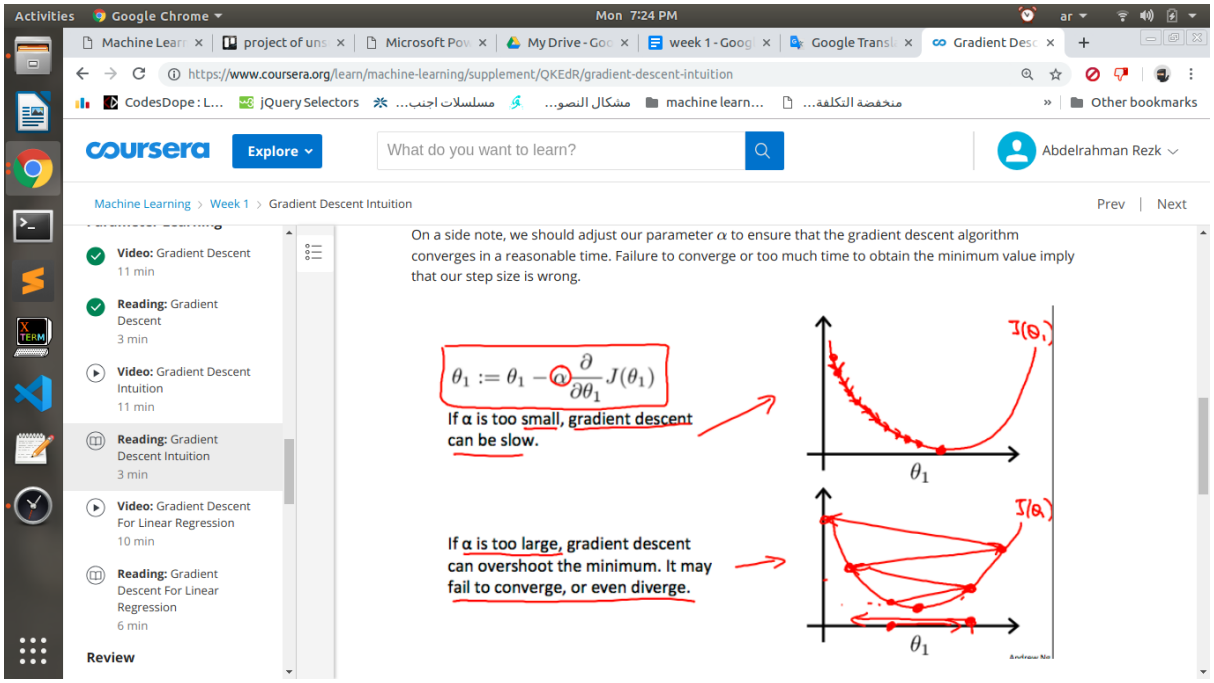


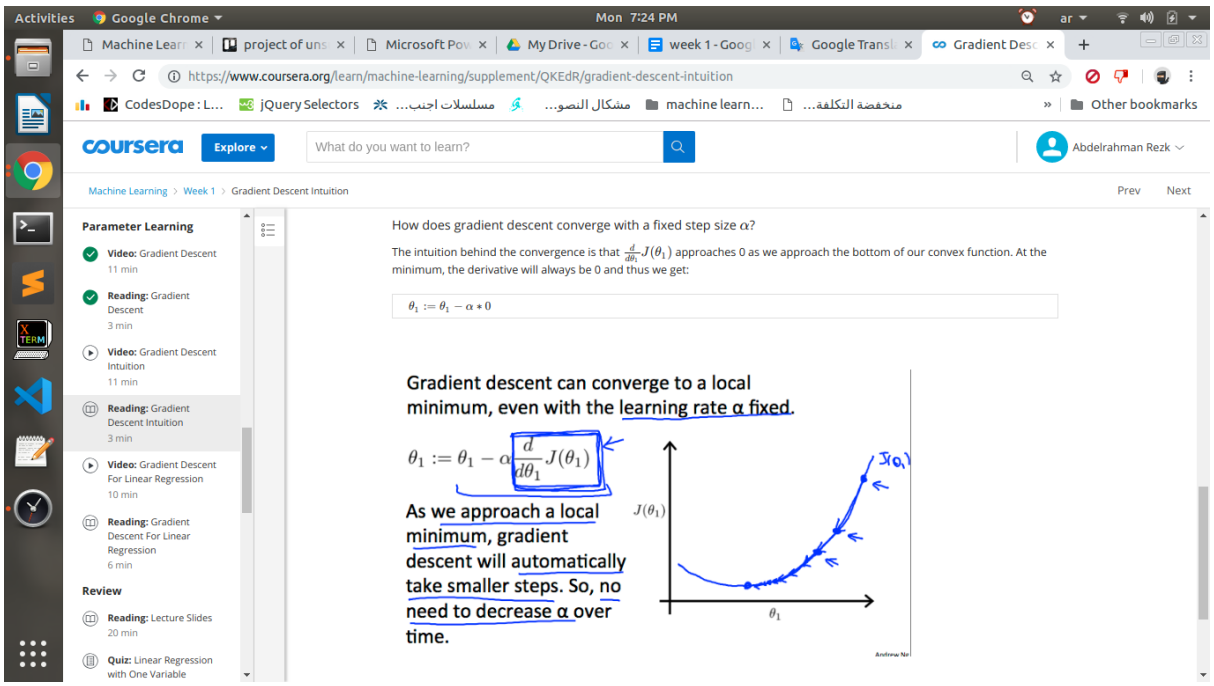


انتا هنا مش محتاج انك تقلل ال learning rate لانك كده كده فى كل مره الخطوه نفسها بتقل بسبب انى قيمة ال derivative بتقل لذلك ال learning rate can be fixed

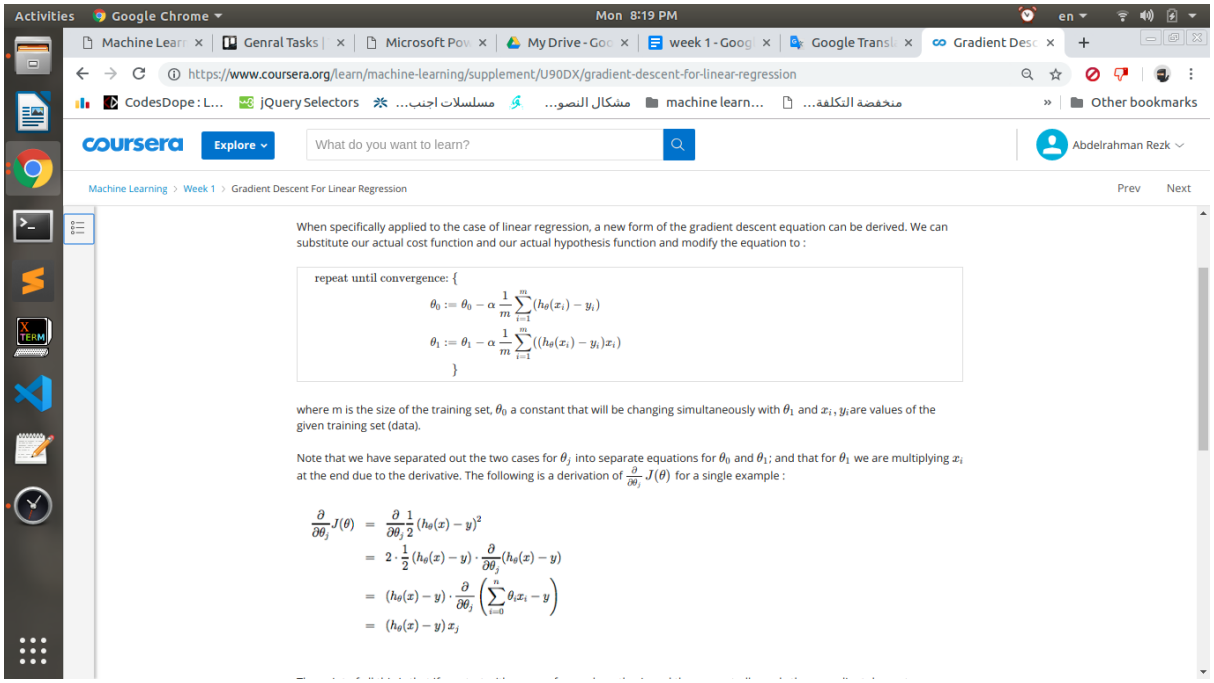

Gradient descent algorithm
repeat until convergence {
$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$
(for $j = 1$ and $j = 0$)
}
Linear Regression Model
$h_\theta(x) = \theta_0 + \theta_1 x$
$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$
Andrew Ng


Regardless of the slope's sign for $\frac{d}{d\theta_1} J(\theta_1)$, θ_1 eventually converges to its minimum value. The following graph shows that when the slope is negative, the value of θ_1 increases and when it is positive, the value of θ_1 decreases.
negative slope
Machine Learning › Week 1 › Gradient Descent Intuition
Prev | Next
Video: Gradient Descent 11 min
Reading: Gradient Descent 3 min
Video: Gradient Descent Intuition 11 min
Reading: Gradient Descent Intuition 3 min
Video: Gradient Descent For Linear Regression 10 min
Reading: Gradient Descent For Linear Regression 6 min
Review



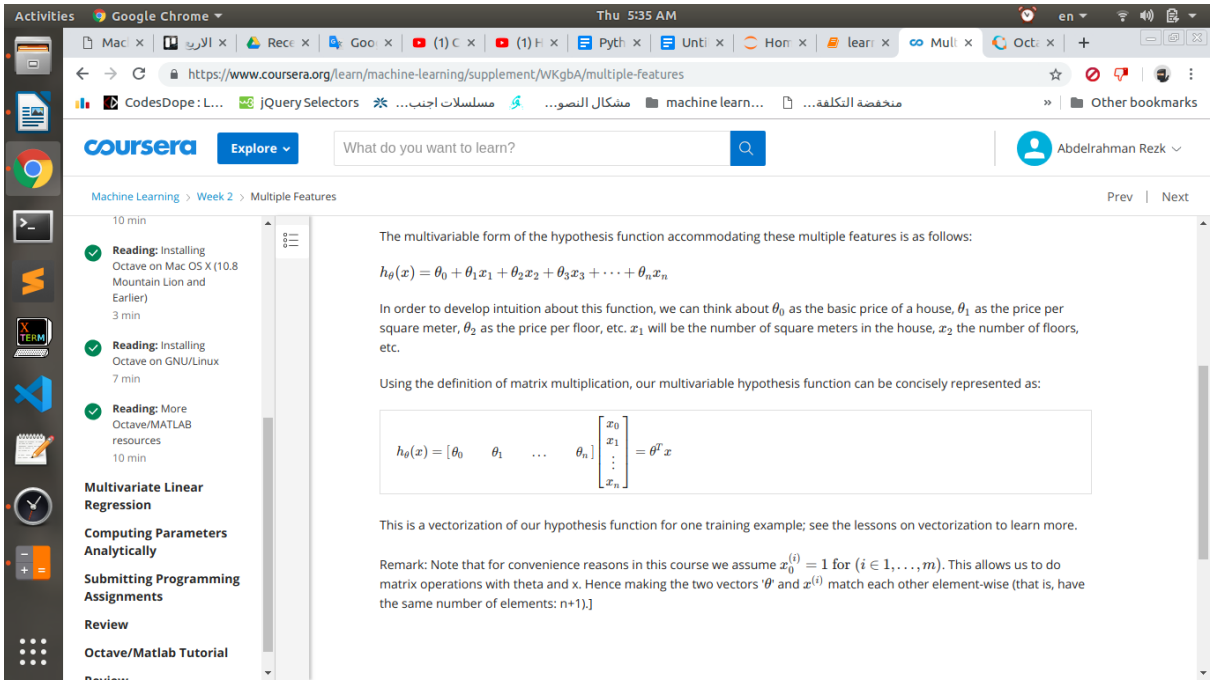


ازاى بقا انى احط ال gradient descent with cost function to fit best line in your data



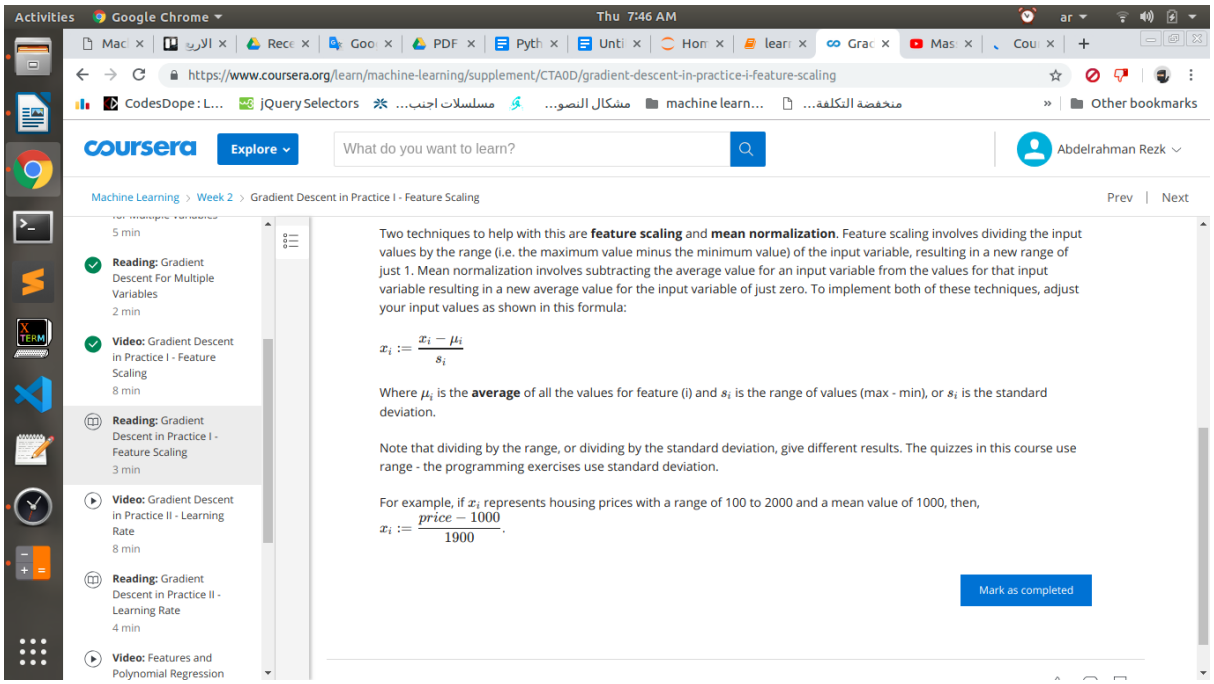
الجزء الكبير فى ال linear regression هو الرياضه او المعادلة الخطيه ومعناها هو العلاقه بين متغيريين حين يكون كلا من هذان المتغيران لهما اوس = 1

استخدام الدوال فى الرياضه مع الماشين ليرننج بيكون على حسب انا شغال على ايه فمثلا ممكن استخدم المعادله الخطيه لو هى دايما فى زياده لكن لو قدام مثلا الحاجه لما بتزيد بتقل بعد شويه فممكن تكون معادله من الدرجه التانيه وهكذا.



Data Rescaling

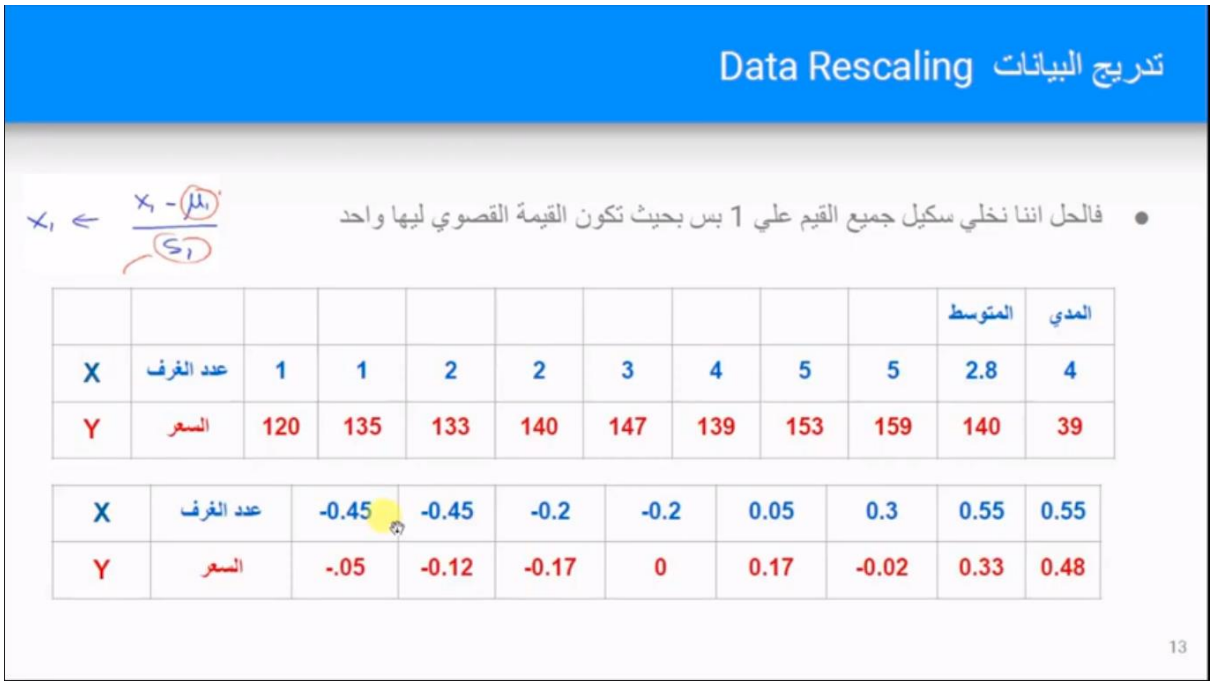
هو انى احاول اخل جميع الداتا الى عندى وفيها ارقام ما بين ال -1 و 1 وده بيساعدنى انى الجراف نفسه بيكون معقول وليه علاقه ببعضه على عكس اما يكون عندى قيم صغيره جدا وقيم كبيره جدا والصورتين الى تحت دول بيوضحوا قبل وبعد ال scaling.

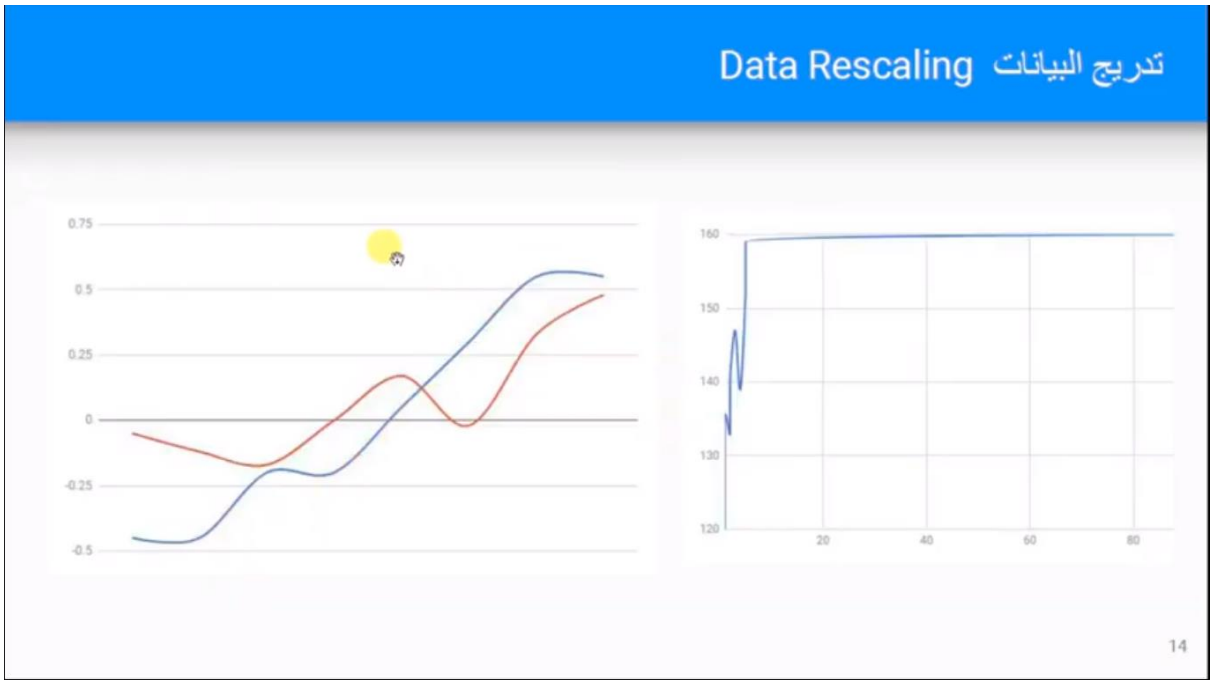


Normal Equation

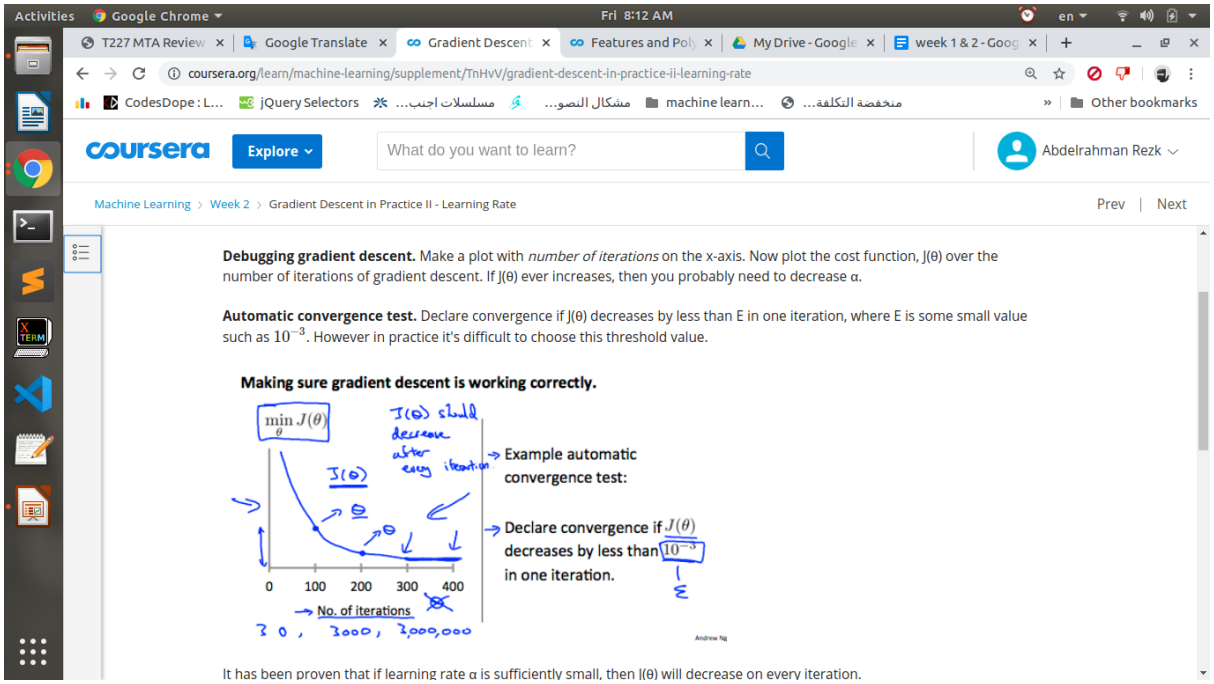
ديه بتحل مشكله انى افترض قيم للثيتات ومن خلالها بقدر اوصل لقيم الثيتا الى انا عايزها عشان اشتغل على ال gradient descent واجيبه فى خطوه واحده بدل المشكله الى بتواجهنى فى تحديده ال learning rate ولكن المشكله هنا بتكون فى ال inverse العمليه بتاعته نفسها بتاخد وقت كبير جدا كل اما الماتركس تكبر على عكس انى احاول مع ال learning rate رغم انى معرفش عدد الخطوات الى هاخدها قد ايه ولكن ممكن فى range ال 10 الاف وده الى هو عدد ال features اشتغل علطول بال normal question.

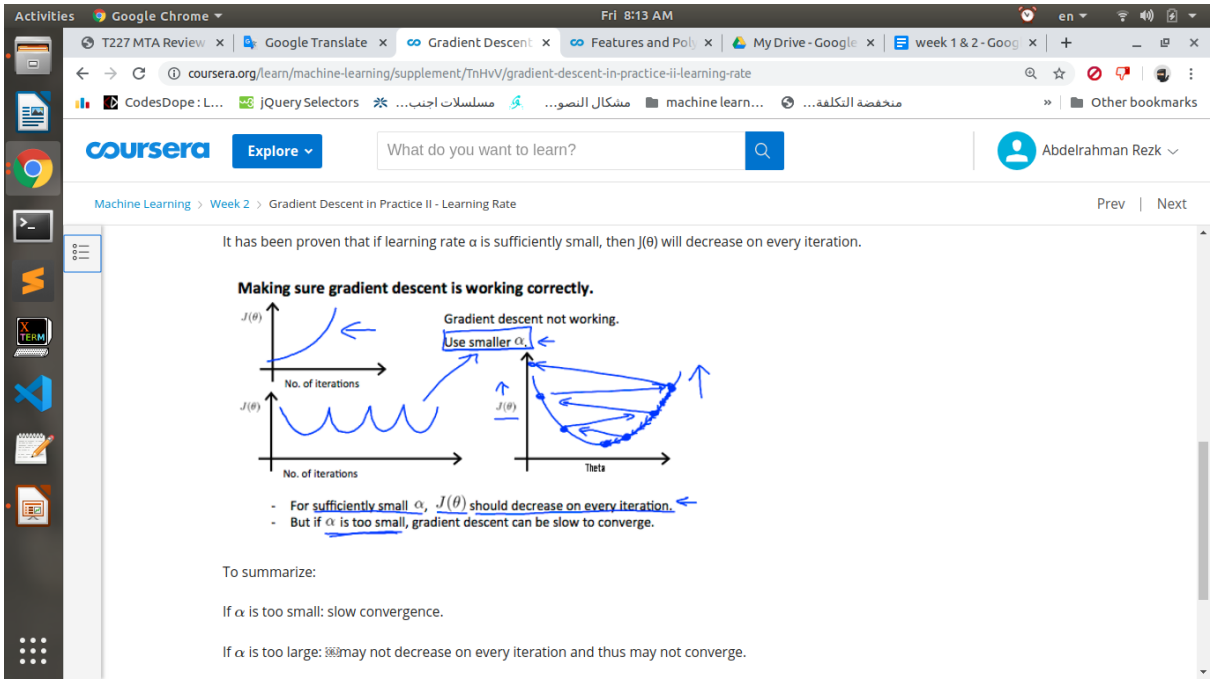
ولازم اخلى بالى لما اجى اجيب ال normal question ممكن لما اجي اجيب ال inverse الاقى الماتركس نفسها singular الى هو ملهاش inverse



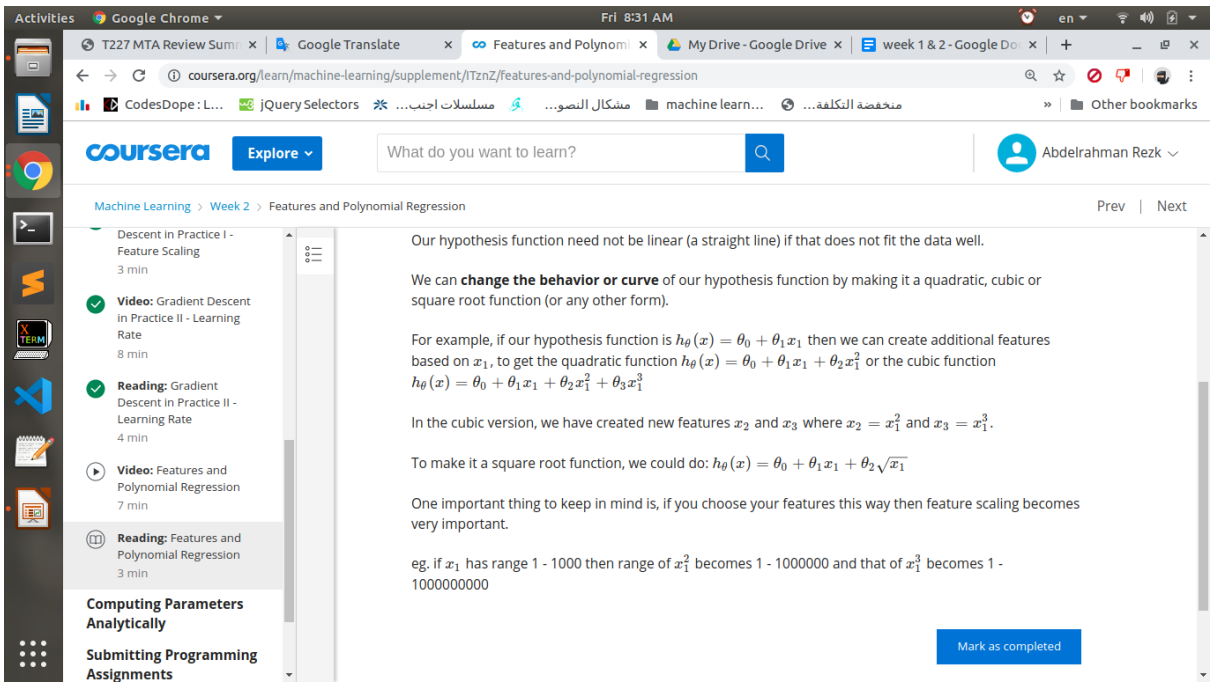


Gradient Descent in Practice II - Learning Rate





إنتا ممكن وانتا شغال لما تشوف ال features تكرر انك تعمل features جديده ممكن تكون تجميع ل لاكتر من feature زى مثلا الطول والعرض بتاع المنزل تخليهم الاتنين فى features واحد وتسميه المساحة وممكن انك ت create feature مش موجود بس يكون ليه علاقه بالحاجة ومؤثر فيها.



ال features الى بتكرر معايا بطريقة متناسبة مع بعضها زى لما قسم حاجه على حاجه او الاتنين يكونوا بيتقسموا على رقم معين غاليا ال features ديه ممكن امسحها كلها واخلي واحد منهم فقط لان يعتبر مفيش غير feature واحد منهم هو الى هاياثر معايا لانهم فى تناسب مع بعضهم وده مش هايساعدنى بحاجه غير وقت و تكلفة على الفاضى يعنى محتاج اخلى بالى وانا شغال من كل حاجه