Abdelrahman Rezk

AOU University

NLP & ML student

بسم الله الرحمن الرحيم

الملخص ده بإذن الله لازم تكون شوفت قبليه اول اسبوعين من Andrew Ng

<https://www.coursera.org/learn/machine-learning>

وفيه مقابل ليه بالعربى مهندس هشام وده بصراحة ربنا يباركله ويجازيه خير على العلم الى بيقدمه والمساعده الجميله وده برده اول اسبوعين فى الكورس فى البلاى ليست بتاعته.

<https://www.youtube.com/watch?v=5_GKIhoXKyI&list=PL6-3IRz2XF5UJE2PbY7UU4SHi7UpV1mXo>

المذكور هنا هايكون جهد شخصى ومحاوله إنى اساعد فى تبسيط المصطلحات والمعادلات مع حل امثله عملية تربط جميع الاجزاء المختلفه خلال اول اسبوعين بإذن الله فالمفروض إنك تكون شوفتهم.

"وهاكون شاكر لو كان هناك دعوة بظهر الغيب"

احنا فى الاخر بنحاول اننا نقلل الفرق بين h(x) الى هى القيمة المتوقعة وال y الى هى القيمه الحقيقية.

**بعض المصطلحات الهامة**

|  |  |
| --- | --- |
| output y | input x |
| 4 | **0** |
| 7 | **1** |
| 7 | **2** |
| 8 | **3** |

m = training examples

عدد الامثله الى هاقدر استخدمها عشان اعمل training للموديل بتاعى فى المثال الى فوق 4

n = number of features

الحاجات المميزه الى بتساعدنى اعمل predict كويس وهنا معنديش الا 1 بس الى هو x

y = actual value or output value

انا فى الاخر بتعامل مع supervised learning فبحاول انى اعمل mapping from input to output معين وده الى انا بخلى الموديل بتاعى يدرب عليه ويحاول يقلل الفرق قدر الامكان بين القيم المتوقعة والقيم الحقيقية.

y^ = h(x) = predicted value

ديه القيمه المتوقعة

∑ summation sign

Θ thetas parameters

قيم ال thetas الاخيره هى الى بتستخدم عشان احاول اعمل predict لحاجه جديدة مشفتهاش قبل كده عن طريق قيم ال thetas الى انا طلعتها من ال training examples .

α alpha called learning rate

ديه الى بتاثر فى تغير تقليل نسبه الخطأ بصورة كبيره او صغيره على حسب قيمتها.

**المعادلات المستخدمة**

**أولا ال Cost function**

J(*θ)* =

vectorized Version of cost function

J(*θ)=*

دلوقت نقدر نطبق على اول حاجه ونعرف انى نفس المعادلتين هايطلعوا نفس الناتج بدل ما اعمل loop ممكن علطول استخدم ال vectorize version ولكن هانطبق على الاتنين ونشوف الدنيا هاتمشى ازاى.

خلينا نفترض انى قيمه theta 0 = 1 and theta 1 = 3 وبما إن معادلة ال h(x) = 1 + 3 X

لازم تكون عارف انى x1  معناها اول مثال فى الداتا يعنى اول row لكن x1  معناها اول feature

فبكده x11  معناها اول feature فى اول مثال وهنا معنديش الا feature واحد لكن 8 training examples .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | H(x) | y | X1 |
| 12 =  1 | 301-300=  1 | h(x) = 1 + 3 X  1+3\*100=301 | 300 | 100 |
| 1 | 1 | 286 | 285 | 95 |
| 1 | 1 | 271 | 270 | 90 |
| 1 | 1 | 241 | 240 | 80 |
| 36 | 6 | 241 | 235 | 80 |
| 121 | 11 | 211 | 200 | 70 |
| 36 | 6 | 211 | 205 | 70 |
| 1 | 1 | 181 | 180 | 60 |
| =  1+1+1+1+36+121+36+1 = 198  Then  J(θ)=  \*  M= 8 training examples so  J(θ)= |

**vectorize version solution of cost function**

J(*θ)=*

Assume = 1 for 8 training examples

حول كل حاجه من المعادله بتاعت ال vectorize ل matrix or vector عشان نقدر نتعامل معاها

لذلك ْ هاتكون بالشكل ده بعد اما افترضنا انى X0 هاتشاوى 1

هاتكون بالشكل ده

Θ ال X وال y

بما ان Θ = 2\*1 وال X= 8\*2 يبقا XΘ ينفع يضربوا لان 8\*2 and 2\*1 والناتج هايكون 8\*1 matrix والى هى تنفع تطرح من y بعدها بيحصلها transpose فتبقا 1\*8 والجزء التانى هايكون 8\*1 فساعتها ققدر اعمل multiplication بين جزئين المعادله والناتج الى هو هايكون رقم واحد يضرب فى 1/2m وده يبقا ال cost function .

من المفترض انك تكون عارف matrix opertions بتتعمل ازاى.

J(*θ)=*

= = - = = transpose it =

Then the same for

هايدينا نفس الناتج بس من غير transpose فهيكون الناتج

so =

وده بيساعدنى جدا فى انى معملش اى loops مجرد بس matrixes operations .

**ثانيا ال Gradient Descent**

**نفس الكلام افترض theta 0 = 1 and theta 1 = 3 وقيمة الفا ب 0.002**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | H(x) | y | X1 |
| 1\*100=100 | 301-300=  1 | h(x) = 1 + 3 X  1+3\*100=301 | 300 | 100 |
| 95 | 1 | 286 | 285 | 95 |
| 90 | 1 | 271 | 270 | 90 |
| 80 | 1 | 241 | 240 | 80 |
| 480 | 6 | 241 | 235 | 80 |
| 770 | 11 | 211 | 200 | 70 |
| 420 | 6 | 211 | 205 | 70 |
| 60 | 1 | 181 | 180 | 60 |
| = 100+95+90+80+480+770+420+60=2095  Then  0.002\*  =.52  Then  هنا المفروض انى theta 0 بتطرح قبل ما اضرب فى Xi الاخيره يعنى المعادلة بتكون كالتالى  فالمجموع بدل ما كان 2095 بعد الضرب هايبقا قبل الضرب 28 لما يضرب فى .002/8 الناتج هايكون 0.007 لما نطرح theta 0 الى هى 1 هايكون الناتج  Theta 0 = 1 - 0.007 = 0.993  لكن  Theta 1 = 3 - ((0.002 / 8) \* (2095) = 2.48 |

وبعد كده بقا عن طريق قيم ال thetas الجديده ابتدى انى اعمل iterations لحد اما اوصل لل global minimum .

**ال vectorize version of gradient descent**

Assume = 1 for 8 training examples

حول كل حاجه من المعادله بتاعت ال vectorize ل matrix or vector عشان نقدر نتعامل معاها

لذلك ْ هاتكون بالشكل ده بعد اما افترضنا انى X0 هاتشاوى 1 لما اجى اضربها فى theta فقط لكن فى العادى هى 8\*1

هاتكون بالشكل ده

Θ ال X وال y

بما ان Θ = 2\*1 وال X= 8\*2 يبقا XΘ ينفع يضربوا لان 8\*2 and 2\*1 والناتج هايكون 8\*1 matrix والى هى تنفع تطرح من y والناتج هايكون vector of 8\*1 والجزء الاول هايكون 1\*8 بعد ما عمل transpose ل فساعتها ققدر اعمل multiplication بين جزئين المعادله والناتج الى هو هايكون رقم واحد يضرب فى وبعدها نطرح من قيم ال thetas الى اكبر من 0 لان theta 0

=

-

2095 =

*بعد كده بقا بقدر انى اجرب عدد من الخطوات لحد اما اوصل للى global minimum.*

*الى فات ده كده كله كان طبيق لما يكون عدد ال features واحد فقط.*

*تطبيق عملى على Multivariable Regression .*

*لما يكون عندى اكتر من feature فهنا عندى 3 features بيمثلوا مميزات السياره وال y وهى سعر السياره بالنسبه لل 3 feature الخاصين بيها فهنا مثلا اول مثال هو*

*X1 = 5 and X2 = 20 and X3 = 6*

*بمعنى انى اول سياره الى هى X1 على حسب ال features الى مذكوره فوق سعرها بيكون هو y=12 .*

*فهنا عندى 5 training example and 3 features لذلك m=5 and n=3 .*

|  |  |  |  |
| --- | --- | --- | --- |
| *Y* | *X3* | *X2* | *X1* |
| *السعر* | *الاسطوانات* | *القدرة* | *العمر* |
| *114* | *6* | *20* | *5* |
| *120* | *6* | *35* | *5* |
| *123* | *8* | *38* | *6* |
| *121* | *8* | *40* | *7* |
| *135* | *10* | *46* | *7* |

*بعد كده هافترض قيم thetas مقابل كل feature ولكن هايكون مقابل theta 0 هايكون X0 =1 لكل مثال.*

*فلنفترض قيم thetas*

*اولا ال Cost function for Multivariable Regression هاتكون*

*J(θ) =*

*كل trainging example بيمثل عندى vector of 4\*1 الى هو بعد افتراض انى X0 = 1 + ال 3 feature هايكون 4 بالنسبه لمثال واحد هايكون 4 \* 1 لكن لو بتكلم عن X كلها هنا بتمثل 4\*5 الى هو 4 features فى عدد ال training data الى هما 5 وبما انى theta = 4\*1 فلما اعملها transpose هاتبقا 1\*4 وهنا ققدر اضربها فى ال X الى ههى 4\*5 والناتج هايكون vector 5\*1 هايتمطرحه من y الى هى 5\*1 وبعدها نربع كل رقم من ال vector ونجمع ثم نضرب فى .1/2m*

*J(θ) =*

*اولا*

*هانعمل transpose عشان نقدر نطرح الارقام ديه من y فهيكون الى هى المعالة ديه*

*+ = 15521 \* = 15521 \**

**vectorize version solution of cost function with multivariable regression**

**بدام كده كده بضرب فى الاخر فى matrix of X فى vector of theta فسواء عملت transpose لاى فيهم وضربته فى التانى الناتجين هايكونوا زى بعض الاختلاف فقط هايكون فى deamination**

J(*θ)=*

**ثانيا ال Gradient Descent هايتم التطبيق فى المعادلات الخاصه بيه سواء كان vectorize or not**

**Vectorize version**

**واخلى بالى انى فى حاله** 0  هى الواحيده الى بجبها لوحدها بالمعادله نفسها عدا الضرب فى Xi .

**بإذن الله الجاى هايكون نفس التطبيق على Logistic regression واتمنى من الله اكون قدرت اساعد بحاجه بسيطة وإن شاء الله ققدر اكمل باقى الكورس على خير.**