كلية الهندسة المعلوماتية

تعليم الكتروني

المحاضرة : الثانية

عدد الصفحات: 22

السنة الخامسة

GATE

Greatness Achieved Throught Excellence

في المحاضرة السابقة تكلمنا عن عدة تعاريف لتعلم الالة وأنواع خوارزميات التعلم, وتحدثنا بشكل مفصل عن التعلم بإشراف وانواعه. تكلمنا بشكل موسع عن الـ linear regression وذكرنا انه عبارة عن مجموعة متغيرات مستقلة عن بعضها البعض independent variables/Features/attributes (مثل مساحة الغرفة, عدد الغرف, ...) واريد إيجاد قيمة متغير معتمد على المتغيرات السابقة Dependent variable (سعر المنزل).

تكلمنا أيضا عن عملية التدريب اعتمادا على الداتاست لكي ينتج مودل ومن خلال هذا المودل نقوم بالتنبؤ بقيمة المتغير (سعر المنزل) الذي يعتمد على المتغيرات المستقلة (عدد الغرف, مساحة المنزل ...) والمودل يكون عبارة عن معادلة من درجة معينة ولكل معادلة منحني بياني خاص بها (تحدثنا عن الشكل البسيط وهو معادلة من الدرجة الأولى ويكون شكلها عبارة عن خط مستقيم, وقلنا ان هذه الحالة تمثل linear regression).

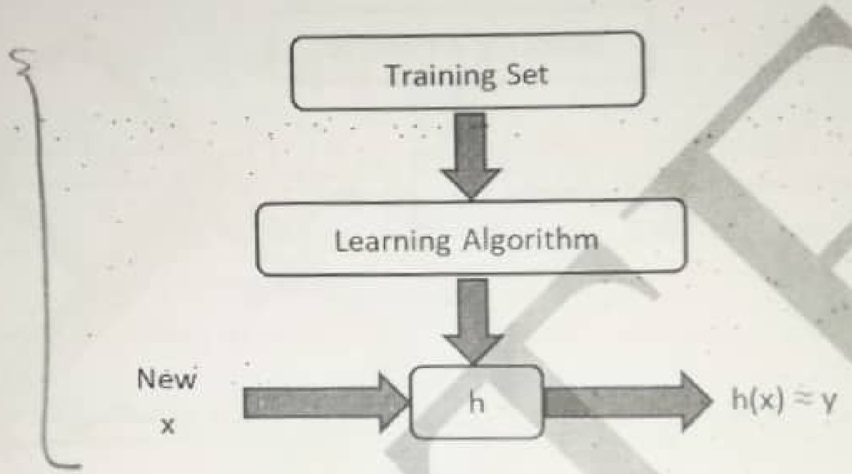
***

إذاً ماهو الـ supervised Learning ؟

خوارزمية نقوم بإعطاءها بيانات (تسمى هذه البيانات بالداتاست Data Set أو Training set) و تحوي هذه الداتاست الدخل والخرج. أي نعطي الخوارزمية الدخل والخرج الصحيح, ومهمة الخوارزمية أن **تتعلم** كيف تحسب الخرج (تتنبأ به) أي ان توجد علاقة ربط بين x و y من خلال تابع رياضي.

وهذا التابع نستخدمه فيما بعد للتنبؤ بقيم جديدة وكلما كانت علاقة الربط قوية أي يتم انتاج قيم متنبأ بها قريبة من القيم الحقيقية كانت الخوارزمية افضل, ويسمى هذا التابع بالمودل او hypothesis واختصارا h .

في مثال تسعير المنازل نعطي خوارزمية التعلم المساحة والسعر ومهمتها معرفة كيف تسعّر المنازل عندما نقوم بإعطاءها منازل جديدة لتتنبأ بأسعارها.

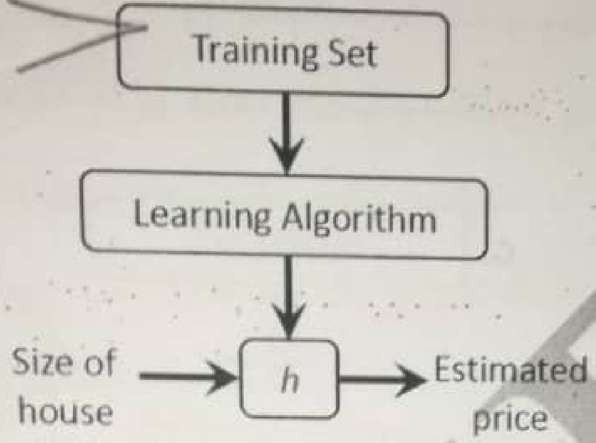


New x يمثل مساحة المنزل الجديد التي تم إدخالها, أي هو أي دخل x جديد (يسمى في المراجع Unsean example) لم تتدرب عليه الخوارزمية, اما h يمثل المودل وهو خلاصة تعلم الخوارزمية, اما h(x) يمثل التنبؤ من قبل الخوارزمية (أي السعر الذي تنبأت به الخوارزمية من اجل مساحة المنزل المدخلة, ويجب ان يكون اقرب مايمكن الى القيمة الحقيقية y).

من الرسم السابق نلاحظ انه يتم **ادخال** الداتاست او مجموعة التدريب او داتاست التدريب او Training set (نفس المسمى) الى خوارزمية التعلم لكي تقوم بالعمل الداخلي (إيجاد العلاقة بين X و y) وإعطاء المودل في النهاية, وبعد الحصول على المودل h , نستطيع التنبؤ, حيث نقوم بإدخال دخل جديد او مثال جديد الى المودل لكي يقوم بإيجاد القيمة المتنبأ بها.

تعديل الرسم حسب المثال السابق (الشكل مهم وقد يأتي سؤال في الامتحان) :

البيانات
الحل الصحيح



***

من خلال المثال الخاص بالتنبؤ بسعر المنزل, نلاحظ ان ابسط علاقة تربط x بـ y هي معادلة مستقيم :

$$y = ax + b$$

والمستقيم السابق تابع لـ a,b أي هذه القيم هي التي تحدد شكل المستقيم.

ذكرنا سابقاً ان المودل h يمثل المستقيم المرسوم ونستطيع إعادة كتابة المعادلة السابقة بالشكل:

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

حيث $\theta_0 = b, \theta_1 = a$

وهذا المستقيم تابع للبارامترات θ_0, θ_1 أي هذه القيم هي التي تحدد شكل المستقيم (سميت بارمترات parameters وليست متحولات او متغيرات لأننا نحتاج ان نعدل قيمهم حتى نصل للقيم المناسبة) وعند تغير هذه القيم يتغير شكل المستقيم.

للاختصار نستطيع كتابة المعادلة السابقة :

$$\boxed{h\,(x) = \theta_1 x + \theta_0}$$

ملاحظة:

لمعرفة القيمة المتنبأ بها نقوم باسقاط النقطة على المستقيم وننظر الى مسقطها على محور y ليظهر لنا السعر المتنبأ به.

$y = ax + b$

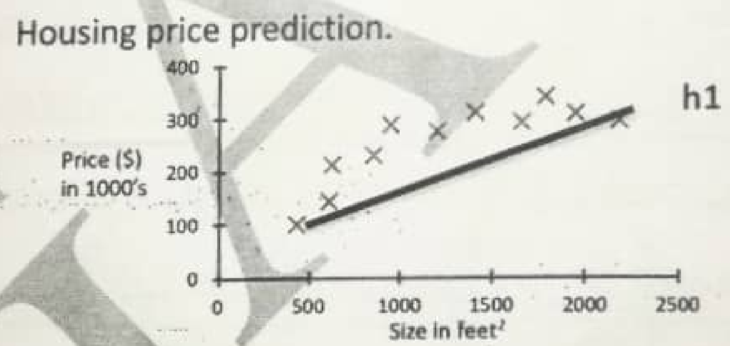
من المحاضرة السابقة نعرف ان العلاقة التي تربط بين x و y ليس بالضرورة ان تكون مستقيم بل قد تكون منحني بسيط او معقد.

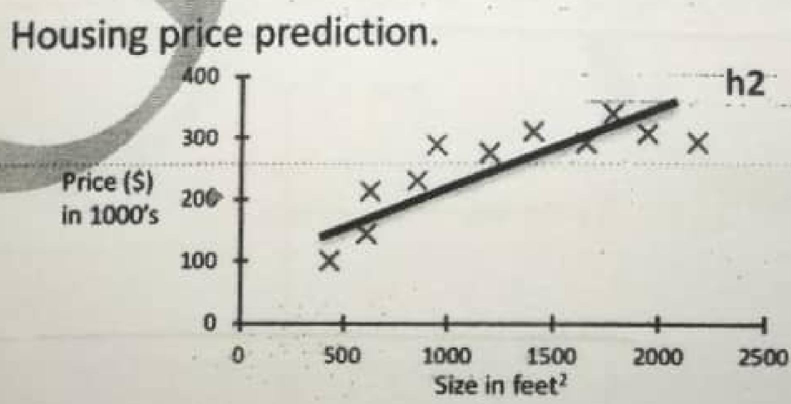
$h_\theta(x) = \theta_1 x + \theta_0$

المسائل التي سنتعامل معها حاليا تحوي في الداتاست اما عمود واحد او عدة أعمدة حسب عدد الخصائص وعمود y (عمود واحد فقط مهما كانت عدد الخصائص).

***

تذكر من المحاضرة السابقة اننا قمنا برسم شكلين مختلفين كما يلي:





نلاحظ أن خوارزمية التعلم الثانية تنبأت بسعر المنزل بشكل أفضل أو أكثر دقة من الخوارزمية الأولى.

b a

وبما ان المستقيم يمثل المودل h, ومعادلة المودل تابعة لـ θ_0, θ_1 , أي كلما تغيرت قيم البارمترات او التيتات, تغير شكل المستقيم, لذلك المستقيمين السابقين لهما قيم مختلفة للبارمترات.

مثلا اذا كانت لدينا قيم مختلفة لـ θ_0, θ_1 كيف سنعرف ماهو المودل الأفضل بين h1 و h2 ؟ أو كيف سنعرف ماهو المودل (المستقيم) الذي يمر بالنقاط بأفضل شكل ممكن كما ذكرنا سابقاً ؟ او كيف سنعرف ماهي افضل القيم لـ θ_0, θ_1 ؟

نحتاج لطريقة نستطيع بها تقييم مدى جودة كل من التابعين لمعرفة التابع أو المودل الأفضل.

ويتم ذلك من خلال حساب الكلفة, والمودل الذي يملك كفلة او خطأ اقل يكون اكثر دقة, آلية حساب الكلفة تعتمد بشكل أساسي على حساب الفرق بين القيمة الحقيقية والقيمة المتنبأ بها وبما انه يتم ادخال عدد كبير من امثلة التدريب (الداتاست) الى الخوارزمية, سنقوم بحساب الفرق بين القيم المتنبأ بها والقيم الحقيقية من اجل كل مثال تدريب, وهذا الفرق نستطيع تسميته بالخطأ أو الكلفة. وهذا هو جوهر عمل تابع الكلفة كما سنرى.

***

Cost function

فهم

نقوم بتحقيق الآلية السابقة من خلال تابع خاص يقوم بقياس أداء المودل عن طريق حساب الكلفة أو الخطأ, ويجب أن تكون هذه الكلفة أقل ما يمكن min. وبما أن المودل h تابع لـ θ_0, θ_1 إذاً الهدف هو إيجاد افضل قيم للبارمترات θ_0, θ_1 التي تجعل الكلفة أقل ما يمكن, وبكلام آخر اختيار قيم بارمترات التابع h بحيث تكون القيمة المتنبأ بها قريبة قدر الإمكان على القيمة الحقيقية من أجل كل أمثلة التدريب, او ان يملك تابع الكلفة اصغر قيمة ممكنة.

كيف سنحصل على هذا الهدف؟

تذكر أولاً أن القيم المختلفة للبارمترات ستعطى توابع او مستقيمات مختلفة.

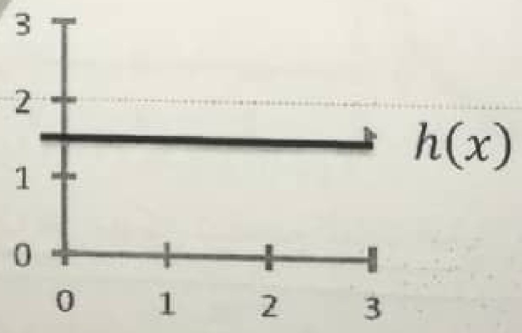
مثال :

1- عندما تكون $\theta_1 = 0$, $\theta_0 = 1.5$, نعوض بالتابع كما يلي:

$$h(x) = 0 * x + 1.5$$

$$h(x) = 1.5$$

سيكون الشكل كما يلي:



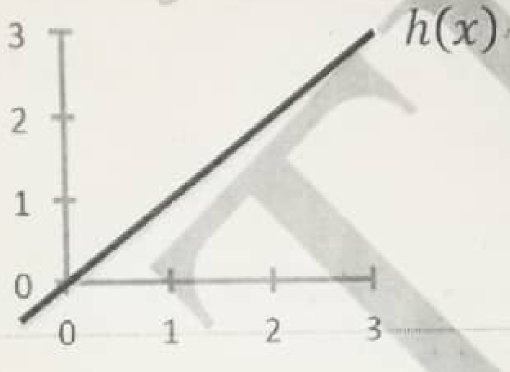
2- عندما تكون $\theta_1 = 0.5$, $\theta_0 = 0$, سيكون التابع كما يلي:

$$h(x) = 0.5 * x + 0$$

$$h(x) = 0.5 * x$$

الميل صفر
المستقيم منصف الربع الأول والثالث

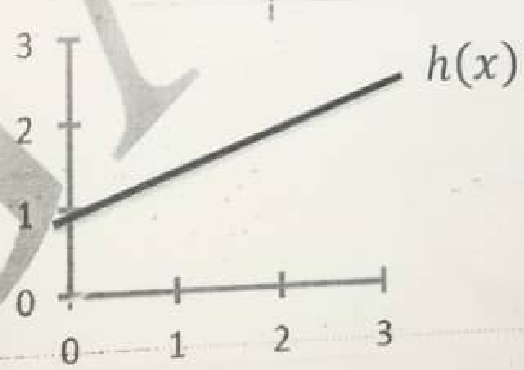
سيكون الشكل كما يلي :



3- عندما تكون $\theta_1 = 0.5$, $\theta_0 = 1$, سيكون التابع كما يلي:

$$h(x) = 0.5 * x + 1$$

سيكون الشكل كما يلي :



لتحقيق هدف ان تكون الكلفة اقل مايمكن نستخدم المعادلة التالية:

$$\boxed{Minimize\ \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right)-y^{(i)}\right)^2}$$

حفظ بصم

نقوم الان فقط بإضافة رمز تابع الكلفة cost function لنحصل على الشكل النهائي للمعادلة, وهو يمثل مقياس metric يعطي مدى جودة المودل وهو عبارة عن Mean square error او MSE ويرمز له بـ J ومعادلته على الشكل التالي:

دقة المودل (دقة الخوارزمية)

cost

$$\boxed{J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2}$$

أي وظيفة هذا التابع هو إيجاد مجموع مربعات الفرق بين القيمة المتنبأ بها والتي تمثل $h(x^{(i)})$ والقيمة الحقيقية $y^{(i)}$ وذلك لكل أمثلة التدريب أي من i=1 أي من المثال الأول الى m والتي هي حجم الداتاست او عدد امثلة التدريب.

وبضرب الناتج بـ $\frac{1}{2m}$: لحساب المتوسط $(\frac{1}{m})$, وجعل العمليات الحسابية أسهل قليلاً $(\frac{1}{2m})$.

ويجب أن يكون هذا التابع أصغر ما يمكن لكي يتصف المستقيم او المودل النقاط بأفضل شكل ممكن.

معادلة المودل (القيمة المتنبأ بها) من اجل مثال التدريب الأول i=1:

$$h(x^{(1)}) = \theta_0 + \theta_1 x^{(1)}$$

معادلة المودل (القيمة المتنبأ بها) من اجل مثال التدريب الثاني i=2:

$$h(x^{(2)}) = \theta_0 + \theta_1 x^{(2)}$$

معادلة المودل (القيمة المتنبأ بها) من اجل كل امثلة التدريب او كامل الداتاست:

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

$$\hat{Y} = b + ax$$

في الحالة المثالية يكون J=0 , أي القيم المتنبأ بها مطابقة للقيم الحقيقية, وتكون في حالة أن النقاط على استقامة واحدة والمستقيم ينصف كل النقاط.

إذاً أصبح هدف خوارزمية التعلم إيجاد قيم البارمترات θ_0, θ_1 الخاصة بـ h التي تجعل القيمة المتنبأ بها اقرب ما يمكن الى القيمة الحقيقية من أجل أمثلة التدريب(x,y) او نفس الكلام ان يكون J أصغر ما يمكن.

الهدف النهائي:

$$\underset{\theta_0, \theta_1}{Minimize}\ J(\theta_0, \theta_1)$$

وتسمى هذه المشكلة minimization cost أي ان عملية الـ minimization هي حساب اصغر قيمة للكلفة بناء على علاقة معطاة وقيم مختلفة للبارمترات, بحيث نختار اصغر قيم للكلفة ونحتفظ بقيم البارمترات التي جعلت الكلفة اقل مايمكن. وبالتالي خرج عملية الـ minimization هو قيم البارمترات المثالية او الأفضل.

***

شرح تابع الكلفة بكلام بسيط

لنفرض في البداية انه يوجد مثال تدريب واحد فقط لتوضيح الفكرة, فرضا مساحة المنزل 200 متر مربع وكان السعر الحقيقي للمنزل 4000 ليرة, في حال قمنا بإدخال هذا المثال التدريبي الى مودل مدرب سابقا (لايجاد القيمة التي سيتنبأ بها من اجل مساحة منزل 200 متر مربع) وهذا المودل تنبأ بسعر المنزل بـ 3800 ليرة.

تذكر في الحالة المثالية تكون القيمة المتوقعة مطابقة للقيمة الحقيقية أي اذا توقع المودل سعر المنزل 4000 ليرة, هنا يكون السعر الحقيقي مطابق للسعر المتنبأ به وبالتالي نقول ان المودل دقيق بنسبة 100 بالمية, اما في حال توقع المودل سعر المنزل 3800 ليرة نقول ان معدل دقة المودل مثلا 80 بالمية, لان القيمة الحقيقية مختلفة عن القيمة المتنبأ بها.

اما في حال توقع المودل سعر المنزل 600 ليرة نقول ان معدل دقة المودل مثلا 13 بالمية وستكون خوارزمية التعلم سيئة, وهكذا.

الان لكي نعرف دقة المودل نحتاج لحساب الكلفة او الخطأ الذي ارتكبه ولإيجاد الخطأ نقوم بطرح القيمة الحقيقية التي هي 4000 ليرة من القيمة المتنبأ بها والتي هي 3800 ليرة لينتج الخطأ والتي قيمته تساوي 200 ليرة.

مثلا في حال كان هناك مودل تنبأ بسعر المنزل 3800 ومودل اخر تنبأ بسعر المنزل 2800, لايجاد المودل الأكثر دقة نبحث عن المودل الذي يملك اصغر قيمة للخطأ وفي حالتنا يكون هو المودل الأول لان قيمة الخطأ لديه 200 ليرة اما المودل الثاني قيمة الخطأ لديه 1000 ليرة, اذا المودل الأول يستطيع التنبؤ بسعر المنزل بدقة أكثر من المودل الثاني (أي تكون القيمة المتنبأ بها قريبة على القيمة الحقيقية).

الان الكلام السابق منطبق على مثال تدريب واحد.

اما في حال وجود 100 مثال تدريب, لايجاد الكلفة الخاصة بالمودل اطبق نفس الكلام السابق تماما, أقوم بطرح القيمة الحقيقية من القيمة المتنبأ بها من اجل المثال الأول ثم المثال الثاني ثم الثالث ... وفي النهاية أقوم بجمع القيم لكي نحصل على الكلفة الاجمالية الخاصة

بالمودل. (المثال السابق تم تبسيط فكرة حساب الكلفة لسهولة الفهم اما في المعادلة نقوم بطرح القيمة المتنبأ بها من القيمة الحقيقية ونقوم بالتربيع وغيره من العمليات).

شرح تابع الكلفة

في البداية من أجل التبسيط لنفرض أن $\theta_0 = 0$ (هذا البارامتر يمثل الازاحة على محور x وهو يمثل b أيضا في حال تم استخدام الرموز w,b) وبالتالي ستكون معادلة المودل :

$$h(x) = \theta_1 x$$

وبالتالي أصبح تابع الكلفة تابع لـ θ_1 فقط أي :

$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

أي بدلاً من معرفة أفضل قيمة للبارمترين سنكتفي بمعرفة أفضل قيمة لـ θ_1 فقط لكي نقوم بتصغير تابع الكلفة لأقل ما يمكن.

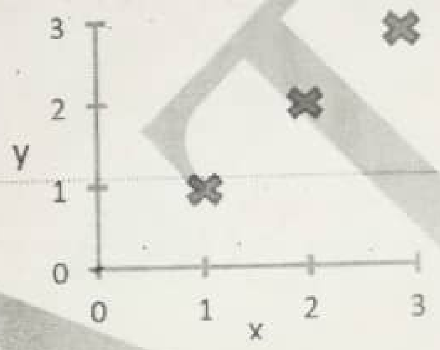
لدينا الآن تابعين:

- تابع المستقيم او المودل h(x)
  - وهو تابع لـ x (مساحة المنزل في مثال التنبؤ بسعر المنزل)
- تابع الكلفة $J(\theta_1)$
  - وهو تابع للبارمتر θ_1 . نفرض $\theta_0 = 0$

والهدف هو

$$\underset{\theta_1}{Minimize}\, J(\theta_1)$$

فهم عوا مثال : ليكن لدينا الداتاست المرسومة بالشكل التالي حيث تحوي 3 نقاط (1,1),(2,2),(3,3) على الترتيب:

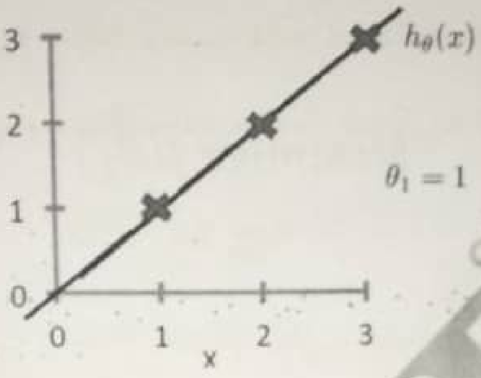


لنقوم الآن بتعويض بعض القيم , مثلاً عندما تكون $\theta_1 = 1$ سيكون

$$h(x) = \theta_1 x = 1 * x = x$$

وبالتالي لرسم هذا **التابع** كما يلي (h(x)) :

المستقيم منطبق على النقاط لانه من اجل x=1 كانت h(x) تساوي 1 ومن اجل x=2 كانت h(x) تساوي 2 ... (تعويض مباشر في القانون)



$h(x) = x$
معادلة مستقيم ينصف الربعين الأول والثالث

- نلاحظ أن المستقيم أو المودل منطبق على النقاط وبالتالي الدقة هنا 100% أي القيمة المتوقعة مطابقة للقيمة الحقيقية. يعاني من over fitting

نقوم الآن بالتعويض في تابع الكلفة كما يلي :

$$J(1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

نعوض $h\left(x^{(i)}\right) = \theta_1 x^{(i)}$ (تذكر اننا قمنا بحذف ثيتا 0) وبالتالي :

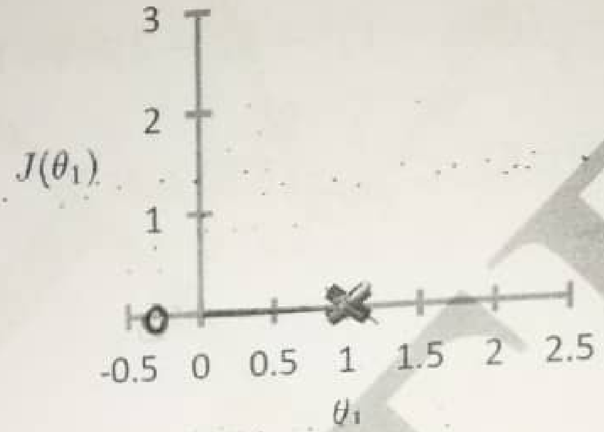
$$J(1) = \frac{1}{2m}\sum_{i=1}^{m}\left(\theta_1 x^{(i)} - y^{(i)}\right)^2$$

$$J(1) = \frac{1}{2m}\left((1-1)^2 + (2-2)^2 + (3-3)^2\right)$$

$$J(1) = \frac{1}{2m}\left(0^2 + 0^2 + 0^2\right) = 0$$

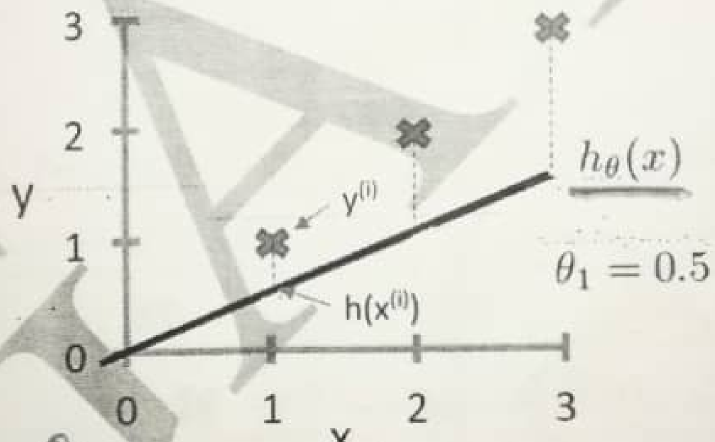
$$J(1) = 0$$

لنقوم برسم النقطة على الشكل (الذي يمثل **التابع** $J(\theta_1)$) :



نلاحظ ان قيمة الكلفة هنا تمثل 0 عندما تكون $\theta_1 = 1$ وهي الحالة المثالية..

مثال لتعويض نقطة ثانية على سبيل المثال : $\theta_1 = 0.5$. نعوض في معادلة التابع $h(x)$:



$h(x) = \theta_1 x + \theta_0 \rightarrow = 0$

$h(x) = 0.5\,x$

لاحظ أن المستقيم غير منطبق على النقاط في الداتاست.

(1,1) نصف الواحد $= 0.5$

(2,2) نصف الاثنان $= 1$

(3,3) نصف الثلاثة $= 1.5$

لنقوم الآن بالتعويض في تابع الكلفة:

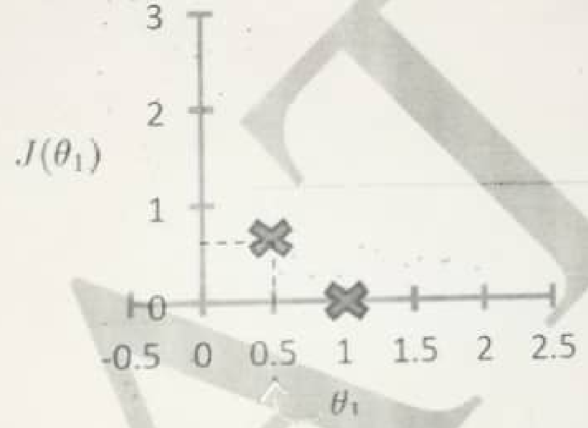
$$J(0.5) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$J(0.5) = \frac{1}{2m}((0.5-1)^2 + (1-2)^2 + (1.5-3)^2)$$

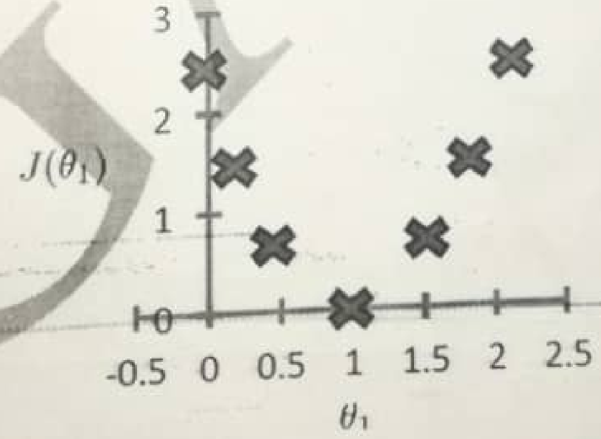
$$J(0.5) = \frac{1}{2*3}((0.5-1)^2 + (1-2)^2 + (1.5-3)^2)$$

$$J(0.5) = \frac{1}{2*3}(3.5) \approx 0.58$$

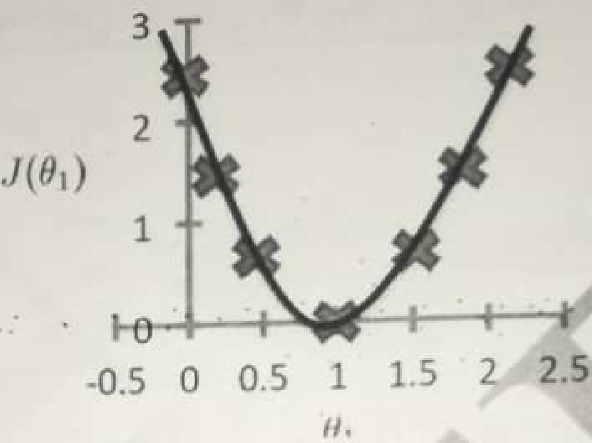
للتعويض في الشكل :



وهكذا بعد تعويض العديد من النقاط بنفس الطريقة قيم تابع الكلفة التالية:



ينتج ما يسمى الـ polynomial أو متعدد الحدود او قطع مكافئ (تابع تربيعي) كما يلي :



ملاحظة:

لاحظ من الشكل أن تابع الكلفة $J(\theta_1)$ يملك أصغر قيمة أي $J(\theta_1) = 0$ عندما يكون $\theta_1 = 1$, وعندما $\theta_1 = 1$ تذكر أن شكل المودل يكون :

أي المودل يمر بالنقاط أو ينصف نقاط الداتاست تماماً . أي دقة الخوارزمية 100% . وعندها يكون تابع الكلفة يساوي الصفر وهي الحالة المثالية global minimum, وبالتالي $\theta_1 = 1$ هي أفضل قيمة او القيمة المثالية.

اذاً كل قيمة لـ θ_1 تكافئ مودل مختلف أو خط مستقيم مختلف , و تكافئ قيمة جديدة لتابع الكلفة $J(\theta_1)$. و هدف خوارزمية التعلم هي اختيار قيمة θ_1 التي تصغر قيمة الكلفة لأقل مايمكن (حيث نستطيع القول ان تابع الكلفة مسؤول عن مدى جودة البارمترات).

***

$y = ax + b$

$h(x) = \theta_1 x + \theta_0$

$y = wx + b$

لنعود الآن للحالة العامة وهي البارمترين θ_0, θ_1 . او w,b

W تمثل θ_1 اما b تمثل θ_0

فرضية

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

Parameters: θ_0, θ_1

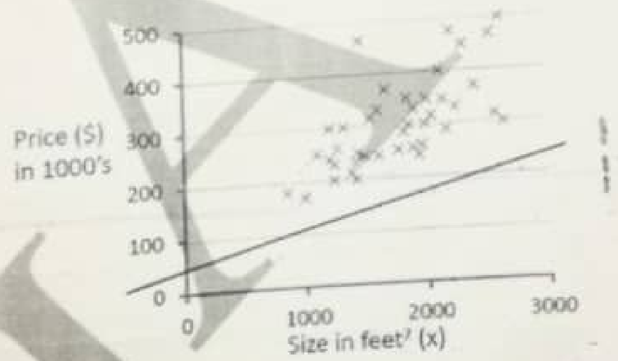
Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$ mean square error — متوسط مربع الخطأ

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

كما في السابق تابع المودل بالشكل التالي :

$h_\theta(x)$

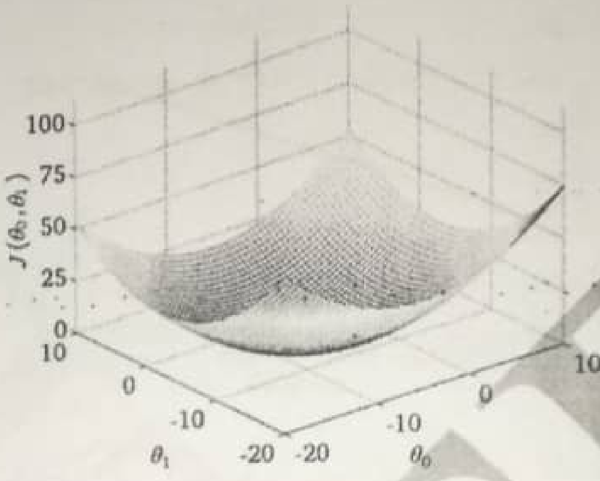
(for fixed θ_0, θ_1, this is a function of x)



$h_\theta(x) = 50 + 0.06x$

عند تعويض بعض القيم للبارمترين و رسم تابع الكلفة $J(\theta_0, \theta_1)$ سينتج الشكل التالي (للتشبيه مثل الصحن المجوف) :

قيم بارامترين ← رسم تابع الكلفة ← زبدية



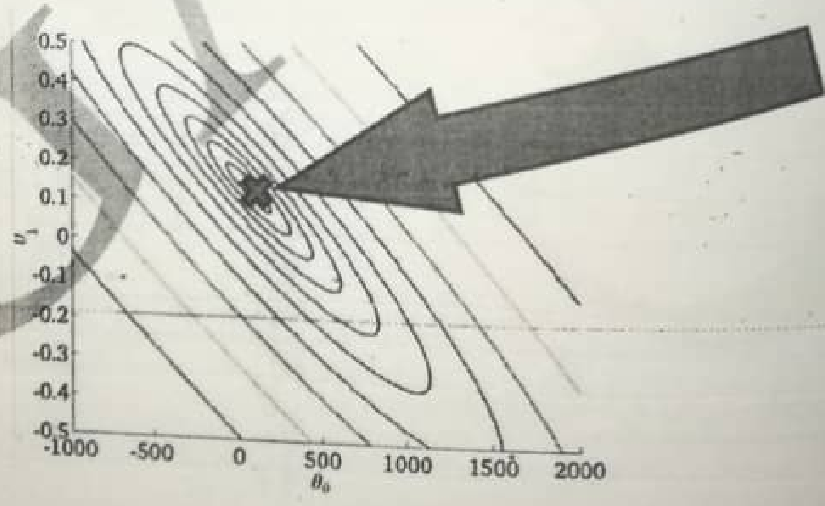
هو شكل ثلاثي الأبعاد و يوجد 3 محاور :

$$X = \theta_1 , Z = \theta_0 , Y = J(\theta_0, \theta_1)$$

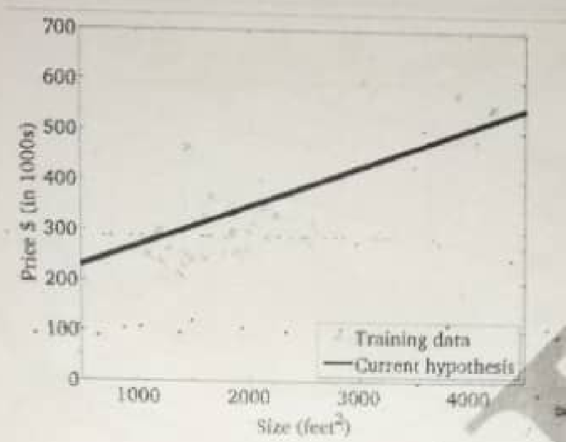
من الشكل السابق يملك التابع أصغر قيمة عندما يكون في الأسفل (أسفل نقطة في باطن الشكل).

لنأخذ مقطع بسيط منه للتوضيح

أفضل قيمة ستكون مكان النقطة التالية global minimum:



وكما في السابق عندما تكون قيمة تابع الكلفة مكان النقطة سيكون المستقيم ينصف النقاط بالحالة المثالية .

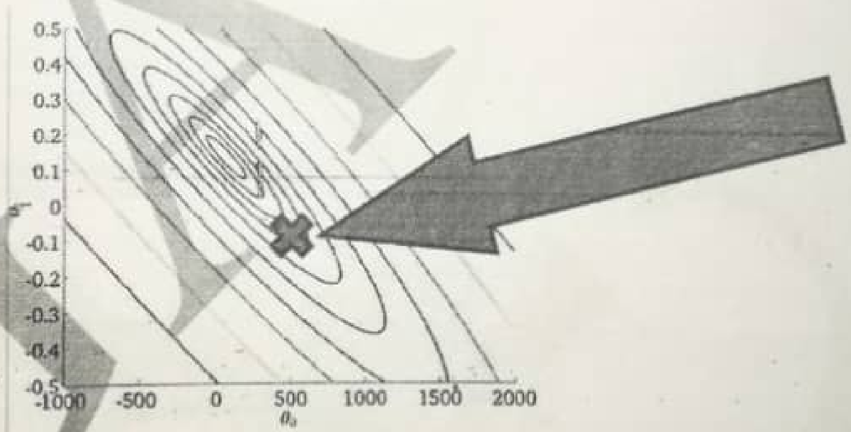


وكلما ابتعدت هذه النقطة عن المنتصف سيختلف شكل المستقيم ولن ينصف نقاط الداتاست بالشكل المطلوب.

مثلا إذا كانت قيمة تابع الكلفة:

$J(\theta_0, \theta_1)$

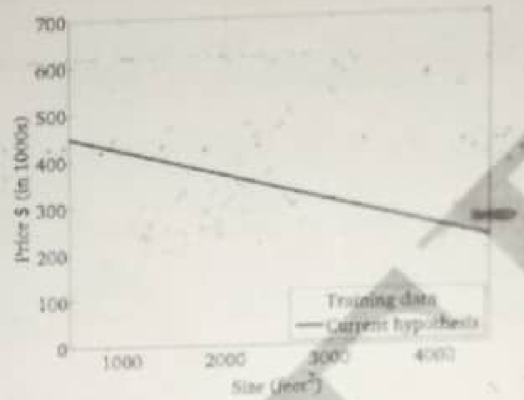
(function of the parameters θ_0, θ_1)



سيكون المستقيم كما يلي :

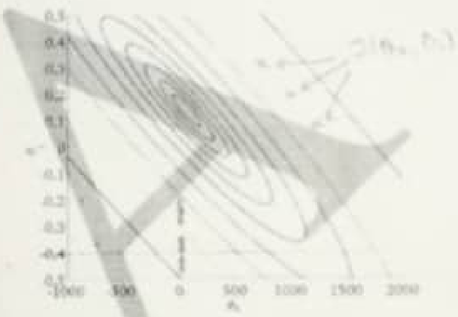
$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



قيمة تابع الخطأ ستكون كبيرة، وبالتالي لن يمر المستقيم بالنقاط بأفضل شكل ممكن.

بعض القيم الأخرى للكلفة :



السؤال الأهم هنا كيف سنصل إلى هذه القيم أو كيف سنعرف قيم البارمترات التي تجعل المستقيم يمر بالنقاط بأفضل شكل أو تجعل تابع الكلفة أصغر ما يمكن ؟

من خلال خوارزمية خاصة عبارة عن learning algorithm تسمى Gradient descent سنراها في المحاضرة القادمة.

***

ملاحظة هامة:

رأينا في الفكرة السابقة كيف قمنا بحذف بارمتر (تيتا 0) لتوضيح الفكرة, وكان الشكل عبارة عن ثنائي البعد, اما بعد إضافة التيتا 0 اصبح لدينا شكل ثلاثي الابعاد, وعند إضافة خصائص جديدة سيزداد البعد وسيزداد الشكل تعقيدا وقد يكون من المستحيل تصوره او رسمه, مع العلم انه ليس كل الخصائص تساهم في تقديم معلومات مفيدة بنفس الدرجة, مثلا لنأخذ مثال التعرف على الوجه, وكان لدينا الخصائص لون العيون والعمر ولون الشعر. وبعد التجربة استطعنا معرفة ان خاصية العمر تساهم في تقديم معلومات مفيدة بشكل اكبر من خاصية لون العيون لذلك نستطيع حذف خاصية لون العيون لكي يتم تبسيط المودل دون خسارة معلومات بشكل كبير وهذا المبدأ يسمى Dimensionality reduction أي تقليص الابعاد من خلال حذف الخصائص التي تعقد المودل ولا تحمل معلومات مفيدة لذلك نتخلص منها بدون التضحية في فقدان كمية كبيرة من المعلومات.

***

supervised Learning Al—

Logistic ~~regression~~ – classification

binary class
multi_class

في المحاضرة السابقة تحدثنا عن الـ supervised learning وبشكل خاص عن الـ linear regression وذكرنا أننا نتنبأ بالخرج y في مجال مستمر مثل درجة الحرارة, كمية المبيعات, المعدل ... وأن المودل يجب ان يكون عبارة عن خط مستقيم أو منحني يمر بالبيانات أو النقاط بأفضل شكل ممكن.

أما مفهوم الـ classification ويسمى أيضاً logistic regression, ويكون مجال الخرج y الذي سنتنبأ به متقطع و ليس مستمر مثل مسائل تشخيص الأمراض مثلاً هل

الشخص مصاب أم لا و تصنيف الايميل هل هو سبام أم لا , هل الطقس غائم أو ماطر أو مشمس , وهكذا ... أي مجال الخرج متقطع.

Binary classification

هو نوع من الـ classification (binary class problem) وهو أبسط الأنواع, بحيث يكون مجال الخرج y هو قيمتين فقط (0 أو 1) , أي يوجد كلاسين فقط 0,1 : $y \in \{0,1\}$, حيث عندما يكون y=1 أي وجود الشيء (الشخص مصاب, الايميل سبام ...) أما عندما تكون قيمة y=0 أي غياب الشيء (الشخص ليس مصاب , الايميل ليس سبام ...).

اما اذا كان هناك أكثر من كلاسين او صنفين مثلا التنبؤ بـ 4 أصناف مثلا التنبؤ بالصورة هل تحوي شاحنة او سيارة او دراجة او غير ذلك. وفي هذه الحالة نسمي المسألة multi-class classification وليس binary classification قد نقوم بدراستها لاحقا.

في مسائل التصنيف نعامل القيمة المتنبأ بها على انها قيمة احتمالية (احتمال ان يكون المثال المدخل الى المودل تابع للصف كذا) وسنقوم بشرحها بالتفصيل في المحاضرات القادمة.

نهايــة المحاضرة

Ali Mannoun