

مواضيع هامة

Handling Error

التعامل مع الأخطاء وهو ازاي قدر اتعامل مع الحاجات الى بتعابلى او الحاجات اصلا الى بتسبب اخطاء وان بشغل زى انحراف ال **data**.

مهم جدا انك تكون فاهم الشغل ماشى ازاي او ايه الى المفروض تستخدمه فى الجزئية دى وليه استخدمته ولو حصل حاجه معينه زى ال **over fitting** تبqa فاهم ده حصل ليه يعني مش مجرد شويه **algorithms** واکواد و **cost** تبقى كده تمام لا المفروض تبqa فاهم الموضوع بشكل أعمق ليه بتسخدم وده وليه ال **error** ده وليه قيمة ال **function** كبيرة مثلا فى ال **test data** طب هل زدت **features** الوضع هيكون احسن وليه اخترت اصلا انى ازود **features** طب لو خليت المعادلة تكون **polynomial** وليه برضه خلتها **polynomial** وهكذا.

فلازم إنتاكمهندس **machine learning** يكون عندك فن وتقدير ووضع الحاجات فى مكانها المناسب. ممكن يكون فيه اكتر من تكنيك و **algorithm** لكن اختيارك لواحد معين يعتمد على العديد من العوامل زى الوقت والميموري وال **complexity** وغيرها.

وفيه طريقة اسمها **machine learning diagnostic** وهى طريقة تتم داخل ال **algorithm** و هو شغال وبتساعدنى اعرف ايها المناسب وغير المناسب والى شغال فعلا ومش شغال وازاي ممكن قدر اطور ال **algorithm** بتابعى.

Evaluating Hypothesis

هى إنى اقيم ال **algorithm** بتابعى ومدى دقته يعني هل تمام اكمل شغل بيه ولا هو بيسبب انى يكون عندي over fitting ومن خلال ده بقدر احدد انى يحتاج اعمل ايه او اطلب ايه بطلب.

التعامل مع الأخطاء

تقييم الإفتراض

- طبيب نيدأ بما يسمى بـ Evaluating a Hypothesis او تقييم الفرضية

- المقصود بيهما ازاي او لا اعمل تقييم للخوارزم اللي تم عندي عشان اعرف مدي جودته ، وقد ليه هو كوبس ولا ، وقل فيه OF OF ولا LA
- فقطلا في المسألة ديه ، عندا عدد كبير من الـ features ، فده غالبا بيسبب الـ OF والجراف لازق في النقاط
- يشكل مستقر ، فيهجي نسبة خطأ توقع النقطة الجائحة أقل كثير

Fails to generalize to new examples not in training set.

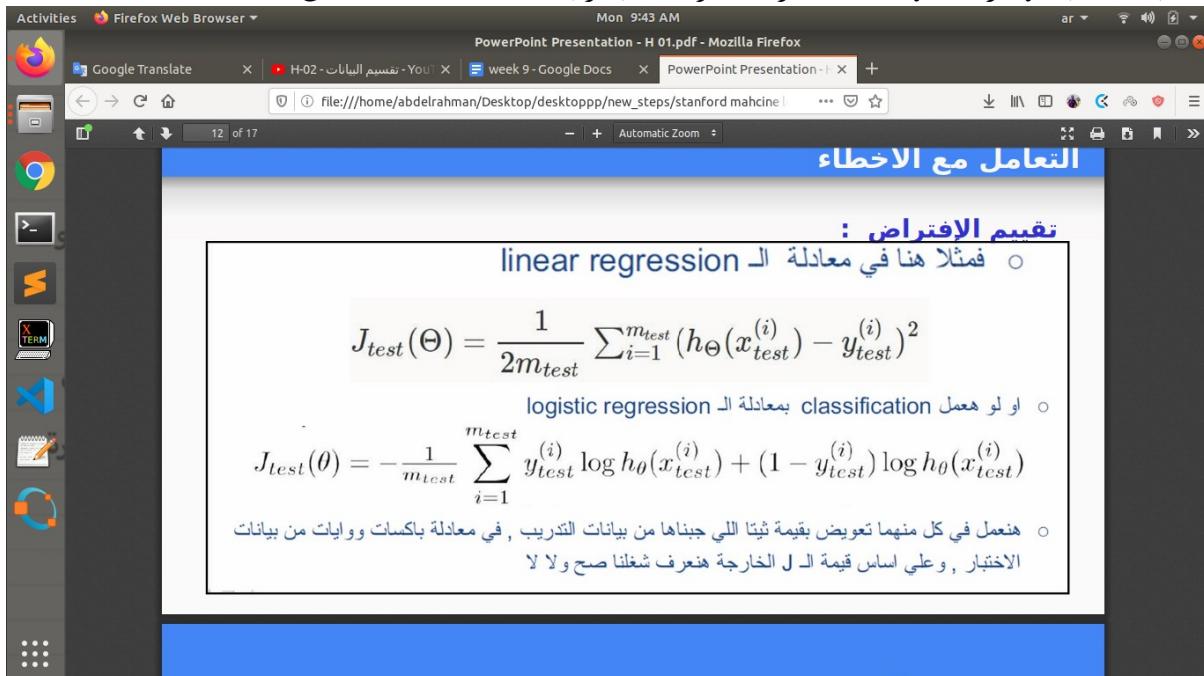
$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + \theta_3 x_3^3 + \theta_4 x_4^4$

$x_1 = \text{size of house}$
 $x_2 = \text{no. of bedrooms}$
 $x_3 = \text{no. of floors}$
 $x_4 = \text{age of house}$
 $x_5 = \text{average income in neighborhood}$
 $x_6 = \text{kitchen size}$
...

x_{100}

وده بيكون من خلال انى بعد ما عملت **training data** ابتدى بقى اشوف الـ **testing data** و ساعتها مقدر احدد ال **algorithm** بتابعى كوبس ولا من خلال ال **cost function** وكمان من قبل كده لو انلعملت ممكن اعرف ساعتها لو كان ال **algorithm** بيعمل over fitting ولا زى

الصوره الى فوق ف ساعتها هاكون عارف قبل ما اعمله **testing** انى ها يكون فيه نسبة خطأ كبيرة . من المفترض برضه إنك تكون عارف وبصيت على الداتا الى انتابتعامل معها قبل ما تيجي تنصف ال **data** الى عندك ل **training and testing** وعلى حسب النسبة الى هتقسم بيهما ممكن تكون 80 ل 20 او 70 ل 30 ولكن المهم فى كده انى الداتا نفسها متكونش مترتبة من الاصغر للابiger او العكس يعني عموما ال **data** كون متلغيطه فى بعضها او حاصلها **shuffle** والا هيكون انتابدرس على حاجه تانيه خالص اصلا.



فيه برضه مشكله تانيه بتبدأ تظهر وهى ال **hyper parameter** وهو القيم الى انا بحتاج احددها زى **lambda** وقيم **alpha** الأولى وقيمة ال **theta** طيب لو قررت انى اعمل **polynomial equation** ها عمل لل **d** كام . وهذا مثلا ممكن اعمل **d=10** مثلا ومع كل درجة احسب قيمة ال **theta** وقيمته الخطأ كام و ساعتها اشوف قلهم خطاء وابتدى اشتغل فيها ولكن اخللى بالى انى ممكن مع قيمه الأخطاء القليلة جدا يكون فيه **overfitting** وهذا ممكن مع كل **d** اعمل **graph** لشكل ال **.data**

Cross-Validation-Data

هى طريقة بتستخدم انى بدل ما كنت بقى ال **dataset** الى عندي لمجرد فقط لا هنا اذليقها لثلاث اقسام **training,cross and testing** وده يساعدنى فى انى بدل ما اوصل لل **algorithm** واكتشف انى قيمه ال **cost function** كبيره عشان مثلا ال **algorithm** كان بيعمل **over fitting** لا بلحق نفسى واغير الحاجات الى عايز اغيرها فى ال **cross validation** عشان بعدها اكون خالص ظبط دينتى وال **algorithm** يشتغل على ال **testing data** وانظممن . وال فكرة الأساسية من ال **cross-validation** هو انه بيعالج مشاكل ال **hyper parameter** الى اذليقها .

وتحديد النسبة بتاعت كل جزء من ال **raining, cross and testing** يختلف بإختلاف كمية ال **data** نفسها.

تقسيم البيانات

- استخدام بيانات الضبط
 - المقصود بيه ان هنقسم البيانات ، بدل ما كانت 70% تدريب ، و30% اختبار ، ه تكون 60% تدريب ، 20% اختبار
 - بيانات الضبط، يستخدم عشان تساعدني في ضبط قيم ثيتا ، ودرجة المعادلة ، عشان اتأكد ان الخوارزم سليم ، عشان لما يجيء ببيانات الاختبار او اي بيانات جديدة ، تبقى كفانته عالية
 - وميزته انه بيفصل بين OF و UF لو شاف اي حاجة فيه زي الصورة ديه

تقسيم البيانات

- يعني من غير بيانات الضبط باعمل التالي :
- اشوف قيم ثيتات من كل درجة (اللي بتجيبي اقل قيمة للـ J)
- اطبق الثيتات ديه في بيانات الاختبار عشان اشوف انه هي اقل لـ J
- وده مشكلته ان مع بيانات حقيقية بلاقي مشاكل
 - لكن مع بيانات الظبط
- اشوف قيم ثيتات من كل درجة (اللي بتجيبي اقل قيمة للـ J)
- اطبق الثيتات ديه في بيانات الضبط عشان اشوف انه هي اقل لـ J
- اشف قيمة الـ J مع بيانات الاختبار عشان اتأكد انها قليلة

Bais & Variance

الانحراف والتنوع
الـ **bais** هو ببساطة بيكون مع مشكله الـ **underfitting** وهو انى فيه ابعد او انحراف كبير جدا للنقط عن الـ **best fit line** لى انا عملته فتلقيه بيظهر كتير لما بتكون شغال بمعادلات **linear**.
لكن الـ **variance** بيظهر مع مشكله الـ **overfitting** ويكون ساعتها **high variance** وده بيظهر لما اجي استخدم معادلات من الدرجة العليا وهو انى ساعتها يكون عملت **fitting** لـ **training data** بشكل كويسي وده بيكون بسبب انى الـ **algorithm** مرن جدا وبيعدى على كل النقط لكن مع الـ **testing data** او **cv data** بتظهر المشاكل.

فلما بتكون عندي الـ **cost function** قليلة جدا فى الـ **training data** ده بيكون **overfitting** وفى نفس الوقت تكون ساعتها الـ **high variance** لأن بيكون ساعتها الـ **algorithm** مرن جدا وبيعدى على كل النقط.

ولما تكون قليله جدا فده معناه انى فيه انحراف كبير وهو انى النقط بعيده جدا عن ال best fit line .under fitting ب تكون كبيرة جدا فيحصل ما يسمى ال high bias .

الإنحراف والتلوّع

- التحيز والتلوّع
 - طيب لو تم اختيار الدرجة المناسبة لكن تظل الحسابات فيها مشاكل ، فممكن نشك في موضوع الانحياز Bias او Variance
 - التحيز الزائد يجعل UF لأن الخوارزم يتتجاهل كثير من البيانات ، او يجعل لها اهتمام قليل شوية ، و ده يؤدي لـ UF لأن الخوارزم مش هيكون بمثلي اغلب البيانات ، وده غالباً بييجي من الدرجات الاقل للمعادلة (المعادلة الخطية مثلاً) ، او صفر حجم الثبات
 - التلوّع الزائد معناه حساسية عالية لاي تغير في البيانات ، فالخوارزم يحال بجمع كل البيانات مع بعض في جراف واحد ، فيعمل OF ، وده غالباً بييجي من الدرجات العالية للمعادلة ، او زيادة قيمة الثبات

الإنحراف والتلوّع

- وعشان نفهم الكلام ده عملينا ، تعالى نقارن بين قيمتين
 - قيمة J_{train}(θ) = $\frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$ لـ الخاصية ببيانات التدريب
 - قيمة J_{cv}(θ) = $\frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$ لـ الخاصية ببيانات الضبط
- هنلاحظ في الرسم ده حاجة مهمة :

 - ان اولاً محور لكس هو عبارة عن درجة المعادلة المستخدمة ، بينما محور واي قيمة الخطأ
 - ان كل ما يزيد قيمة الدرجة كل ما كان الرسم رايج لـ OF وكل ما قلت كل ما كان رايج لـ UF

وال over fitting بيحصل بسبب انى فى ال training data بيعمل انه يصل لقيم الثيتات المناسبة جدا بضربيها فى الاكسات فنلقى ده بيظهر اكتر لما نيجي نزود ال degrees بتاعت المعاadle .
نلاحظ كمان انى لما بيكون قيمة ال cost function مع ال training data صغيره جدا ومع ال cv data كبيرة جدا ده بيكون ساعتها over fitting لان بيكون ال fitting algorithm عمل قوى جدا مع ال training data لكن مع بيانات جديدة بيحصل انى الخطأ بيكون كبير جدا .
ولكن لما تكون قيمة ال cost function بالنسبه لـ training and cv data كبيرة وكمان متقاربين مع بعض فده بيكون under fitting وبتلقيه لما نيجي نستخدم مثلاً معادلة من الدرجة الاولى وال algorithm بتعنى height bias مش عامل يعمل على data بشكل كويس فبيسبب under fitting او ما يسمى

Mon 10:56 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | H-04 - حجم العينة | PowerPoint Presentation - | +

8 of 17

الإنحراف والتنوع

- و لما يكون عندي مشكلة في البيانات ، وانا مش قادر احدد هي UF ولا OF ، ساعتها اشوف الرسم الجاي

لو كان كل من الضيـط والتـريـب كـبـيرـة بـيـقـي اـعـرـف ان درـجـةـ المـعـادـلـةـ قـلـيلـةـ ، وـهـدـ معـناـهـ انهـ UF

لو كان لـ الضـيـطـ كـبـيرـةـ ، وـلـ التـريـبـ قـلـيلـةـ ، بـيـقـيـ اـنـاـ فـيـ درـجـةـ عـالـيـةـ ، وـلـ هـ سـبـبـ المـشـكـلـةـ

Mon 10:57 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | H-04 - حجم العينة | PowerPoint Presentation - | +

9 of 17

الإنحراف والتنوع

- و عـلـاقـةـ التـنـوعـ Var~ianceـ بـالـاحـيـازـ عـلـاقـةـ عـكـسـةـ تـامـاـ فـيـ الـبـادـيـةـ الـدـرـجـاتـ الـقـلـيلـةـ تـنـوعـ قـلـيلـ (ارـقـامـ غـيرـ مـتـماـشـيـةـ مـعـ الـمـنـغـرـاتـ) ، وـالـاحـيـازـ كـبـيرـ عنـ الـحـقـقـةـ ، بـيـنـماـ فـيـ الـدـرـجـاتـ الـكـبـيرـةـ ، تـلـقـيـ تـنـوعـ كـبـيرـ (الـخـارـزـمـ بـيـرـوـحـ وـيـجـيـ مـعـ كـلـ نـقـطـةـ) ، لـكـنـ اـنـحـيـازـ قـلـيلـ لـانـهـ جـاـبـ الـأـرـقـامـ بـالـظـلـيـطـ

وبرهه قيمة ال lambda نفسها بتفرق في ال under fitting over fitting وال over fitting لانها في الآخر هتضرب في ال theta لى انلاصلا بحاول قللها عشان تقلل ال cost function فيحصل انى في الآخر ممكن توصل لل 0 وده بيسبب over fitting لانى مبكنتش استخدمت ال requilrization والعكس لو كانت كبيرة فلما تضرب في قيم ال theta لى انلاحاول قللها هاتطلع قيم ال under fitting theta جدا وبالتالي ده هايسبب انها تكبر ال cost function ويحصل

Mon 11:02 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | H-04 - حجم العينة | PowerPoint Presentation - | +

Activities Firefox Web Browser ▾

12 of 17

الإنحراف والتباين

و عشان ندرس تأثيرها على كل من الانحراف والتباين ، عايزين نعرف هي المدما بتعمل ايه اساسا

- زي ما شافين في المعادلة اللي فوق ، اللمنا بتضرر في مجموع الثبات سكوير (باستثناء ثبات صفر) ، والخوارزم بيحاول بقل قيمة L للاكبر قدر

Price

Size

Large $\lambda \leftarrow$

\rightarrow High bias (underfit)

$\rightarrow \lambda = 10000, \theta_1 \approx 0, \theta_2 \approx 0, \dots, h_{\theta}(x) \approx \theta_0$

Price

Size

Intermediate $\lambda \leftarrow$

"Just right"

Price

Size

\rightarrow Small λ

High variance (overfit)

$\rightarrow \lambda = 0$

Andrew Ng

Mon 11:01 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | H-04 - حجم العينة | PowerPoint Presentation - | +

Activities Firefox Web Browser ▾

13 of 17

الإنحراف والتباين

- فلو كانت اللمنا كبيرة جدا ، الخوارزم هيبطر يقل جدا قيمة كل الثباتات عدا ثبات صفر اللي مش مضروبة فيها ، وده هيخلني قيمة كل الثباتات تقريبا بصفر ، الا ثبات صفر اللي هنكون بقيمة عالية ، فيكون Hx يساوي تقريبا ثبات صفر ، فعنكون معادلة مش خطية ، لكن هن تكون خط افقي مستقيم يساوي ثباتا 0 ، زي الشكل الشمال ، وهن تكون UF
- ولو اللمنا قلت جدا ، فده معاناها ان الـ Regularization مش موجود تقريبا ، وبالتالي هنفضل الخوارزم OF زي ما في الشكل اليمين
- اما لو اللمنا متوسطة ، فمش هتقل الثباتات كثير ، لكن هتظميها بحيث يتلاشي الـ OF من غير ما ندخل في الـ UF

Mon 11:02 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | YouTube - حجم العينة | PowerPoint Presentation - | +

الإنحراف والتتنوع

- طيب نحدد قيمة لمدا المناسبة زي؟
- الاول نبدا نحدد قيمة لمدا اللي هجريبيها ، ويفضل نبدا برقم صغير زي 0.01 ونسميها لمدا 1، بعدها نضاعفه يكون 0.02 ويكون لمدا 2، ونضاعفه ثانى وهكذا ، لغاية لما نوصل لرقم كبير زي 10 اللي ه تكون لمدا 12

1. Try $\lambda = 0$
2. Try $\lambda = 0.01$
3. Try $\lambda = 0.02$
4. Try $\lambda = 0.04$
5. Try $\lambda = 0.08$
- ⋮
12. Try $\lambda = 10$

بعدها نجريها واحدة واحدة في مع بيانات التدريب ، عشان نجيب قيمة ثيتا اللي هتعمل اقل قيمة لـ J ، في المعادلة الخاصة بالرجو لاشن اللي هي

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^m \theta_j^2$$

هام جدا جدا

Mon 11:03 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | YouTube - حجم العينة | PowerPoint Presentation - | +

الإنحراف والتتنوع

- كدة هيكون عندينا ثيتا 1 اللي جت لما عملنا لمدا 1 ، وثيتا 2 اللي جت لما عملنا لمدا 2 ، وهكذا
- مش هبص على اقل قيمة لـ J دلوقتي ، لأن زي ما شفنا انها بتعمل مشاكل ، همسك كل ثيتا من غير ما اخذ المدا الخاصة بيها ، واحطها في نفس المعادلة اللي فوق مع بيانات الـ CV مثل التدريب ، بس المعادلة المرادي هتون من غير قيمة للمدا ، ونشوف اقل J فيهون ه تكون كام

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2.$$

- الـ J الاقل تدل على افضل ثيتا ، وبالتالي ، المدا المرتبطة بيها دي هي الافضل في الاستخدام
- وعشان اتأكد ، هجري نفس الثيتا في بيانات الاختبار ، بنفس المعادلة اللي فوق اللي من غير لمدا
- يعني لما هاجي اطبق بشكل حقيقي مع بيانات العمل ، هاستخدم المدا دي اللي جابت افضل ثيتا ، اللي جابت اقل J ، لكن احنا هنا موقفنا بنوقف تطبيق المدا عشان اتأكد هل الثيتا اللي خرجت هي بالفعل المثالية اللي مش هتعمل ولا UF ولا OF

وهو انى لما بعمل **cv** and **regularization** عمله فقط مع الـ **training data** اروح اجرب ساعتها فى الـ **.testing**

حجم العينة

مهم جدا اكون عارف تأثير حجم العينة يعني ممكن يكون فيه over fitting لـ **cost function** لـ **fitting** لكن لما ازود اكتر القى انى قيمة الـ **cost function** لـ **algorithm** قليله جدا بتزيد وده بسبب انى الـ **algorithm** بيكون بيحاول ان يعمد **fitting** لكل الـ **data** فى **graph** واحد.

ولكن ده فى الـ **cv data** ده بيكون كوييس جدا لان ساعتها حتى لما تكون قيمة الـ **J** زادات بسبب انى البيانات كبيرة ده بيساعد انه يقدر يعمل **prediction** كوييس لبيانات مشفتهاش قبل كده.

Mon 11:27 AM

PowerPoint Presentation - H 04.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | H-04 - حجم العينة | Google Search | مواعيد الصالات | PowerPoint Presentation - | +

6 of 13

Automatic Zoom

حجم العينة m

- فلاحظ في الرسم أن كل ما تزيد قيمة m كل ما تزيد نسبة الأخطاء لـ في بيانات التدريب

- لكن من البديهي ان بيانات الضبط CV وضعاها مختلف ، لأن مع العدد القليل جدا من بيانات التدريب ، صحيح ان لـ التدريب قليلة (ان الحراف لا يرق في النقط) لكن فعليا لـ الضبط هتني كبيرة جدا ، ليه ؟ لأن اساسا الخوارزم المعول بـ m قليل ، مستحيل يقدر يتوقع اي قيم مشاهدات ، وبالتالي قيمة لـ الضبط ه تكون ضخمة جدا
- كل ما تزيد قيمة m لبيانات التدريب ، كل ما تقل قيمة لـ لبيانات الضبط لأن الخوارزم يعني اقدر انه يتوقع بيانات مشاهدات

وأخل بالى لما بيكون عندي مشاكل الى هي over fitting ف ساعتها ايجاد data جديده واعملها الى هو زيادة نسبة الـ training data بيتسبب انى الـ cv data تقل فيها نسبة الـ error لكن على عكس الـ under fitting زبادة الـ data ش بيحل المشكلة ف ساعتها بشوف مثلا حاجه تانية زي الـ polynomial degree فسواء كده او كده لازم اكون عارف امتا استخدم ايه او المشكله ديه بتطلب انى اعمل ايه بظبط.

Mon 11:36 AM

PowerPoint Presentation - H 04.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | H-04 - حجم العينة | Google Search | مواعيد الصالات | PowerPoint Presentation - | +

13 of 13

Automatic Zoom

حجم العينة m

- طيب بعد ما تعلمنا عدد من التقنيات لعلاج المشاكل ، عاليزين نعرف ايه بيعتخدم لاي :

- زيادة العينة شوية وتجنب صيغه اكتر :
- يستخدم بين في حالة الـ OF و ان عندي تنوع كبير ، لكن لو التنوع قليل او الانحراف كبير ، وعندى UF فزيادة العينة يعني تضييع وقت
- تقليل عدد الـ features :
- برضه في نفس حالة الـ OF
- زيادة عدد الـ features :
- غالبا بتحل مشكلة الـ UF و الانحراف الكبير ، لأن زيادة عدد الـ features بيخلي الخوارزم عنده القدرة انه يتبع النقط بشكل افضل
- زيادة عدد و اس الاقسات polynomial :
- وده برضه هيل مشكلة الـ UF و الانحراف الكبير ، لأن كل ما بيزيد اس المعادله ، كل ما يكون لها القدرة انها تتباين اكتر و تتجنب الـ UF
- زيادة قيمة لمدا :
- تحل مشكلة الـ UF
- تقليل قيمة لمدا :
- تحل مشكلة الـ UF

Skewed Classes

يقصد بها إنى مش معنى انى الـ algorithm بتاعى بيكول انى الـ accuracy عندي يعني نسبة الصح كبيرة جدا فده شيء كويش لأن فى بعض الأحيان يحتاج انحراف او اروح ناحيه حاجة معينه او قيمه معينه اكتر من الثانية فمثلا لو بنتكلم انى آل algorithm بتاعى شغال انى يعمل predict للشخص مريض ولا مش مريض

ساعتها ممكن حاجه تزيد على حساب حاجة تانية وهو انى حتى لو توقعت انى الشخص ده مريض هيبقى احسن بكثير جدا من انى ققول انه مش مريض وساعتها ممكن اصلا تكون حاله حرجه ويحصل إنديمومت او غيره فهنا المفروض انى اكون عارف انى آل algorithm بتابعى هيروح ناحية ال precision وهى نسبة الدقة أو الظبط اكتر بمعنى آخر مثلا لو بنتكلم فى logistic regression وكنت بقول انه اكتر من او يساوى 5. هاتوقع حاجه معينه انه مريض مثلا ولو اقل هاقول لأ ده مش مريض هنا فى المثال ده مينفعش امشى كده فاحتاج انى أعلى قيمة ال threshold وهي العتبة او الحاجة الى من عندها ببتدئ اخد قرار كأنها Decision boundary فممكן هنا بقاللو عليت شويه تكون مثلا 7. يعني فيما فوق ال 7 هيكونوا لامش مرضى وغير كده هيكونوا مرضي وهنا الى انتابتصنفهم انهم مرضى في الآخر اكتر من او يساوى 7. وده طبعا مش صح لأنك في حاله زى كده تحتاج تعكس الموضوع.

فهنا بيظهر عندي مفهوم جديد ليه علاقة بحساب الدقة لأن مبتكونش بطريقه انى اجيبي الصح على التوتال لأنها average or accuracy لا بيكون فيه طريقه ثانيه وهي Matrix بقدر اعرف من خلالها الحاجة الى انا شغال عليها فلو هنا انا عايز احسب الناس المرضى فعلياً فهنا عشان اشتغل خلينا نقول مثال هو انى عندي 1000 واحد مشتبه فيه انه مريض سرطان لكن فعلياً فى 1000 واحد المشتبه فيه دوى فيه 5أشخاص مرضى سرطان فعلياً وجه ال algorithm بتاعىتوقع منهم 4 مرضى سرطان ولكن فى الحقيقى فيه 2 منهم فقط هما الى مرضى واتنين تانية اصحابه فهنا بقليتبدى احسب ال matrix ديه عن طريق اربع حاجات الى هما True Positive ,True Negative, False Positive and False Negative.

طيب هاتتحسب على اساس ايه ديه بتحتختلف على أساس القيمة الى انلغايز اتوقعها يعني انا هنا عايز اتوقع الاشخاص اللي هما مرضى سرطان فببتدى احسب اول حاجه الاشخاص الى انا توقعت انهم مرضى سرطان وفعلا هما مرضى سرطان وديه بتكون ال **True Positive** وفي المثال الى فوق هتلافقى انى آل **algorithm** توقع ان فيه 4 مرضى سرطان بس فعلا فيه اتنين منهم فقط هما الى مرضى فهنا فى الحالة ديه هيكون 2 فقط ال **algorithm** توقع انهم مرضى سرطان وهمما فعلا مرضى سرطان يعني $2 = \text{True positive}$

تاني حاجه ال True Negative وهى الاشخاص الى انا توقعت انهم مش مرضى وهم فعليا مش مرضى هنا
الباقي من 1000 واحد بعد اما توقعت انى فيه 4 مرضى هو 996 ولكن في حقيقة رغم انى ال algorithm
توقع انهم مش مرضى وهم فعليا مش مرضى لكن فيه خطأ وهو اصلا عارف انى فيه فعليا 5 مرضى سرطان
فلما توقعت فوق انى فيه 2 وطلعوا فعليا مرضى فهنا باقى من 5 تلاته فاكيده ال algorithm توقع 3 غلط انى
هما مش مرضى بس فعليا هما مرضى فتكون النتيجه انى آل True Negative هي 993.

ثالث حاجه هى ال False Positive ودية تعتمد برضه على الى انا بحاول اتوقع انا بحاول هنا اتوقع المرضى
فال False Positive بت Shawf القيم الى انا توقعتها صح لكن في حقيقة الأمر هى مش صح فلو لاحظت فوق انى
توقعت انى فيه 4 مرضى لكن في حقيقة الأمر هما كانوا اتنين مرضى واتنين لا فهنا ال False Positive تعتمد
على كام واحد توقعت انه مريض وهو اصلا مش مريض فنلاقى انهم 2 من ضمن ال 4 توقعت انهم مرضى بس
فعلا هما اصحابه.

آخر حاجه وهى ال False Negative وهى اتوقع انى هم مش مرضى لكن فعليا هما مرضى ودول هما ال 3 الى انا توقعهم من ضمن ال 996 غلط انهم مش مرضى كلهم ولكن كان فيه منهم ثلاثة مرضى.

نیجی بعد کده مفهومین مهمین الا وهما ال precision وال recall . ال precision ببساطة هى لو في مثال انى بحاول اتوقع المرضى فهى بتكون النسبة بين المتوقع الصحيح على مجموع كل المتوقع.

TP/ TP+FP

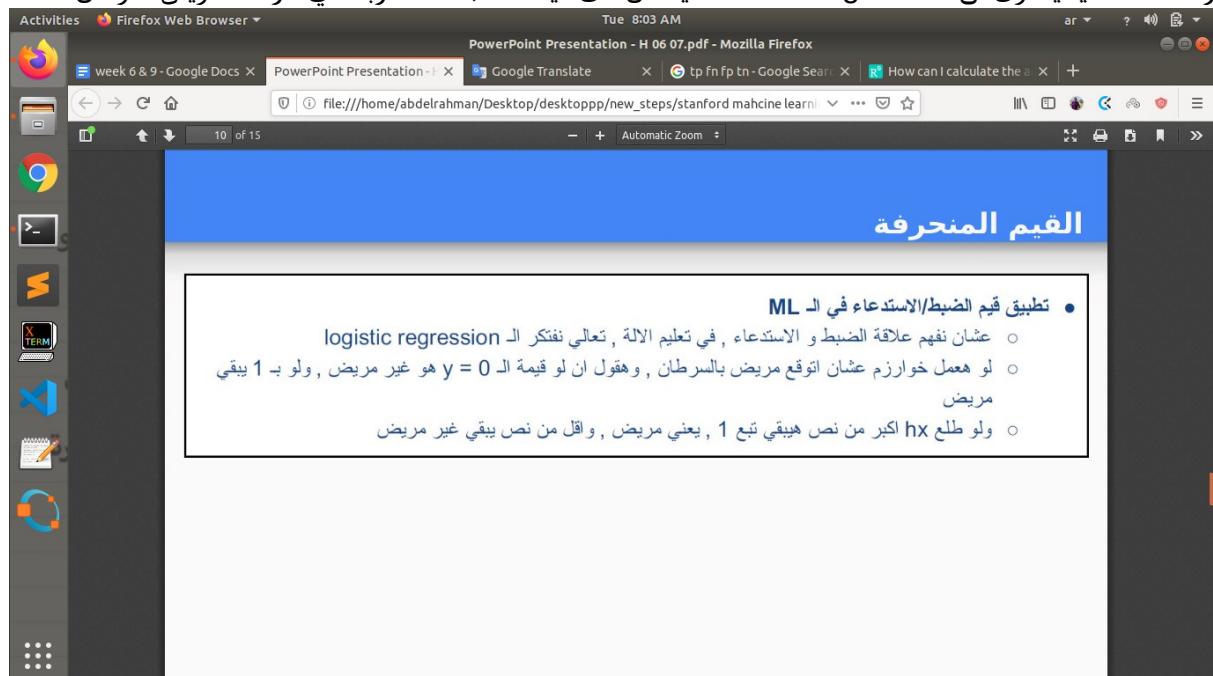
فه تكون على المثال اللي فات انا عندي TP الى هى الاشخاص الى انا توقعتهم مرضي وهما فعلياً مرضي 2 والاشخاص الى انا توقعتهم مرضي بس هما مش مرضي برضه 2/4 وبكده بيطلع عندي 50% من الى

انقللت عليهم عندهم سرطان هما فعلاً عندهم سرطان فتلاقي انى آل algorithm بتعاي ضعيف جداً لأن ده معناه نسبة قد ايه من ال algorithm بتعاي قال انه صح او مريض طلع فعلاً انه مريض فكتئه هنا 50% فقط.

لكن ال recall بيكون هو نسبة بقا المتوقع الصحيح بالنسبة للفعلى الى هو بضبط نسبة قد ايه من المفروض هو يتوقعه هو فعلاً توقعه وعرفه فهنا مثلاً ال algorithm المفروض يتوقع الناس المرضى بالسرطان وهو فعلاً كانوا 5 فنلاقي ان هو في الآخر معرفش يتوقع الا 2 فقط مريض سرطان لأنه لما توقع انهم 4 مطلعش ده صح ده طلع انى 2 مريض واتنين لا وال حقيقي او الفعلى الى عندي انلعارف انى فيه 5 مريض من ال 1000 المشتبه فيهم.

TP/TP+FN

الى هو ال algorithm توقع قد ايمريض وهو فعلاً مريض على نفس الكلام كام واحد قال انه مريض وهو فعلاً مريض + كام واحد مش مريض وهو قال انه مريض فنلاقي انها $\frac{2}{5}$ وده طبعاً رقم قليل جداً جداً حوالي 40%. معناه منطقياً يا ترى ال algorithm اكتشف اد ايه من ال فيه الحاجة المطلوبه الى هو هنا مريض سرطان.



ساعات كتيره على حسب انا شغال على ايه بحتاج زى ما كنت وضحت فوق انى ممكن تيجي تقول لا والله انا بحتاج فعلاً مقولش على حد انه مريض الا لما يكون فعلاً مريض فهنا نسبة ال threshold بتعلى فمثلاً لو كنت بتتوقع انى اكبر من او يساوى 5. ده كده هايكون مريض لا باخلى النسبة مثلًا 7. او اكبر فهنا فعلياً انا كده كبيرة نسبة ال precision الى هو بقا لما بتتوقع انى حد مريض فهو فعلياً مريض لكن مثلاً لو شاكك فى بعض الناس زى مثلاً النسبة بتاعتته 6. فهنا هاقول انه مش مريض فانا هنا برفع نسبة ال precision وده لو لاحظت هايزيود قيمى ال 0 الى هى الناس الى مش مريض ويقلل قيمة المرضي ولكن فى الآخر ده كده ممكن يعرض نسبة بعض الاشخاص للخطر الحقيقي الى هو لو كان 67. هايقول انه مش مريض معن ممكن يكون المريض ده مصاب بمرض خطير يسبب للوفاه فلازم اخلى بالى انا شغال على ايموهنا هلاحظ انى قيمة الظبط هاتزيد لان هاقل عدد الاصحاء وهاقول انهم مريض على عكس ال recall هاتقل لان ممكن يكون فعلياً يكون فيه مريض وانا مش هاقولهم زى النسبة ال 20% الى ان رفعتها من 5. لـ 7.

Tue 8:15 AM

PowerPoint Presentation - H 06 07.pdf - Mozilla Firefox

Activities Firefox Web Browser week 6 & 9 - Google Docs PowerPoint Presentation - H 06 07.pdf - Mozilla Firefox Google Translate tp fn fp tn - Google Search How can I calculate the a +

11 of 15

Automatic Zoom

الحالة الأولى :

- بفرض ان لسبب معين , مهم جدا اني مقولش لاي حد صحيح انه مريض , ومفيش مشكلة لو فيه مريض مقولتش انه مش مريض , يعني بالعربي مش هقول احد انه مريض الا لو كنت متتأكد جدا ساعتها هرفع قيمة الـ **threshold** اللي هي العتبة المحددة , بدل ما تكون نص هتكون 0.7 , وده هي عمل حاجة مهمة
- ان بعض اللي كنت شاكك انهم مرضى (0.6 مثلا) مش هقولهم انهم مرضى , وده هيزود قيم الـ 0 و يقلل قيم 1 شوية , صحيح انه هينقدر مجموعة من الاصحاء من اني اخضمهم , لكن هيعرض حياة بعض المرضى الحقيقيين للخطر اني مش هقولهم
- دلوقتي لأن قيمة الظبط معناها (نسبة قد ايه من اللي الخوارزم قال انه صحيحة . طبع فعلا صحيحة) فكدة قيمة الضبط هترزيد شوية لاني هقلل عدد الاصحاء اللي هقولهم انتو مريض
- لكن لأن قيمة الاستدعاء معناها (يا ترى الخوارزم اكتشف نسبة قد ايه من اللي فيه الحاجة المطلوبة) , فقيمة الاستدعاء هتقل شوية , لأن ممكن يكون فيه مرضى وانا مش هقولهم
- وطبعاً لو خليت العتبة 0.8 يبقى قيمة الضبط هترزيد اكتر و الاستدعاء هتقل اكتر
- فزيادة قيمة العتبة بيعلي الضبط و يقلل الاستدعاء

Tue 8:15 AM

PowerPoint Presentation - H 06 07.pdf - Mozilla Firefox

Activities Firefox Web Browser week 6 & 9 - Google Docs PowerPoint Presentation - H 06 07.pdf - Mozilla Firefox Google Translate tp fn fp tn - Google Search How can I calculate the a +

12 of 15

Automatic Zoom

الحالة الثانية: (العكس)

- بفرض ان لسبب ثاني , مهم جدا اني مفوتتش اي حد مريض , يعني اي حد شاكك فيه لازم اقوله انه مريض , ومفيش مشكلة لو فيه صحيح فلتله انت مريض مش ازمه كبيرة , يعني بالعربي اي حد هشك فيه هقوله انت مريض من سبيل الاحتياط
- ساعتها هقلل قيمة الـ **threshold** اللي هي العتبة المحددة , بدل ما تكون نص ه تكون 0.3 , وده هي عمل حاجة مهمة
- ان بعض اللي كنت شاكك انهم اصحاب (0.4 مثلا) هقولهم انهم مرضى , وده هيزود قيم الـ 1 و يقلل قيم 0 شوية , وده هينقدر مجموعة من المرضى انهي هعرفهم المرض, لكن هيخضر مجموعة اكتر من الاصحاء اللي شاكك في مرضهم و هقولهم و خلاص
- دلوقتي لأن قيمة الظبط معناها (نسبة قد ايه من اللي الخوارزم قال انه صحيحة . طبع فعلا صحيحة) فكدة قيمة الضبط هتقل شوية لاني هزود عدد الاصحاء اللي هقولهم انتو مريض
- لكن لأن قيمة الاستدعاء معناها (يا ترى الخوارزم اكتشف نسبة قد ايه من اللي فيه الحاجة المطلوبة) , فقيمة الاستدعاء هترزيد شوية , لأن تقريبا كل المرضى هيعرفوا
- وطبعاً لو خليت العتبة 0.2 يبقى قيمة الضبط هتقل اكتر و الاستدعاء هيزيد اكتر
- فتقليل قيمة العتبة بيعمل الضبط و ويزيود الاستدعاء

Tue 8:16 AM

PowerPoint Presentation - H 06 07.pdf - Mozilla Firefox

week 6 & 9 - Google Docs | Google Translate | tp fn fp tn - Google Search | How can I calculate the a | +

13 of 15 | Automatic Zoom |

القيم المترفة

Precision

Recall

threshold = 0.09

threshold = 0.01

- محور اكس الاستدعاء , محور واي هو الضيـط
- لما كانت العتبة كبيرة (فوق على الشمال) كان الضيـط كبير و الاستـدعـاء قـليل
- العـكـس ، لما كانت العـتبـة قـليلـة (تحـتـ على الـيمـين) الضـيـطـ قـلـ و الاستـدعـاءـ كـبـيرـ
- ولو هـذـيـ العـتبـةـ فـيـ النـصـ ، كـلـ مـنـ الضـيـطـ وـ الاستـدعـاءـ هـيـقـضـلـوـ مـتوـسطـينـ
- وـ بـيـنـهـمـ مـمـكـنـ الـجـرـافـ يـقـيـ خـطـيـ اوـ كـيـرـفـيـ فـوـقـ اوـ تـحـتـ حـسـبـ اـتـجـاهـ الـقـيـمـ

Tue 8:16 AM

PowerPoint Presentation - H 06 07.pdf - Mozilla Firefox

week 6 & 9 - Google Docs | Google Translate | tp fn fp tn - Google Search | How can I calculate the a | +

15 of 15 | Automatic Zoom |

القيم المترفة

- نفس المـوضـعـ ، لوـ العـتـبةـ زـادـتـ ، هـنـقـلـ قـيـمةـ الـاستـدعـاءـ وـ يـزـيدـ الـضـيـطـ ، وـ يـطـلـعـ مـتوـسطـ جـمـيلـ لـكـنـ الـكـفـاءـةـ
- وـ حـشـةـ
- فالـحـلـ اـسـتـخـدـمـ قـيـمةـ ماـ يـسـمـيـ F1 Scoreـ الـيـ مـعـادـلـهـ كـهـدـهـ :
$$F = \frac{2PR}{P+R}$$
- هناـ هـنـلـاقـيـ انـ الـقـيـمـ الـنـهـائـيـةـ ، مـعـتمـدـهـ عـلـىـ قـيـمـ الـضـيـطـ وـ الـاسـتـدعـاءـ معـ بـعـضـ ، لوـ وـاحـدـ فـيـهـمـ قـلـيلـ جـداـ ، الـإـفـ
- هـنـقـلـ بـرـضـهـ وـمـمـكـنـ تـوـصـلـ لـصـفـرـ
- وـ الـقـيـمـ الـمـثـالـيـةـ ، انـ الـخـواـرـجـ يـكـشـفـ كـلـ الـمـرـضـيـ ، وـيـسـتـعـدـ كـلـ الـأـصـحـاءـ ، وـسـاعـتـهـاـ كـلـ مـنـ الـضـيـطـ
- وـ الـاسـتـدعـاءـ هـيـكـوـنـوـ بـ 1ـ ، وـقـتـهاـ هـيـكـوـنـ Fـ بـوـاحـدـ
- ولوـ اـيـاـ مـنـ الـاسـتـدعـاءـ اوـ الـضـيـطـ بـصـفـرـ ، الـ Fـ هـيـكـوـنـ صـفـرـ فـورـاـ
- قـيـمـةـ الـ Fـ تـنـتـراـوـحـ بـيـنـ الصـفـرـ وـ الـوـاحـدـ ، وـهـوـ اـفـضـلـ مـمـثـلـ عـنـ كـفـاءـةـ الـخـواـرـجـ

تجهيز البيانات

احتياجيـ لـلـ dataـ بـيـخـتـلـفـ مـنـ مـوـقـعـ لـلـتـانـيـ بـمـعـنىـ أـنـ زـيـادـ الـبـيـانـاتـ مـمـكـنـ يـكـونـ مـفـيدـ وـسـاعـاتـ لـاـ يـعـنـىـ لـماـ اـكـونـ بـتـعـالـمـ مـعـ الـمـوـقـعـ الـيـ فـيـهـ predictـ لـلـكـلـمـةـ الـقـادـمـةـ فـيـ textـ اوـ voice recognitionـ وـغـيـرـ مـنـ الـاـمـثلـةـ الشـبـيـهـهـ هـنـاـ زـيـادـ الـبـيـانـاتـ بـيـكـونـ مـفـيدـ جـداـ اـنـهـ الـ algorithmـ يـشـوـفـ أـمـثـلـةـ مـخـتـلـفـ وـكـثـيرـ لـكـنـ فـيـ حـاجـهـ زـىـ تـوـقـعـ سـعـرـ بـيـتـ زـيـادـ الـبـيـانـاتـ لـوـ كـانـ الـ featuresـ لـىـ عـنـدـيـ قـلـيلـهـ جـداـ سـاعـتـهـاـ الـزـيـادـهـ دـيـهـ مـشـ هـتـفـيدـ فـهـنـاـ فـيـ نـمـوذـجـ بـشـريـ بـيـتـعـلـمـ إـلـىـ هـوـ لـوـ حـاجـهـ إـلـيـانـ الـبـشـرـيـ يـقـدرـ يـسـتـنـتـجـهـاـ فـتـمـاـ غـيرـ كـهـدـهـ يـبـقاـ زـيـادـ الـبـيـانـاتـ مـشـ صـحـ.

وكمان زيادة البيانات يؤدي انى كفاءة ال algorithm بتكون اعلى بكثير جدا وده مع اختلاف الحاله الى احنا بنتعامل معها.

وده بيقود انى هنا فيه اختلاف يعني امته تكون زيادة البيانات شيء كويسيس وامثلًا والمثال ده بيوضح .

تجهيز البيانات

الحالة الأولى

- لو المشكلة عندي اني همل فراغ في كلمات مثل زي :
In the breakfast I ate ----- eggs
- وكانت الاختيارات اني هحط واحدة من الكلمات ديه :
Too - two - to
- فهنا لو ان الخوارزم شاف كمية كبيرة من البيانات زي ديه , هيدقير يستنتج اي مسألة زي كدة

الحالة الثانية

- لو المشكلة عندي اني عايز اجيب سعر بيت عن طريق مساحته بس , من غير اي معلومات تانية , هنا عدد الـ features قليل , ولو جبت 500 الف بيانات زي كدة , الخوارزم مش هيعرف يجيب البيانات بدقة ابدا
- يعني هنا زيادة البيانات مش هيغيفد حاجة الا ضياع الوقت فيها

تجهيز البيانات

- طبعاً ازاي بينهم , ونعرف امته زيادة البيانات مفيدة ولا لا :
- في الحالة الأولى لو عرضناها لبشرى متخصص (مدرس انجليزي) وقلنا له يا ترى هنحط كلمة ايه , هيعرف يجاوبها فورا , وهنا زيادة البيانات هتفيد
- في الحالة الثانية لو عرضناها لبشرى متخصص (مسار) وقلنا له سعر البيت الـ 200 متر , مش هيعرف يجاوب طبعاً لانه يحتاج يعرف بيانات تاني و هنا زيادة البيانات مش هتفيد

ونختم بحاجة مهمة :

- البيانات الكبيرة , بتكون مفيدة في حالة عندي features كثيرة , لأن البيانات الكبيرة بتقلل التتواء Variance
- والـ features الكثيرة بتقلل الانحراف bias
- والحاله المثلية , ان يكون كل من التتواء والانحراف قليلين مع بعض , فاتجنب الـ OF & UF مع بعض
- بينما لو واحدة كثيرة وواحدة قليلة هقع في مشكلة OF او UF

Anomaly Detection

القيم الشاذة هي بعض القيم التي تخرج بره إطار معظم ال data الى عندي وديه لازم اخذ بالى منها لانها ساعات بتؤدى لأضرار أو انها حاجه كويسيه فى حالات معين ولازم اهتم بالحالات الى زي كده.
وهنا إنتا بتحدد القيم الشاذة ديه ومدى اهتمامها بناء على إنتا شغال على ايم بظبط.

يصعب جداً إنك تحدد القيم الشاذة ديه أثناء عملية ال training ولكنها بتظر بشكل واضح مع ال testing او البيانات الى ال model مشفتهاش قبل كده وكمان لو كان فيه بعض القيم الشاذة أثناء ال training فممكنا انها تتمسح لأنها ممكنا تؤدي ل misleading في ال .data

Anomaly detection is a technique used to identify unusual patterns that do not conform to expected behavior, called outliers.

Ex:

Point anomalies: A single instance of data is anomalous if it's too far off from the rest. *Business use case:* Detecting credit card fraud based on "amount spent."

Gaussian decent or Normal Distribution

هـى الطريقة الى بقدر من خلاـلها أعرف القيم الشاذـة وهي عبارة عن طـرـيقـة بـيـتم حـاسـبـ فيها meue and feature على حـدـه عـشـان قـدرـ اـحـدـ الـ rangeـ بـتـاعـكـ اوـ الـ sigmaـ لـكـلـ هـتـكـونـ بالـنـسـبـهـ لـيـاـ anomaly pointsـ واـخـلـىـ بـالـيـاـ اـكـونـ عـارـفـ انهـيـ feaـtـureـ بـخـتـارـهاـ مشـ اـىـ حاجـهـ وـخـلـاصـ لـانـ لوـ اـخـتـارـتـ وـخـلـاصـ دـهـ بـيـقـودـ غالـباـ لـ misleadingـ



Activities Firefox Web Browser ▾ Wed 1:13 PM

PowerPoint Presentation - H 10.pdf - Mozilla Firefox

(2) H-12 - Google Docs | Google Translate | PowerPoint Presentation - | +

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ ... Automatic Zoom

القيم الشادة والـ GD

(2) حساب قيمة ميو μ :

يعني لو عندي الف عميل , وكل عميل عنده 50 معلومة , فهجيب متوسط كل معلومة فيهم , عن طريق اني اجمع قيمة المعلومة الاولى (الطول مثلا) لลألف عميل , واقسمها على الالف , بعدها اجمع قيمة المعلومة الثانية (الوزن) و اقسمها على الالف وهكذا

$$\underline{\mu_j} = \frac{1}{m} \sum_{i=1}^m x_j^{(i)}$$

عشان كدة مكتوب ان عشان اجيب ميو 5 مثلا (المعلومة الخامسة او العود الخامس) هجمع كل قيم اكس في العمود الخامس لكل الصفوف , واقسمها على عدد الصفوف m

Activities Firefox Web Browser ▾ Wed 1:19 PM

PowerPoint Presentation - H 10.pdf - Mozilla Firefox

(2) H-12 - Google Docs | Google Translate | PowerPoint Presentation - | +

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ ... Automatic Zoom

القيم الشادة والـ GD

(4) إيجاد قيمة الإحتمالية :

يعني عندي عميل بنك جديد , فيه أربع خصائص معلومة feature اللي عنده , واجيب احتمالية كل معلومة بناء على الميو و السيجما الخاصة بالمعلومة ديه

يعني اطبق القانون ده $p(x_j; \mu_j, \sigma_j^2)$ على مثلا قيمة حسابه الدولاري (feature) معين و استخدم الميو و السيجما اللي حسبتهم من الحسابات الدولارية للألف عميل سابق

اكرر خطوة الاحتمالية لباقي الـ features (كل الأعمدة) II

اضرب كل الاحتماليات في بعض عن طريق الـ II

اخيرا , لو لقيت ان قيمة P النهائية (الاحتمالية) اقل من ايسلون ϵ اللي تم تحديدها , يبقى العنصر ده في مشكلة و ابدا اخذ بالي منه

الـ feature تكون هي ال range الى بقع فيه النقط بتاعتي وده بيختلف مع كل σ

Activities Firefox Web Browser ▾ Wed 1:23 PM

PowerPoint Presentation - H 10.pdf - Mozilla Firefox

(2) H-12 - Google Docs | Google Translate | PowerPoint Presentation - | +

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ ... Automatic Zoom

القيم الشادة والـ GD

العلاقة بين متغيرين :

- دلوقتي لما نقيم نقطة جديدة ، هيكون ليها اكس 1 و اكس 2 ، فلو كانت النقطة في قلب الزحمة ، مثلاً نقطة 1 :

- هتلaci ان اساقطها هيكون في قلب الجرس 1، والجرس 2، يعني في قلب الجرس المجمم ، يبقى الـ P اكابر من ε
- بينما لو لقينا نقطة بعيدة شوية زي 2 ، هي موجودة في جرس 1 ، بس بعيدة عن حدود جرس 2 ، ساعتها هتخرج برة الجرس المجمم ، يعني الـ P اقل من ε
- معنى كده ان اي نقطة ه تكون تحت سطح الجرس المجمم ، تبقى مقبولة ، اي نقطة فوق سطحه ، تبقى شادة

ازايقدر الاحد او الكشف عن البيانات الشادة الى عندي وايماصلا الى يخليني قول انى ديه قيم شادة.
لما يكون عندي القيم الشادة مثلاً هى عدد قليل جداً من النقط بالنسبة للـ **data** ثلا 20 نقطة من وسط آلاف
فمساحتها يعتبر فعلاً انى ديه قيم شادة وابتدى اشوفها على حده لكن لو كان مثلاً 20% من البيانات قيم شادة
فهنا أنا مش بتعامل معها كأنها قيم شادة ولاحتاج انى اكتشفها لأنها بتكون جزء اصلاً من الـ **data** بتاعتنى
وبتعامل معها عادي زي اى **algorithm** هو الى يحددها لكن فى حالة انى فى حالة اى فى مثلث 20 نقطة
من وسط آلاف النقاط فهنا لازم اخلى بالى وامشى بطريقه من الطريق وهى:
انى لو كان عندي مثلاً 20 نقطة شادة من وسط 10000 نقطه فهنا هبتدى قسم الـ **data** بتاعتنى عن طريق الـ
cross validation مثلاً 60% و 20% و 20% والـ 60% الى هما دول هاخدتهم من الداتا الى
مفهاش اى شواز يعني ه تكون الداتا العادي جداً الى مفهاش اى حاجه وها قسم الـ **cv** 20% والـ 20%
هافقسم فيهم القيم الشادة نصين نص **testing validation** ونص للـ **testing** فتكون هنا مثلاً 2010 و 2010 للـ **.testing and validation**

Wed 1:33 PM

PowerPoint Presentation - H 11.pdf - Mozilla Firefox

(2) H-12 - Google Docs | week 6 & 9 - Google Docs | Google Translate | PowerPoint Presentation - | +

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/

5 of 10 Automatic Zoom

○ يعني ممكن لو البيانات السليمة 10 الاف ، والغلط 20 تكون كدة :

Training set: 6000 good engines
 CV: 2000 good engines ($y = 0$), 10 anomalous ($y = 1$)
 Test: 2000 good engines ($y = 0$), 10 anomalous ($y = 1$)

• والخطوات تكون كالتالي :

- أولا يتم تقسيم البيانات بالشكل الموضح اعلاه
- ثانيا اتناول بيانات الـ training (التي مفيهاش اي شواد) عشان اخلي الخوارزم ، يقدر يحدد قيم ميو و سيجما
- ثالثا امسك بيانات الـ CV او test (اللي فيها شواد) و ابدأ اعمل اختبار ليهم ، وهبطع حاجة من اربع حاجات
 - اما True Positive يعني هي اصلا بايطة ، والخوارزم اكتشفها
 - او false positive الخوارزم قال انها بايطة بس هي اصلا سليمة
 - او true negative يعني هي سليمة والخوارزم قال انها سليمة
 - او false negative يعني الخوارزم قال انها سليمة بس هي بايطة
- رابعا احسب قيم الـ Precision Recall , F1 Score

• ولاحظ ان احد الطرق اللي بحد فيها قيمة ايسلون ε هي اني اعمل فيه مختلفة ليها في المثال اللي فات ، وشووف انها F1 Score فايم بتعمل اعلى

اخلى بالى برضه انى فى الاخر القيم الشاذة ديه مبيكونش ليها pattern عين إنتلاقتدر تتبعه او تمشى عليه او تحلله وتعلمه معادله او حاجه لكن بتكون مختلفه اختلاف كلها كل نقطة عن الثانية ممكن لو هنتكلم فى معين الي هو الطبيعي مثلًا من 100 ل 1000 نلاقى قيمة مثلًا بتكون في رنج 0 ل 5 وقيم تانية فى رنج 100000 ل 1000000 فهي بتختلف من قيمة للثانية مش كلها كده قيم شاذة بتقع فى محيط واحد ولو كانت بتقع فى محيط واحد او range معين فهنا انا مش هاعتبرها قيم شاذة اصلا واتعامل معها عادي.

Wed 1:40 PM

PowerPoint Presentation - H 11.pdf - Mozilla Firefox

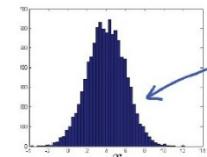
(2) H-12 - Google Docs | week 6 & 9 - Google Docs | Google Translate | PowerPoint Presentation - | +

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/

9 of 10 Automatic Zoom

• ضبط عناصر خوارزم القيم الشاذة

○ في بعض الاحيان ، تكون قيم الافتراضات لدينا جاهزة ، و يتم رسمنها ببساطة بشكل جرس جوسينيان زي كدة



○ و حتى لو لمكانش الشكل دقيق 100 % ، لكن هو يقاس على جوسينيان ، وبالتالي مش هاعمل فيه حاجة ، وهجيب الميو و السيجما بشكل طبيعي

○ لكن ممكن اجي ارسمها الاقبها بشكل مختلف زي ده



Activities Firefox Web Browser ▾

PowerPoint Presentation - H 11.pdf - Mozilla Firefox

Wed 1:41 PM

(2) H-12 - Google Docs X week 6 & 9 - Google Docs X Google Translate X PowerPoint Presentation - H 11.pdf - Mozilla Firefox

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ ... Automatic Zoom +

اكتشاف القيم الشاذة

- بقى هنا عايزين نعمل تغيير في معاملات الدالة ، عشان تتضبط ، وده اللي اسمه ضبط عناصر خوارزم القيم الشاذة
- بشكل عام عشان احول من الشكل ده \rightarrow ، لده \rightarrow ، محتاج العاب في الدالة نفسها
- يعني ممكن بدل ما هي اكشن ، اخليها حاجة من دول مثل :

 - $\log x$
 - $\log(x + c)$ ===== c is constant
 - \sqrt{x} ===== whatever the root
 - x^3 ===== whatever the power

- يعني لنا الامكانية بتغيير دالة اكشن باي تعامل رياضي ، عشان اوصل لجرس جوسين
- وطبعاً لما اوصلناها ، بقى اي قيمة شاذة عشان اكتشف عليها ، هعمل لها نفس التحويل

بتقابلنا مشكلة تانية وهي تحديد قيمة الشوادع مع اكتر من تغير لأن ساعتها بيكون في اختلاف كون القيمة شاذة ولا لأن ممكن يكون كل feature لي حداً متقبل القيمة ديـة بـس ارتباطـ الـ features مع بعض بيخلـي النقطـ ديـه تـبـان انـها شـاذـة فـهـنـا الـ correlation بيـوضـلـي مـدى اـهمـيـة اـكـتـر مـن feature ليـ النقـاط الـى عنـديـ مجردـ انـها اـشـوفـها بالـنـسـبة لـكـل قـيـمة عـلـى حـدـاـ.

Activities Firefox Web Browser ▾

PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

Wed 3:39 PM

week 6 & 9 - Google Docs X Google Translate X PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ ... Automatic Zoom +

استخدام أكثر من متغير

- لو قلنا دلوقتي ان عندي قيمة جديدة لونها اخضر ، عايز اعرف هي عاديـة ولا شـاذـة

x_1 (CPU Load)	x_2 (Memory Use)
0.4	1.5
0.5	0.3
0.6	0.4
0.7	0.5
0.8	0.6
0.9	0.7
1.0	0.8
1.1	0.9
1.2	1.0
1.3	1.1
1.4	1.2
1.5	1.3
1.6	1.4
1.7	1.5
1.8	1.6

زيـ بـطـبـطـ النـقـطـةـ الخـضرـاءـ واضحـ انـهاـ جـوهـ الـقـيمـ الخـاصـةـ بـ x_1 ـ لـوحـدهـاـ وـنفسـ الـكـلامـ بـالـنـسـبـهـ لـ x_2 ـ ولـكـ مجردـ النـظـرـ إـلـيـهاـ بـالـنـسـبـةـ لـلـ graphـ 2dـ معـ بـعـضـ بـتـلـاحـظـ انـهاـ بـعـيـدةـ جـداـ عـنـ تـكـلـ النـقـطـ حتـىـ تـزاـيدـ النـقـطـ نفسـهاـ بـيـزـيدـ بـطـرـيقـةـ مـعـيـنةـ لـماـ x_1 ـ تـزاـيدـ الـ x_2 ـ بـتـزاـيدـ وـهـكـذاـ لـكـنـ النـقـطـةـ الخـضرـاءـ لـوـ لـاحـظـناـ أـنـهاـ رـغـمـ اـنـ x_1 ـ قـلـيلـةـ جـداـ وـلـكـنـ قـيـمةـ x_2 ـ كـبـيرـةـ جـداـ فـهـنـاـ اـرـتـبـاطـ النـقـطـتـيـنـ بـبعـضـ بـيـخـلـىـ النـقـاطـ الشـاذـةـ تـوضـحـ عـلـىـ عـكـسـ كـلـ featureـ لـوحـدةـ.

Wednesday 3:45 PM

PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

Activities Firefox Web Browser week 6 & 9 - Google Docs X Google Translate X PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox Problem Motivation - St X graph - Google Search X + file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ 7 of 31 Automatic Zoom

استخدام أكثر من متغير

من الرسم واضح أنها شادة ، لأن استهلاك الـ CPU قليل مع استهلاك ذاكرة كبير ، وحتى بادئه انها بعيدة عن الرسم

x_2 (Memory Use)

x_1 (CPU Load)

Wednesday 3:45 PM

PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

Activities Firefox Web Browser week 6 & 9 - Google Docs X Google Translate X PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox Problem Motivation - St X graph - Google Search X + file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford machine learning/ 8 of 31 Automatic Zoom

استخدام أكثر من متغير

لكن المشكلة أن عرض القيمة دي لوحدها في اكس 1 او 2 اتنين هيبيان أنها طبيعية (النقطة الخضراء هنا و هنا)

$p(x_1; \mu)$

x_1 (CPU Load)

$p(x_2; \mu)$

x_2 (Memory Use)

وهنا ببتدى انى احسب قيمة ال p لكل ال features مع بعضها

استخدام أكثر من متغير

- و السيجما غالبا تكون **diagonal matrix** مصفوفة قطرية , كل عناصرها اصفار ما عدا القطر اللي هو سيجما 1 , سيجما 2 , وهكذا (امحرافات معيارية)

$$\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n \end{bmatrix}$$

- تعالى نشوف رسوم عملية عشان نفهمها كويس :
- لو قلنا ان عندي اكس 1 و اكس 2 , وان الميو بصفرين , والسيجما زي هنا

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

استخدام أكثر من متغير

- فيكون الحل , اتنا منحسبش كل Σ لوحدها , لكن Σ المجمعة ليهم

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

- القانون الغريب ده , بيجمع كل تفاصيل الاكتسات مع بعض , ويهجيب الـ Σ بشكل دقيق
- و متنساش ان الـ Σ هي المصفوفة اللي اتكلمنا عنها قبل كدة , وابعادها $n \times n$
- كمان ان $|\Sigma|$ معناها قيمة المصفوفة و بتتعمل بامر \det في اوكتيف او ماتلاب

$$\frac{1}{m} \sum_{i=1}^m x^{(i)}$$

- متنساش ان الميو بتكون فيكتور $n \times 1$ و قيمتها تساوي

$$\frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

- بينما السيجما بتكون مصفوفة $n \times n$ و قيمتها

Wednesday 3:58 PM

استخدام أكثر من متغير

- التطبيق في الكشف عن الشواد :

- هنستخدم ما ذكرناه عن القيم المتعددة في الكشف عن الشواد
- والخطوات هتكون كالتالي :
 - تحديد قيم ميو و سيجما من القوانين الخاصة بيهم

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$
$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

Wednesday 3:58 PM

استخدام أكثر من متغير

- ايجاد قيمة P عن طريق القانون الخاص بيه

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$

- في حالة عدم تواجد ابسلون ، تحديد قيمتها عبر تكرار فرض قيم لها و اختيار اعلى قيمة لـ **F1 Score**
- تحديد اي نقطة يكون قيمه P فيها اقل من ابسلون عشان تكون نقطة شاذة

Wednesday 3:59 PM

PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

week 6 & 9 - Google Docs | Google Translate | PowerPoint Presentation | Problem Motivation | graph - Google Search | ... | 04 - Google Drive | +

رسم الموديل المناسب بناء على قيم ميو و سيجما

- متباين في حالة متغير واحد ، القيمة النهائية لـ P تكون حاصل ضرب قيم p الصغيرة

$$p(x) = p(x_1; \mu_1, \sigma_1^2) \times p(x_2; \mu_2, \sigma_2^2) \times \cdots \times p(x_n; \mu_n, \sigma_n^2)$$

- بينما في حالة اكتر من متغير

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$

<https://drive.google.com/drive/folders/10pUUwl7sr4mQhdJM8evvKqxX9KGC1gTU>

هام جدا اللينك الى فوق فيه توضيح اكتر.

Recommendation System

نظام الترشيحات هو نظام يعتمد بشكل عام على ال Reinforcement learning الى هو بناء على ال action اللى انت باتخادها ببتدئ ارشحلك حاجة معينه زى لما تيجي تخش على اليوتيوب هنا اليوتيوب فى الاول خالص لو انتا لسه عامل اكونت جديد مبيكونش عارف حاجه عنده فبببتدئ يرش لك حاجات randomly بعد فتره من استخدام اليوتيوب لو انتابتسيرش على حاجات معينة او بتحش على فيديوهات معينة باستمرار وتتابع بعض الفيديوهات للاخر فى مجال معين او بتحب تسمع فيديوهات دينية او فيديوهات خيال علمي هنا اليوتيوب بببتدئ يعملك بقا recommendation ل حاجات قريبه جدا من الى انتا بتتشوفها او بتدور عليها وده بيساعد جدا فى انه النظام يخليك تستخدمه اطول وقت ممكن ومتقدرش تستغنى عنه زى الفيسبوك. وفي نفس الوقت تقييمات الأشخاص مثلا للأفلام ممكن حد يكون دايما بيحب يتفرج على أفلام أكشن وتلقيه بيعمل rate الى لبعض الأفلام الأكشن الى شافها وعجبته بينما الرومانسى ممكן يعمل ضعيف جدا ولو اتفرج على الفيلم ممكן يقدر يقدم الفيلم او يشوف النهاية علطول فهنا حتى لو معملش system rate ال يسنتنچ انك مش بتحب النوعية دي من الأفلام او بتحبها بناء على انك شوفت الفيلم كله ولا لا طيب هل قدمت كتير من الفيلم عشان تشويف اخره وخلاص على كده هل عملت rate فهنا سواء انتا عملت rate او معملتش بس اتفرجت على نوعية معينة من الأفلام ال system بببتدئ يفهم ميولك ناحية ايهم برتبط حتى الحاجات الى بتتميل ليها فيه منها بببقة كتير جدا بس انتا لسه مشفتهوش فيبدأ يعملها rate حتى لو انتا لسه مشفتهاش بالتقريب مع نفس الحاجات الى انتا شوفتها اصلا وحييتها من نفس النوعية ويملى القيم الفارغة فببتدئ يحصلك بالحالات الى لسه مشفتهاش من نفس النوعية ويبعد عن الحاجات الى مش بتحبها.

The screenshot shows a Firefox browser window with the title bar "H 14.pdf - Mozilla Firefox". The address bar displays the URL "file:///home/abdelrahman/Desktop/desktoppp/new_steps/stanford mahcine learr". The main content area shows a slide from a presentation. The slide contains the following text in Arabic:

■ فلما دا ML يعمل التقييم و يظهر عنده انه كان ممكن يعمل تقييم كبير لافلام كذا كذا ،
يعرضها عنده و يخفي الافلام اللي كان هي عمل عليها تقييم سلبي

■ مش بس التقييم هو الاساس (احنا واخدinه مثل عشان واضح) ممكن يكون من ضمن العناصر ، انه شاف فيلم لآخر ، او انه شافه مرتين ، او انه بيبحث عن اسم فيلم معين ، او بيفتح نوعية افلام و هكذا

○ و يجب ان تعرف ان :

- عدد المستخدمين نسميه N_u
- عدد المستخدمين نسميه N_m
- قيمة $r_{(i,j)}$ ستكون بصفر اذا لم يتم التصويت ، و 1 اذا تم التصويت
- قيمة $y^{(i,j)}$ ستكون بقيمة التصويت اذا تم التصويت

- و بناء نظام الاقتراحات له العديد من التحاولات , منها ما يسمى Content Based Recommendation الترشيح بناء على المحتوى
- الترشيح بناء على المحتوى CBR
 - نمسك المثال السابق , ونعمل دائرة على القيم الناقصة
 - اي فيلم بيكون فيه عدد كبير من الـ features , نقول ان عندنا 2 , هما :
 - مقدار الرومانسية في الفيلم X_1
 - مقدار الأكشن في الفيلم X_2
 - نقول ان كل فيلم من الافلام الخمسة , فيه مقدار كذا من X_1 و مقدار كذا من X_2

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)	π_1	π_2
Love at last	5	5	0	0	0.9	0
Romance forever	5	?	?	0	1.0	0.01
Cute puppies of love	?	4	0	?	0.99	0
Nonstop car chase	0	0	5	4	0.1	1.0
Swords vs. karate	0	0	5	?	0	0.9

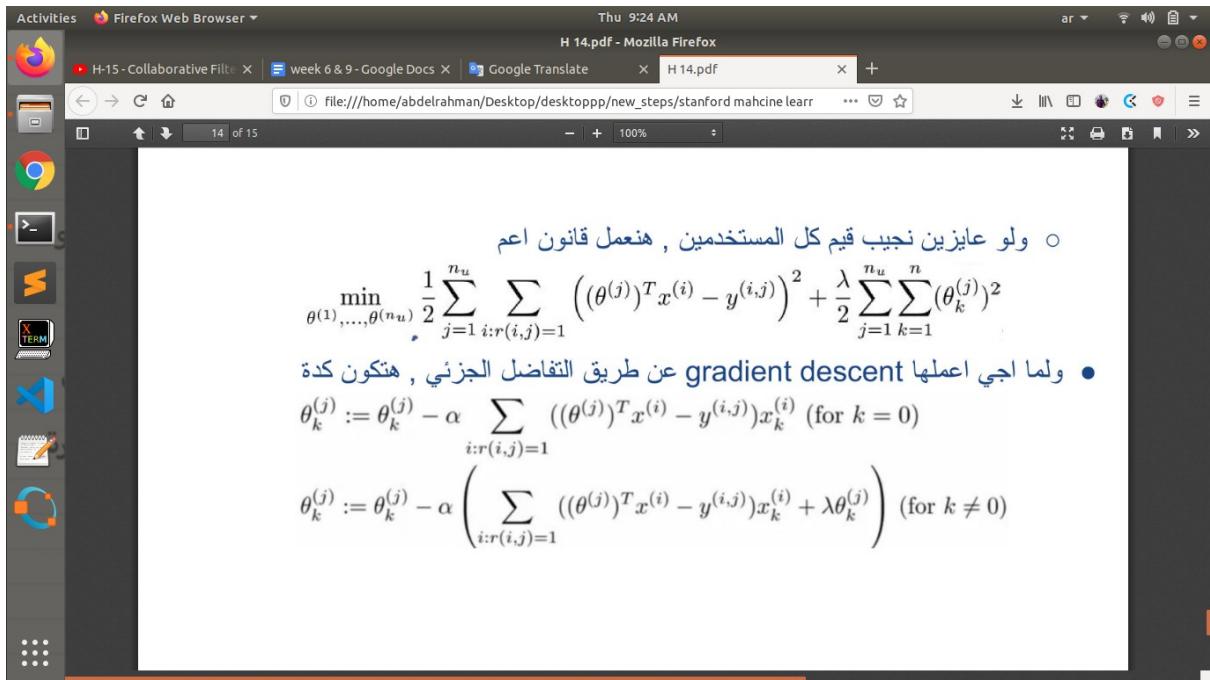
- نبدأ جمع الـ **features** الخاصة بكل فيلم ، هنقول إن X^1 (يعني فيتشرز الفيلم الأول) هتكون

$$\begin{bmatrix} - \\ 0.9 \\ 0 \end{bmatrix}$$

وهنا ده بيعتمد على ال linear regression algorithm الخاص بال

- حيث : $\theta^{(j)}$ هي معاملات كل مستخدم
 - $x^{(i)}$ قيم الـ features للفيلم
 - $r^{(i,j)}$ بتكون بوحدة لو المستخدم عمل تصويت الفيلم ، وصفر لو معملش قيمة التصويت
 - العداد تحت السيجما من رقم 1 ، و مرورا بـ m بالارقام اللي تم التصويت ليها عشان كدة كتب $r^{(i,j)}$
 - لاحظ ان تم حذف قيمة m اللي كانت موجودة ، لأنها ثابت في الطرفين ، واحنا عايزيين نقلل القيمة فاختفائها مش هيعمل مشكلة
 - متنساش اننا اضفنا في الآخر قيمة الـ regulation عشان الـ OF

معنى اي x لو قلنا الفيلم الخامس الى هو مثلا $5x$ هنا الفيلم نفسه بالنسبة لمستخدمي النظام كل واحد منهم ليه theta_i معينة خاصة به هو فمعنى كده ان كل i of x بيكون theta_i بالنسبة لكل مستخدم على النظام.



Collaborative Filters

طيب فيه بعض الحاجات مبيكونش فيها انك تشوف الحاجه او المحتوى مثل ازى موقع البيع والشراء بمعنى انى ممكن اخش على المنتج لكن مقدرش قفيس تغيرات ال user زى الفيلم ده مثلًا شافه كله ولا عمال يقدم وهو بيتفرج وجاب نهاية الفيلم علطول فققر هنا اخد value من كل الحاجات ديه لكن حاجات زى ال online stores وكمان بيتدى يظهر مفهوم تانى اسمه collaborative filters.

ال collaborative filters ى طريقة كانها عامله زى الداير بتروح وترجع لنفس المكان بمعنى انها بتغير قيم θ بناء على قيم ال θ .

<https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26>