

Abdelrahman Rezk

Web Developer

NLP & ML student

AOU University

# مواضيع هامة

## Handling Error

التعامل مع الأخطاء وهو ازاي قدر اتعامل مع الحاجات الى بتقابلنى او الحاجات اصلا الى بتسبب اخطاء وانا بشتغل زى انحراف ال data .

مهم جدا انك تكون فاهم الشغل ماشى ازاي او ايه الى المفروض تستخدمنه فى الجزئية دى وليه استخدمته ولو حصل حاجه معينه زى ال over fitting تبقا فاهم ده حصل ليه يعني مش مجرد شويه algorithms واکواد وتفقى كده تمام لا المفروض تبقا فاهم الموضوع بشكل أعمق ليه بتستخدم وده وليه ال error ده وليه قيمة ال cost function كبيرة مثلا فى ال test data طب هل زودت الوضع هيكون احسن وليه اخترت اصلا انى ازود features طب لو خلية المعادلة تكون polynomial وليه برضه خلتها polynomial وهكذا .  
فلازم إنتا كمهندس machine learning يكون عندك فن وتقدير ووضع الحاجات فى مكانها المناسب .  
ممكن يكون فيه اكتر من تكتيك 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 ايه المناسب وغير المناسب والى شغال فعلا ومش شغال وازاي ممكن قدر اطور ال algorithm بناء على .

## Evaluating Hypothesis

هي إنى اقيم ال algorithm بناء على مدى دقته يعني هل تمام اكمل شغل بيه ولا هو بيسكب انى يكون عندي over fitting ومن خلال ده بقدر احدد انىحتاج اعمل ايه او اضبط ايه بظبط .

التعامل مع الأخطاء

تقييم الإفتراض

- طلب نيدا بما يسمى بـ Evaluating a Hypothesis او تقييم الفرضية

- المقصود بيه ازاي او لا أعمل تقييم للخوارزم اللي تم عندي عشان اعرف مدى جودته ، وقد ايه هو كويس ولا لا ، وهل فيه OF او UF ولا
- فضلًا في المسألة دي ، عدنا عدد كبير من الـ features ، فده غالباً بيسكب الـ OF والجراف لازق في النقط

شكل مستفز ، فيدخلني نسبة خطأ توقع النقاط الجائحة أقل كثير

Fails to generalize to new examples not in training set.

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

$x_1 = \text{size of house}$   
 $x_2 = \text{no. of bedrooms}$   
 $x_3 = \text{no. of floors}$   
 $x_4 = \text{age of house}$   
 $x_5 = \text{average income in neighborhood}$   
 $x_6 = \text{kitchen size}$   
 $\vdots$   
 $x^{100}$

وده بيكون من خلال انى بعد ما عملت **data training** لـ **data testing** بقى اشوف الـ **cost function** وساعتها هقدر احدد الـ **algorithm** بتناعي كويس ولا لا من خلال الـ **cost function** وكمان من قبل كده لو انا عملت لـ **visualize** ممكن اعرف ساعتها لو كان الـ **algorithm** **over fitting** ولا لا زى الصوره الى فوق ساعتها هاكون عارف قبل ما اعمله **testing** انى ها يكون فيه نسبة خطأ كبيرة.

من المفترض برضه إنك تكون عارف وبصيت على الداتا الى انتا بتعامل معها قبل ما تيجي تتصفح الـ **data** الى عندك لـ **training and testing** وعلى حسب النسبة الى هتقسم بيها ممكن تكون 80 لـ 20 او 70 لـ 30 ولكن المهم في كده انى الداتا نفسها متكونش مترتبة من الاصغر للابiger او العكس يعني عموماً الـ **data** كون متغطيه في بعضها او حاصلها **shuffle** والا هيكون انتا بتدرب على حاجه تانيه خالص اصلاً.

فيه برضه مشكله تانيه بتبدأ تظهر وهي الـ **hyper parameter** وهو القييم الى انا بحتاج احددها زى **lambda** وقيم **alpha** الأولى وقيمة الـ **theta** طيب لو قررت انى اعمل **polynomial equation** ها عمل لـ **d**=**kam**. وهذا مثلاً ممكن اعمل لـ **d=10** مثلاً ومع كل درجة احسب قيمة الـ **theta** كام وساعتها اشوف قلهم خطاء وابتدى اشتغل بيها ولكن اخلى بالى انى ممكن مع قيمة الأخطاء القليلة جداً يكون فيه **overfitting** وهذا ممكن مع كل **d** اعمل **graph** لشكل الـ **data**.

## Cross-Validation-Data

هي طريقة يستخدم انى بدل ما كنت بقييم الـ **dataset** الى عندي لمجرد فقط لا هنا انا بقسمها لثلاث اقسام **training, cross and testing** وده يساعدنى في انى بدل ما اوصل لـ **testing** واكتشف انى قيمة الـ **cost function** كبيره عشان مثلاً الـ **algorithm** كان بيعمل **over fitting** لا بلحق نفسى وغير الحاجات الى عايز غيرها فى الـ **cross validation** عشان بعدها اكون خلاص ظبط دنيتى والـ **algorithm** يشتغل على الـ **testing data** وانا مطمئن.

وال فكرة الأساسية من الـ **cross-validation** هو انه بيعالج مشاكل الـ **hyper parameter** الى انا بحتاج احددها. وتحديد النسبة بتاعت كل جزء من الـ **raining, cross and testing** مختلف بإختلف كمية الـ **data** نفسها.

Mon 10:07 AM

PowerPoint Presentation - H 02.pdf - Mozilla Firefox

Google Translate | H-03 - Variance & Bias | week 9 - Google Docs | PowerPoint Presentation - | +

Activities

تقسيم البيانات

- استخدام بيانات الضبط
  - المقصود ببها ان هنقسم البيانات , بدل ما كانت 70% تدريب , و30% اختبار , ه تكون 60% تدريب , ضبط 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

Mon 10:09 AM

PowerPoint Presentation - H 02.pdf - Mozilla Firefox

Google Translate | H-03 - Variance & Bias | week 9 - Google Docs | PowerPoint Presentation - | +

Activities

تقسيم البيانات

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

## Bais & Variance

الانحراف والتتواء

الـ **bais** هو ببساطة بيكون مع مشكله ال **underfitting** وهو انى فيه ابعد او انحراف كبير جدا للنقط عن ال **best fit line** لى انا عملته فتلقيه بيظهر كتير لما بتكون شغال بمعادلات **linear**.  
 لكن الـ **variance** بيظهر مع مشكله ال **overfitting** وبيكون ساعتها **high variance** وده بيظهر لما اجي استخدم معادلات من الدرجة العليا وهو انى ساعتها تكون عملت **fitting** لـ **training data** بشكل كويوس وده بيكون بسبب انى الـ **algorithm** مرن جدا وبيعدى على كل النقط لكن مع ال **cv data** او **testing data** بتظهر المشاكل.  
 فلما بتكون عندي ال **cost function** قليلة جدا فى ال **training data** ده بيكون **overfitting** وفى نفس الوقت لان بيكون ساعتها ال **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 بالتالى قيمه ال cost function تكون كبيره جدا فيحصل ما يسمى ال high bias الى هو under fitting.

**الإنحراف والتوزع**

- طيب لو تم اختيار الدرجة المناسبة لكن تظل الحسابات فيها مشاكل , فممكن نشك في موضوع الانحياز Bias او التوزع Variance

التوزع الزائد يجعل UF لان الخوارزم يتغاهل كتير من البيانات , او يجعل لها اهتمام قليل شوية , و ده بؤدي لـ UF لان الخوارزم مش هيكون يمثل الغلب للبيانات , وده غالبا بييجي من الدرجات الاقل للمعادلة (المعادلة الخطية مثلا) , او صفر حجم الثبات

التوزع الزائد معناه حساسية عالية لاي تغير في البيانات , فالخوارزم يحال بجمع كل البيانات مع بعض في جراف واحد , فيعمل OF , وده غالبا بييجي من الدرجات العالية للمعادلة , او زيادة قيمة الثبات

**الإنحراف والتوزع**

- و عشان نفهم الكلام ده عمليا , تعالى نقارن بين قيمتين

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

قيمة ل الخاصية بيانات التدريب

$$J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

قيمة ل الخاصية بيانات الضبط

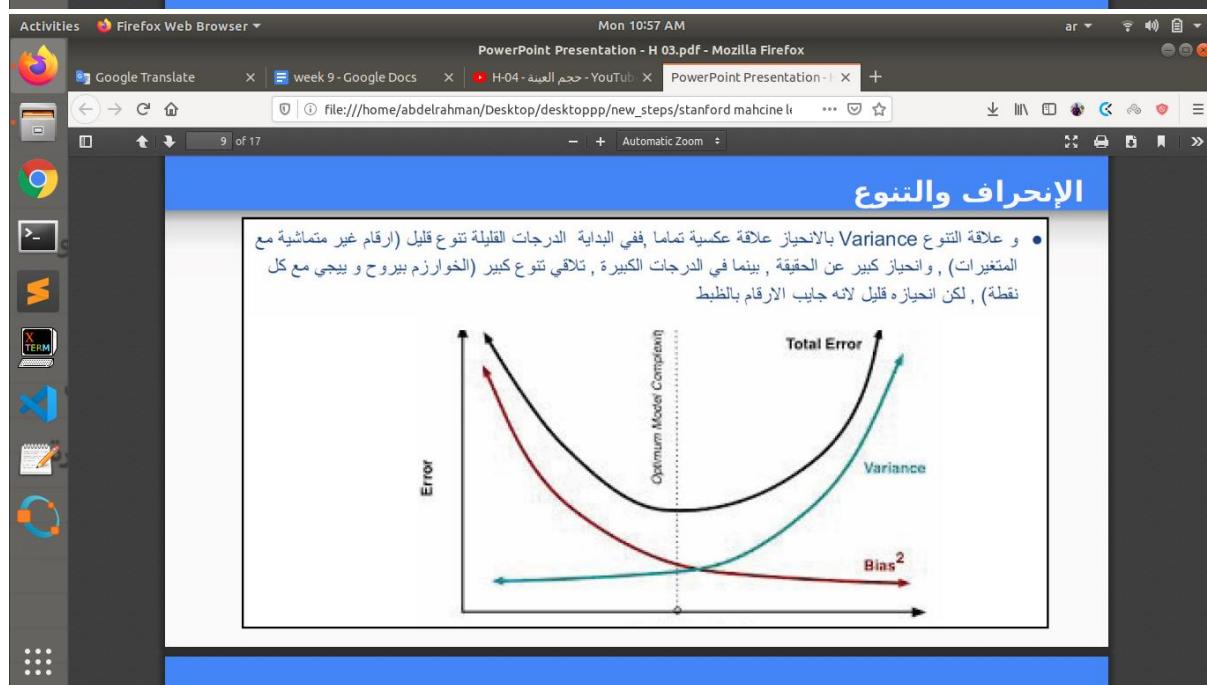
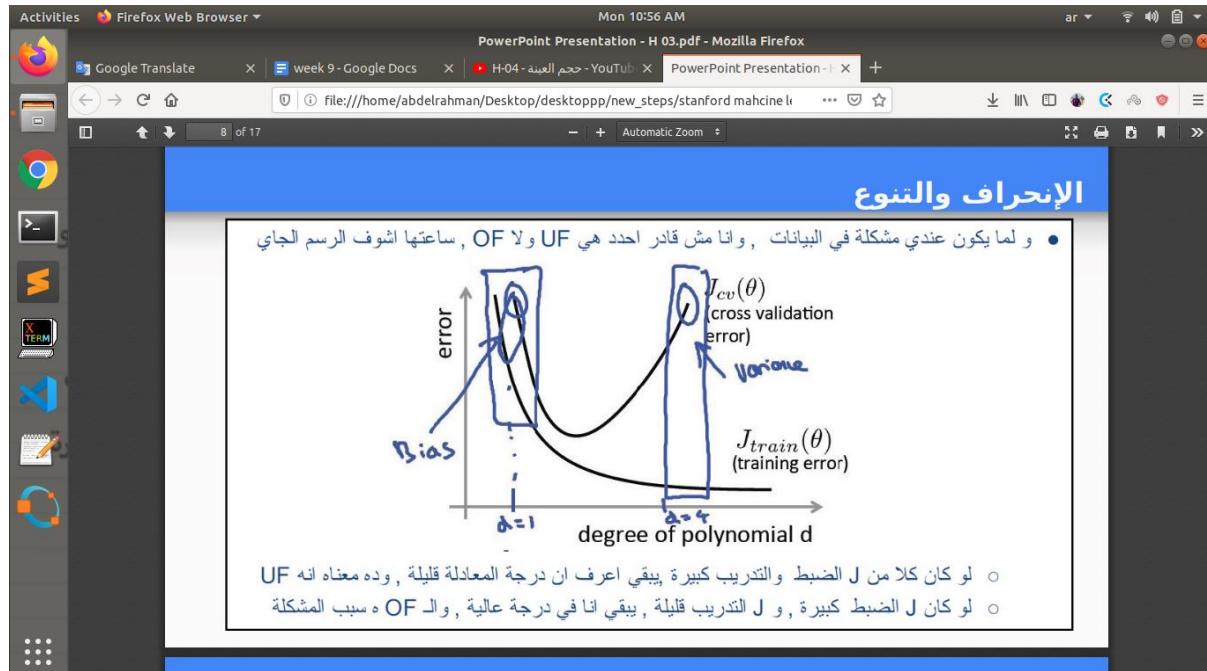
هنالاحظ في الرسم ده حاجة مهمة :

- ان اولا محور اكس 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 ال algorithm بيحاول انه يوصل لقيم الثبات المناسبة جدا بضربيها فى الاكسات فلنقى ده بيظهر اكتر لما نيجي نزود ال degrees بتاعت المعادلة .

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 algorithm عمل قوى جدا مع ال training data لكن مع بيانات جديدة بيحصل انى الخطأ بيكون كبير جدا .

ولكن لما تكون قيمة الـ cost function بالنسبة لل training and cv data كبيرة وكمان متقاربين مع بعض فده يكون under fitting ويتلاقيه لما نيجي نستخدم مثلاً معادلة من الدرجة الاولى وال algorithm بداعي مش عامل .height bais under fitting data أو ما يسمى under fitting.



ويرضه قيمة ال lambda نفسها بتفرق في ال under fitting over fitting لانها في الآخر هتضرب في ال theta لى انا اصلاً بحاول قللها عشان تقلل ال cost function فيحصل انى في الآخر ممكن توصل لل 0 وده بيسبب over fitting لانى مبنوش استخدمت ال regularization والعكس لو كانت كبيرة فلما تضرب في قيم ال theta لى انا بحاول قللها هاتطلع قيم ال theta قليله جداً وبالتالي ده هايسبب انها تكبر ال cost function ويحصل under fitting.

Mon 11:02 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

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

12 of 17

Automatic Zoom

## الإنحراف والتباين

و عشان ندرس تأثيرها على كل من الانحراف والتباين ، عايزين نعرف هي المدما بتعمل ليه أساساً

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

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 - | +

13 of 17

Automatic Zoom

## الإنحراف والتباين

فلو كانت اللمنا كبيرة جدا ، الخوارزم هيفسر يقل جداً قيمة كل الثباتات عدا ثباتاً صفر اللي مش مضروبة فيها ، وده هيخلي قيمة كل الثباتات تقريباً بصفر ، الا ثباتاً صفر اللي ه تكون بقيمة عادي ، فيكون  $Hx$  يساوي تقريباً ثباتاً صفر ، فع تكون معادلة مش خطية ، لكن ه تكون خط افقي مستقيم يساوي ثباتاً 0 ، زي الشكل الشمال ، وه تكون UF ولو اللمنا قلت جداً ، فده معاناها ان الـ Regularization مش موجود تقريباً ، وبالتالي هنفضل الخوارزم زي ما في الشكل اليمنى

اما لو اللمنا متوسطة ، فمش هتقل الثباتات كثير ، لكن هنطبطها بحيث يتلاشى الـ  $OF$  من غير ما ندخل في الـ UF

Mon 11:02 AM

PowerPoint Presentation - H 03.pdf - Mozilla Firefox

Activities Firefox Web Browser

Google Translate week 9-Google Docs YouTube PowerPoint Presentation -

file:///home/abdelrahman/Desktop/desktoapp/new\_steps/stanford machine learning

الإنحراف والتمنع

- طيب نحدد قيمة لمدا المناسبة زي ؟
- الاول نبدا نحدد قيم لمدا اللي هنجريها ، 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

Activities Firefox Web Browser

Google Translate week 9-Google Docs YouTube PowerPoint Presentation -

file:///home/abdelrahman/Desktop/desktoapp/new\_steps/stanford machine learning

الإنحراف والتمنع

- كدة هيكون عندنا ثيتا 1 اللي جيت 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 مش التدريب ، بس المعادلة المرادي هتون من غير قيمة المدا ، ونشوف اقل  $J$  فيهم هنكون 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 ديه هي الافضل في الاستخدام
- عشان اتأكد ، هجرب نفس الثيتا في بيانات الاختبار ، بنفس المعادلة اللي فوق اللي من غير لمدا
- يعني لما هاجي اطبق بشكل حقيقي مع بيانات العمل ، هاستخدم المدا ديه 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  $J$  ولا  $J'$  ولا  $J''$

وهو انى لما بعمل regularization عمله فقط مع ال training data اروح اجرب ساعتها فى ال .testing

## حجم العينة

مهم جدا انى اكون عارف تاثير حجم العينة يعني ممكن يكون فيه over fitting لل training data لكن لما ازود اكتر القى انى قيمة ال cost function لى كانت قليله جدا بتزيد وده بسبب انى ال algorithm بيكون بيحاول انه يعمل fitting لكل ال data فى graph واحد.

ولكن ده فى ال cv data ده بيكون كوييس جدا لان ساعتها حتى لما تكون قيمة ال  $J$  زادات بسبب انى البيانات كبيره ده بيساعد انه يقدر يعمل prediction كوييس لبيانات مشفهاش قبل كده.

Mon 11:27 AM

PowerPoint Presentation - H 04.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | YouTube - حجم العينة (1) | Google Search - مواعيد الصلاة | PowerPoint Presentation - | +

6 of 13 Automatic Zoom

### حجم العينة $m$

- فلاحت في الرسم ان كل ما تزيد قيمة  $m$  كل ما تزيد نسبة الاخطاء لـ في بيانات التدريب

- لكن من الديهي ان بيانات الضبط CV وضعا مختل ، لأن العدد القليل جدا من بيانات التدريب ، صحيح ان لـ التدريب قليلة (لان الجراف لا يرق في النقط) لكن فعليا لـ الضبط هتني كبيرة جدا ، ليه ؟ لأن اساسا الخوارزم المعول بـ  $m$  قليل ، مستحبيل يقدر يتوقع اي قيم مشاهاش ، 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 مشاهاش

واخلى بالى لما بيكون عندي مشاكل اى هى over fitting ف ساعتها ايجاد  $\text{data}$  جديده واعملها الى هو زيادة نسبة الـ training data cv data بيتسبب انى الـ error فيها نسبة الـ error لكن على عكس polynomial degree زياحة الـ data ش بيح المشكلة ف ساعتها بشوف مثلا حاجه تانية زى الـ under fitting فسواء كده او كده لازم اكون عارف امتا استخدم ايه او المشكله ديه بتطلب انى اعمل ايه بطبعه.

Mon 11:36 AM

PowerPoint Presentation - H 04.pdf - Mozilla Firefox

Google Translate | week 9 - Google Docs | YouTube - حجم العينة (1) | Google Search - مواعيد الصلاة | PowerPoint Presentation - | +

13 of 13 Automatic Zoom

### حجم العينة $m$

- طيب بعد ما اتعلمنا عدد من التكتيكات لعلاج المشاكل ، عايزين نعرف ايه بيستخدم لاي :
- زيادة العينة شوية ونجيب صافوف اكتر :
- يستخدم بس في حالة الـ OF و ان عندي تنوع كبير ، لكن لو التنوع قليل او الانحراف كبير ، وعندى UF فزيادة العينة يعني تصبيح وقت
- تقليل عدد الـ features :
- برضه في نفس حالة الـ OF
- زيادة عدد الـ features :
- غالبا يتحل مشكلة الـ UF والانحراف الكبير ، لأن زيادة عدد الـ features بيخلي الخوارزم عنده القدرة انه يتبع النقاط بشكل افضل
- زيادة عدد و اس الـ polynomials :
- وده برضه هيجل مشكلة الـ UF والانحراف الكبير ، لأن كل ما يزيد اس المعادلة ، كل ما يكون ليها القدرة انها تتقارب اكتر و تتجنب الـ UF
- زيادة قيمة لمدا :
- تحل مشكلة الـ OF
- تقليل قيمة لمدا :
- تحل مشكلة الـ UF

## Skewed Classes

يقصد بها إنى مش معنى انى الـ algorithm بتاعى بيقول انى الـ accuracy عندي يعني نسبة الصح كبيرة جدا فده شيء كوييس لأن فى بعض الأحيان يحتاج انى انحراف او اروح ناحية حاجة معينه او قيمة معينه اكتر من الثانية فمثلا لو بنتكلم انى الـ algorithm بتاعى شغال انى يعمل predict للشخص مريض ولا مش مريض ساعتها ممكن حاجة تزيد

على حساب حاجة تانية وهو انى حتى لو توقعت انى الشخص ده مريض هيبقى احسن بكثير جدا من انى ققول انه مش مريض و ساعتها ممكن اصلا تكون حاله حرجه ويحصل انه يموت 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 و كنت بقول انه اكبير من او يساوى 5. هاتوقع حاجه معينه انه مريض مثلا ولو اقل هاقول لا ده مش مريض هنا فى المثال ده مينفعش امشى كده فاحتاج انى أعلى قيمة ال threshold وهى العتبة او الحاجة الى من عندها ببتدى اخد قرار كأنها Decision boundary فممكن هنا بقا لو عليت شويه تكون مثلا 7 يعني فيما فوق ال 7 هيكونوا مش مرضى وغير كده هيكونوا مرضى فهنا الى انتا بتصنيفهم انهم مرضى في الاخر اكبير من او يساوى 7. وده طبعا مش صح لأنك فى حاله زى كده تحتاج تعكس الموضوع.

فهنا بيظهر عندي مفهوم جديد ليه علاقة بحساب الدقة لأن مبنكونش بطريقه انى اجيبي الصح على التوتال كلها average or accuracy لا بيكون فيه طريقه ثانية وهي Matrix بقدر اعرف من خلالها الحاجة الى انا شغال عليها فلو هنا مثلا انا عايز احسب الناس المرضى فعليا فهنا عشان اشتغل خلينا نقول مثلا هو انى عندي 1000 واحد مشتبه فيه انه مريض سرطان لكن فعليا في 1000 واحد المشتبه فيهم دول فيه 5 اشخاص مرضى سرطان فعليا وجه ال algorithm بتابعى توقع منهم 4 مرضى سرطان ولكن فى الحقيقى فيه 2 منهم فقط هما الى مرضى وانتين تانين اصحابه فهنا بقا ببتدى احسب ال matrix ديه عن طريق اربع حاجات الى هما

#### True Positive , True Negative, False Positive and False Negative.

طيب هاتحسب على اساس ايه ديه بختلف على أساس القيمة الى انا عايز اتوقعها يعني انا هنا عايز اتوقع الاشخاص اللي هما مرضى سرطان فببتدى احسب اول حاجه ال الاشخاص الى انا توقعت انهم مرضى سرطان وفعليا هما مرضى سرطان وديه بتكون ال True Positive وفي المثال الى فوق هنلاقى انى آل algorithm توقع ان فيه 4 مرضى سرطان بس فعلا فيه اتنين منهم فقط هما الى مرضى فهنا في الحاله ديه هيكون 2 فقط ال algorithm توقع انهم مرضى سرطان وهما فعلا مرضى سرطان يعني .2 = positive

تاني حاجه ال True Negative و هي الاشخاص الى انا توقعت انهم مش مرضى وهما 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 وهما فعليا مش مرضى لكن فيه خطأ وهو انى اصلا عارف انى فيه فعليا 5 مرضى سرطان فلما توقعت فوق انى فيه 2 وطلعوا فعليا مرضى فهنا باقى من 5 تلاته فاكيد ال algorithm توقع 3 غلط انى هما مش مرضى بس فعليا هما مرضى فتكون النتيجه انى آل True Negative هي 993.

ثالث حاجه هي ال False Positive و ديه تعتمد برضه على الى انا بحاول اتوقع انا بحاول هنا اتوقع المرضى فال False Positive بتشوف القيم الى انا توقعتها صح لكن في حقيقة الأمر هي مش صح فلو لاحظت فوق انى توقعت انى فيه 4 مرضى لكن في حقيقة الأمر هما كانوا اتنين مرضى وانتين 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.

اخرا حاجه وهي ال False Negative و هي انى 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.

نيجي بعد كده مفهومين مهمين الا و هما ال recall وال precision .  
ال precision ببساطة هي لو في مثال انى بحاول اتوقع المرضى فهى بتكون النسبة بين المتوقع الصحيح على مجموع كل المتوقع.

#### TP/ TP+FP

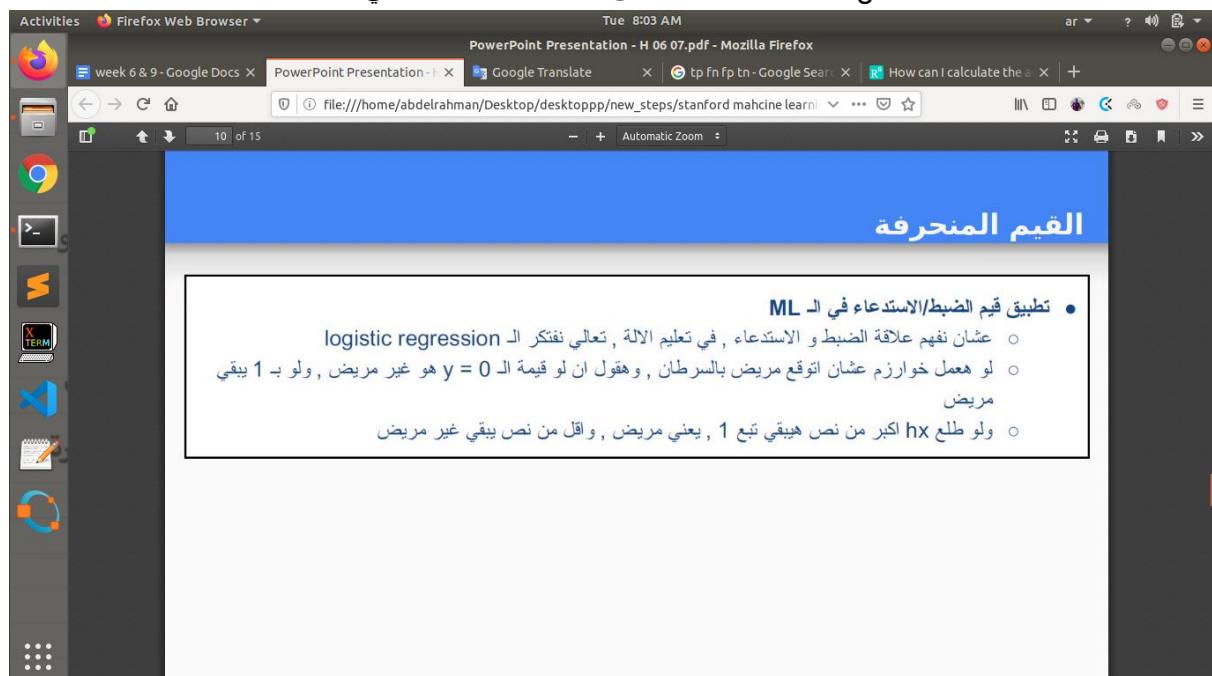
فه تكون على المثال اللي فات انا عندي TP الى هى الاشخاص الى انا توقعهم مرضى وهما فعليا مرضى 2 والاشخاص الى انا توقعهم مرضى بس هما مش مرضى برضه 2 فه تكون 4/2 وبكده بيططلع عندي 50% من الى انا قلت عليهم

عند هم سرطان هما فعلا عند هم سرطان فتلاقي انى آل algorithm بتاعي ضعيف جدا لأن ده معناه نسبة قد ايه من ال algorithm بتاعي قال انه صح او مريض طلع فعلا انه مريض فكتئه هنا 50% فقط.

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

## TP/TP+FN

الى هو ال algorithm توقع قد ايه مريض وهم فعلا مريض على نفس الكلام كام واحد قال انه مريض وهو فعلا مريض + كام واحد مش مريض وهو قال انه مريض فتلاقي انها ½ ودهطبعاً رقم قليل جداً جداً حوالي 40%. ومعناه منطقياً يا ترى ال algorithm اكتشف اد ايه من الى فيه الحاجة المطلوبه الي هو هنا مريض سرطان.



ساعات كتيره على حسب انا شغال على ايه بحاج زى ما كنت وضحت فوق انى ممكن تيجى تقول لا والله انا محتاج فعلا مقولش على حد انه مريض الا لما يكون فعلا مريض فهنا نسبة ال threshold بتتعلى فمثلاً لو كنت بتتوقع انى اكبر من او يساوى 5. ده كده هايكون مريض لا هاخلى النسبة متلا 7 او اكبر فهنا فعليا انا كده كبيرة نسبة ال precision الى هو بقا لما بتتوقع انى حد مريض فهو فعلياً مريض لكن مثلاً لو شاكك في بعض الناس زى مثلاً النسبة بتاعته 6. فهنا هاقول انه مش مريض فانا هنا برفع نسبة ال precision وده لو لاحظت هايزيود قيمى ال 0 الى هي الناس الى مش مرضي ويفاقل قيمة المرضي ولكن في 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 ايه وهذا هلاحظ انى قيمه الظبط هاتزيد لأن هاقلل عدد الاصحاء وهاقول انهم مريضي على عكس ال recall هاتقل لأن ممكن فعلياً يكون فيه مريض وانا مش هاقولهم زى النسبة ال 20% الى ان رفعتها من 5.

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 مقولاش لاي حد صحيح انه مريض , ومفيش مشكلة لو فيه مريض مقولاوش انه مش مريض , يعني بالعربي مش هقول احد انه مريض الا لو كنت متأكد جدا ساعتها هرفع قيمة الـ **threshold** اللي هي العتبة المحددة , بدل ما تكون نص هنكون 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 وبقلل قيم 1 شوية , صحيح انه هينفذ مجموعة من الاصحاء من اني اخضمهم , لكن هيعرض حياة بعض المرضى الحقيقيين للخطر اني مش هقولهم
- دلوقتي لأن قيمة الضبط معناها (نسبة قد ايه من اللي الخوارزم قال انه صحيحة . طبع فعلا صحيحة) فكدة قيمة الضبط هتزيد شوية لاني هقلل عدد الاصحاء اللي هقولهم انتو مريضي
- 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
- وطبعا لو خلية العتبة 0.8 يبقى قيمة الضبط هتزيد اكثر و الاستدعاء هتقل اكتر
- فزيادة قيمة العتبة بيعلي الضبط و يقل الاستدعاء

الحالة الثانية: (العكس)

- بفرض ان لسيب تاني , مهم جدا اني مفوش اي حد مريض , يعني اي حد شاكك فيه لازم اقوله انه مريض , ومفيش مشكلة لو فيه صحيح فلتله انت مريض مش ازمة 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 هتكون 0.3 , وده هي عمل حاجة مهمة
- ان بعض اللي كنت شاكك انهم اصحاب (4 مثلا) هقولهم انهم مرضى , وده هيزود قيم الـ 1 و يقلل قيم 0 شوية , و ده هيمنق مجموعة من المرضى اتهي هعرفهم المرض , لكن هيخضن مجموعة اكبر من الاصحاء اللي شاكك في مرضهم و هقولهم و خالص
- دلوقتي لان قيمة الضبط معناها (نسبة قد ايه من اللي الخوارزم قال انه صح . طبع فعلا صح) ف kedda قيمة الضبط هتقى شوية لاني هزود عدد الاصحاء اللي هقولهم انتو مرضى
- 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 هيعرفو وطبعا لو خلية العتبة 0.2 بيبقى قيمة الضبط هتقى اكتر و الاستدعاء هزيد اكتر
- فتقلى قيمة العتبة بيكيل الضبط و وزود الاستدعاء

القيم المنحرفة

- محور اكس الاستدعاء , محور واي هو الضبط
- لما كانت العتبة كبيرة (فوق على الشمال) كان الضبط كبير و الاستدعاء قليل
- العكس , لما كانت العتبة قليلة (تحت على اليمين) الضبط قلل و الاستدعاء كبر
- ولو هخلي العتبة في النص , كلام الضبط والاستدعاء هيفضلو متواسطين
- وبينهم ممكن الجراف بيقى خطى او كيرفى فوق او تحت حسب اتجاه القيم

القيم المترفة

- نفس الموضوع ، لو العتبة زادت ، هنقول قيمة الاستدعاء ويزيد الضبط ، ويطلع متوسط جميل لكن الكفاءة وحشة
- فالحل استخدام قيمة ما يسمى F1 Score اللي معادلتها كدة :
$$F = \frac{2PR}{P+R}$$
- هنا هنلاقي ان القيمة النهائية ، معتمدة على قيم الضبط والاستدعاء مع بعض ، ولو واحدة فيهم قليلة جدا ، الإف هنقول برضه وممكن توصل لصفر
- و القيمة المثلية ، ان الخوارزم يكتشف كل المرضى ، ويستبعد كل الأصحاء ، و ساعتها كلام من الضبط والاستدعاء هيكونو بـ 1 ، وقتها هيكون F بواحد
- ولو ايا من الاستدعاء او الضبط بصفر ، الـ F هيكون صفر فورا
- قيمة الـ F تترواح بين الصفر والواحد ، وهو افضل ممثّل عن كفاءة الخوارزم

## تجهيز البيانات

احتياجي لل data بيتختلف من موقف للثاني بمعنى انى زياده البيانات ممكن يكون مفيد وساعات لا يعني لما اكون بتعامل مع الموقف الى فيها predict للكلمة القادمة فى text او voice recognition وغير من الامثلة الشبيهة هنا زيادة البيانات بيكون مفيد جدا انه ال algorithm يشوف أمثلة مختلفة وكثيرة لكن فى حاجه زى توقع سعر بيت زيادة البيانات لو كان ال features لى عندي قليله جدا ساعتها زيادة ديه مش هتفيد فهنا فى نموذج بشري بيتعمل الى هو لو حاجة الإنسان البشري يقدر يستنتجها ف تمام غير كده بيقا زيادة البيانات مش صح.

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

وده بيقود انى هنا فيه اختلاف يعني امتى تكون زيادة البيانات شيء كويسيس وامتنانا لا والمثال ده بيوضح.

Tue 8:48 AM

PowerPoint Presentation - H 08.pdf - Mozilla Firefox

week 6 & 9 - Google Docs | Google Translate | H-09 - Anomaly Detection | PowerPoint Presentation - | +

8 of 11 Automatic Zoom

## تجهيز البيانات

- الحالة الأولى
- لو المشكلة عندي اني هلا فراغ في كلمات مثل زي : **In the breakfast I ate ----- eggs**
- وكانت الاختيارات اني حط واحدة من الكلمات ديه : **Too - two - to**
- فهنا لو ان الخوارزم شاف كمية كبيرة من البيانات زي ديه , هيقدر 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 هيغيب حاجة الا ضياع الوقت فيها

Tue 8:50 AM

PowerPoint Presentation - H 08.pdf - Mozilla Firefox

week 6 & 9 - Google Docs | Google Translate | H-09 - Anomaly Detection | PowerPoint Presentation - | +

10 of 11 Automatic Zoom

## تجهيز البيانات

- طب نفرق ازاي بينهم , ونعرف امتي زيادة البيانات مفيدة ولا لا :
- في الحالة الأولى لو عرضناها لبشرى متخصص (مدرس انجليزي) وقلاله يا ترى هنحط كلمه ايه , هيعرف يجاوبها فورا , وهذا زيادة البيانات **هتفيد**
- في الحالة الثانية لو عرضناها لبشرى متخصص (مسار) وقلاله سعر البيت الـ 200 متر , مش هيعرف يجاوب طبعا لانه محتاج يعرف بيانات تاني و **هذا زيادة البيانات مش هتفيد**

- و نختم بحاجة مهمة :
- البيانات الكبيرة , بتكون مفيدة في حالة عندي **features** كثيرة , لأن **Variance** البيانات الكبيرة بتقلل التنوع
- والـ **bias** الكثيرة بتقلل الانحراف
- وبالحاله المثلية , ان يكون كلام من التنوع والانحراف قليلين مع بعض , فاتجنب الـ **OF & UF** مع بعض
- بينما لو واحدة كثيرة وواحدة قليلة هعم في مشكلة **OF او UF**

# Anomaly Detection

القيم الشاذة هى بعض القيم التي تخرج بره إطار معظم ال **data** الى عندي وديه لازم اخذ بالى منها لانها ساعات بتؤدى لأضرار او 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 ايه بظبط.

يصعب جداً إثبات صحة القيم الشاذة في التعلم العملي (training) ولكنها تتطلب شكل واضح مع الـ **testing** أو البيانات إلى الـ **model** مشفههاً قبل ذلك، وكمان لو كان فيه بعض القيم الشاذة أثناء التعلم (training) فممكن أنها تتفسح لأنها ممكن تؤدي لـ **misleading** في الـ **.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

هي الطريقة التي يقدر من خلالها أعرف القيم الشاذة وهي عبارة عن طريقة يتم حساب فيها  $\mu$  (mean) و  $\sigma$  (standard deviation)، على هذه عشان قدر احدد الـ **range** بتابعك أو الـ **epsilon** إلى لو أقيمت قيمة خارجة عنه ه تكون بالنسبة لها **anomaly points** واخلى بالى أنى لازم أكون عارف انهى **features** بختارها مش اى حاجة وخلاص لأن لو اختارت وخلاص ده بيقود غالباً لـ **misleading**.



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/desktopopp/new\_steps/stanford machine learning/

7 of 13 Automatic Zoom

## القيم الشادة والـ GD

### (2) حساب قيمة ميو $\mu$ :

يعني لو عندي الف عميل , وكل عميل عنده 50 معلومة , فهجيب متوسط كل معلومة فيهم , عن طريق اني اجمع قيمة المعلومة الاولى (الطول مثلا) لالاف عميل , واقسمها على الاف , بعدها اجمع قيمة المعلومة الثانية (الوزن) و اقسمها على الاف 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 الصحفوف , واقسمها على عدد الصحفوف  $m$

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/desktopopp/new\_steps/stanford machine learning/

9 of 13 Automatic Zoom

## القيم الشادة والـ GD

### (4) إيجاد قيمة الاحتمالية :

يعني عندي عميل بنك جديد , فيه اجيب الخمسين 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 المولارية للالف عميل سابق

اكرر خطوة الاحتمالية لباقي الـ features (كل الاعدمة)  $\prod$

اضرب كل الاحتماليات في بعض عن طريق الـ  $\prod$

اخيرا , لو لقيت ان قيمة  $P$  النهائية (الاحتمالية) اقل من ايسلون  $\epsilon$  اللي تم تحديدها , بقى العنصر ده في مشكلة و ابدا اخذ بالى منه

ال  $\sigma$  بتكون هي ال range الى بتقع فيه النقط بتاعتي وده بيختلف مع كل feature.

Activities Firefox Web Browser ▾ Wed 1:23 PM

PowerPoint Presentation - H 10.pdf - Mozilla Firefox

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

week 6 & 9 - Google Docs | Google Translate | PowerPoint Presentation - | +

file:///home/abdelrahman/Desktop/desktoppp/new\_steps/stanford machine learning/ ... Automatic Zoom

13 of 13

## القيم الشاذة 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 اكتر من ٤
- 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 ٤
- معنى كده ان اي نقطة ه تكون تحت سطح الجرس المجمم ، تبقى مقبولة ، اي نقطة فوق سطحه ، تبقى شاذة

ازاي قدر الاحظ احدد او الكشف عن البيانات الشاذة الى عندي وايه اصلا الى يخليني قولو انى ديه قيم شاذة.  
لما يكون عندي القيم الشاذة مثلاً هي عدد قليل جداً من النقط بالنسبة للـ **data** ثلا 20 نقطة من وسط آلاف فساعتها يعتبر فعلاً ان ديه قيم شاذة وابتدى اشوفها على حده لكن لو كان مثلاً 20% من البيانات قيم شاذة فهنا أباً مش بتعامل معها كأنها قيم شاذة ولا يحتاج ان اكتشفها لأنها بتكون جزء اصلاً من الـ **data** بnature و بتعامل معها عادي زي اى اسيب الـ **algorithm** هو الى يحددها لكن في حالة انى في مثلاً 20 نقطة من وسط آلاف النقاط فهنا لازم اخل بالى وامشي بطريقة من الطرق وهي:

انى لو كان عندي مثلاً 20 نقطة شاذة من وسط 10000 نقطة فهنا هبتدى قسم الـ **data** بnature عن طريق الـ **cross validation** مثلاً 60% و 20% و 20% الى 60% **training** دول هاخدهم من الداتا الى **testing** اى شواذ يعني هاتكون الداتا العادية جداً الى مفهاش اى حاجه وها قسم الـ **cv** 20% و 20% **validation** هنا مثلاً 2010 و 2010 للـ **testing**.and **validation**

The screenshot shows a Firefox browser window with the address bar pointing to a local file: file:///home/abdelrahman/Desktop/desktoppp/new\_steps/stanford machine learning/. The main content area displays a slide from a PowerPoint presentation titled "PowerPoint Presentation - H 11.pdf". The slide contains the following text:

○ 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 (اللي مفيهاش اي شواد) عشان اخلي الخوارزم ، يقدر يحدد قيم ميو و سigma
- ثالثا امسك بيانات الـ CV او test (اللي فيها شواد) و ابدأ اعمل اختبار ليهم ، وهبطل 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 ابسلون ε هي اني اعمل قيم مختلفة لها في المثال اللي فات ، وشووف انها قيمة بتعمل اعلي

F1 Score

اخلى بالى برضه انى فى الاخر القيم الشاذة ديه مبيكونش ليها pattern عين إنتا تقدر تتبعه او تمشى عليه او تحله وتعمله معادله او حاجه لكن بتكون مختلفه اختلاف كلها كل نقطة عن الثانية ممكن لو هنتكلم فى range معين الي هو الطبيعي مثلا من 100 ل 1000 نلاقى قيمة مثلا بتكون في رنج 0 ل 5 وقيم تانية في رنج 1000000 ل 10000000 فهى بتختلف من قيمة للثانية مش كلها اك 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.

Activities Firefox Web Browser ▾

PowerPoint Presentation - H 11.pdf - Mozilla Firefox

Wed 1:40 PM

(2) H-12 - 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 الأحيان ، تكون قيم الافتراضات لدينا جاهزة ، و يتم رسمها ببساطة بشكل جرس جوسين زي كدة

◦ و حتى لو مكانتش الشكل دقيق 100 % ، لكن هو يفاض على جوسين ، وبالتالي مش ها عمل فيه حاجة ، وهجيب الميلو و السيجما بشكل طبيعي

- لكن ممكن اجي ارسمها الاكيها بشكل مختلف زي ده

Activities Firefox Web Browser ▾

PowerPoint Presentation - H 11.pdf - Mozilla Firefox

Wed 1:41 PM

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

file:///home/abdelrahman/Desktop/desktoppp/new\_steps/stanford machine learning/

10 of 10 Automatic Zoom

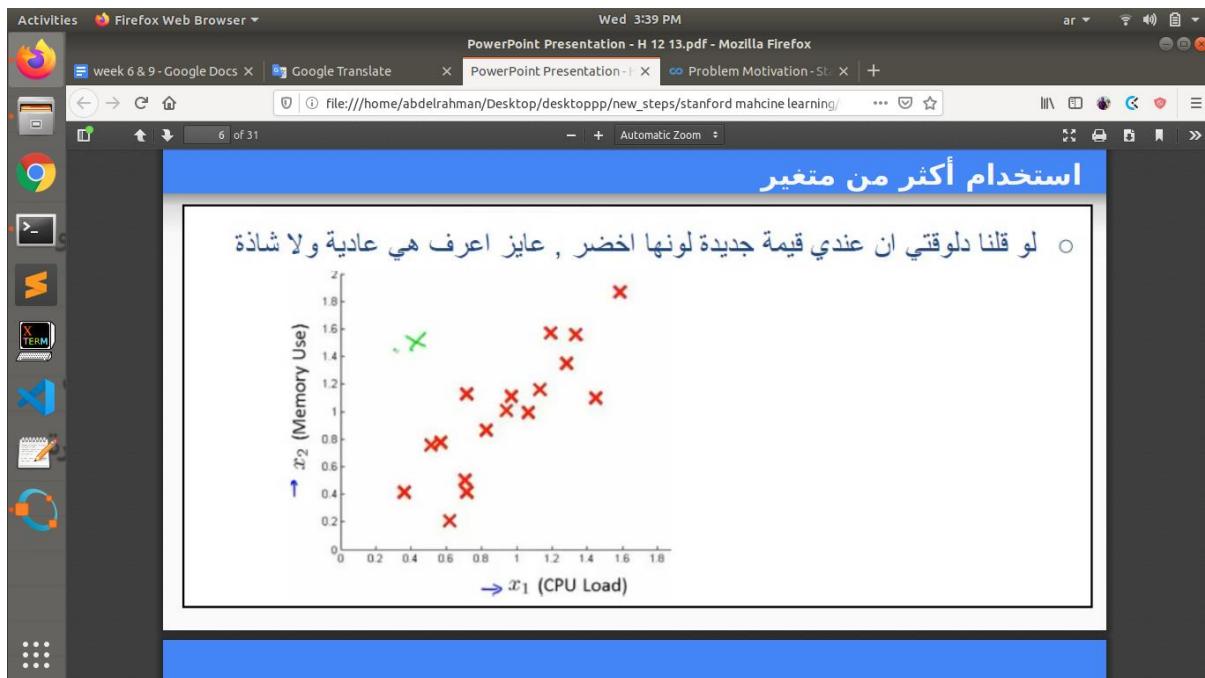
## اكتشاف القيم الشاذة

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

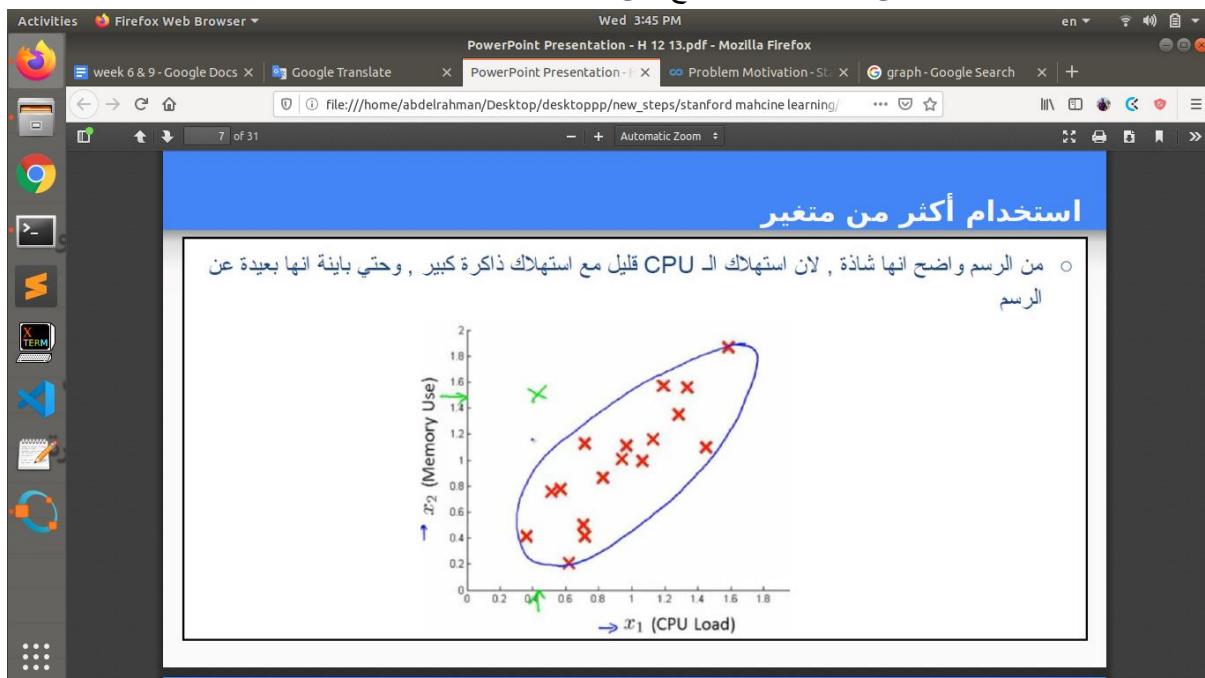
  - $\log x$
  - $\log(x + c) == c$  is constant
  - $\sqrt{x} ==$  whatever the root
  - $x^3 ==$  whatever the power

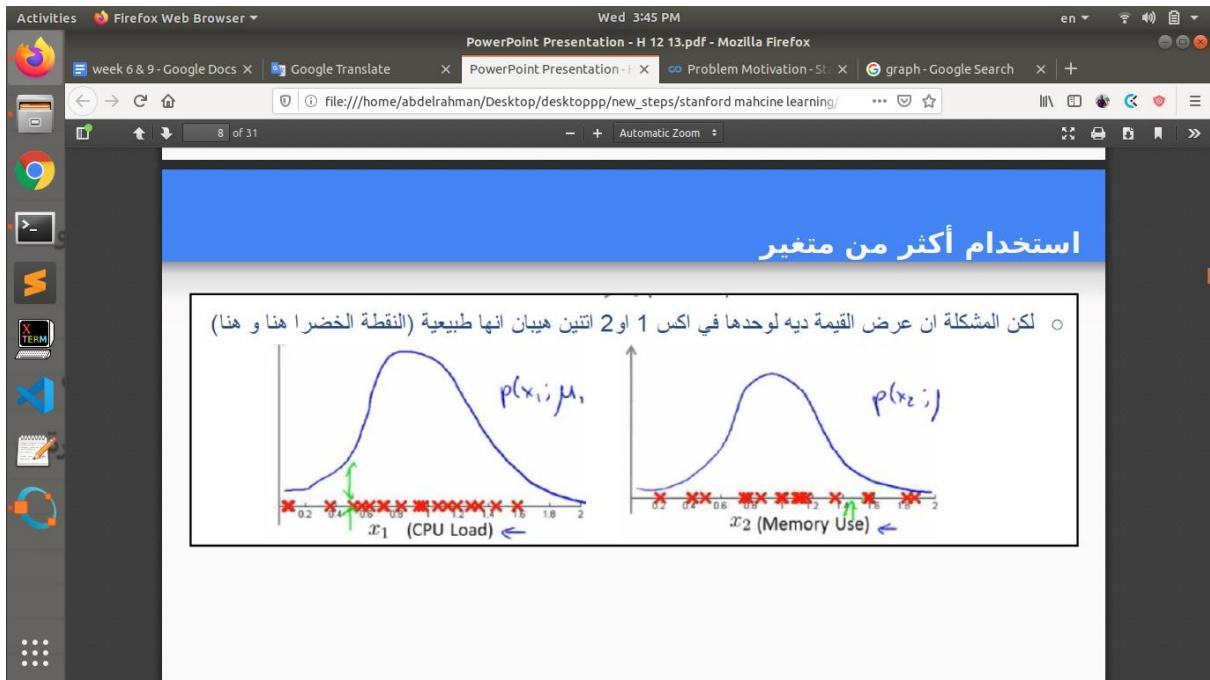
- يعني انا معايا الامكانية بتغيير دالة اكبس باي تعامل رياضي ، عشان اوصل لجرس جوسين
- و طبعاً لما اوصلها ، يبقى اي قيمة شاذة عشان اكتشف عليها ، هعمل لها نفس التحويل

بتقابلينا مشكلة تانية وهي تحديد قيمة الشوادع مع اكتر من تغير لان ساعتها بيكون في اختلاف كون القيمة شاذة ولا لا لأن ممكن يكون كل **feature** لي حدا متقبل القيمة دي بس ارتباط ال **wo features** مع بعض بيخلوي النقطة ديه تبان انها شاذة فهنا ال **correlation** بيوضحلى مدى أهمية اكتر من **feature** لي النقاط الى عندي مش مجرد انى اشوفها بالنسبة لكل قيمة على حدا.



زى بظبط النقطة الخضراء واضح انها جوـه القيم الخاصة بـ $x_1$  لوحدها ونفس الكلام بالنسبة لـ $x_2$  ولكن مجرد النظر إليها بالنسبة لـ graph2 مع بعض بـلـاحـظـانـها بـعـيـدة جداً عن تـكـثـلـالـنـقـطـ حتى تـزاـيدـالـنـقـطـ نفسها بـيزـيدـبـطـرـيقـةـ مـعـيـنةـ لماـ $x_1$  تـزاـيدـالـ $x_2$  بـتـزيـدـهـ وـهـكـذاـ لـكـنـ النـقـطـ الخـضـرـاءـ لـوـ لـاحـظـانـاـ أـنـهـ رـغـمـ اـنـ $x_1$  قـلـيلـةـ جـداـ وـلـكـنـ قـيـمةـ $x_2$  كـبـيرـةـ جـداـ فـهـنـاـ اـرـبـاطـ النـقـطـتـينـ بـعـضـ بـيـخـلـىـ النـقـطـ الشـاذـةـ توـضـحـ عـلـىـ عـكـسـ كـلـ featureـ لـوـحدـةـ.





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

Wednesday 3:53 PM

## 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)

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

- تعالي نشوف رسوم عملية عشان نفهمها كويس :
- لو قلنا ان عندي اكس 1 و اكس 2 ، وان الميو بصفرتين ، والسيجما زي هنا

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

Wednesday 3:51 PM

## استخدام أكثر من متغير

- فيكون الحل ، اتنا منحسب كل  $\mathbf{P}$  لوحدتها ، لكن  $\mathbf{P}$  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 ، بيعمل كل تفاصيل الاكسات مع بعض ، وهجيب الـ  $\mathbf{P}$  بشكل دقيق
- و متنساش ان الـ  $\Sigma$  هي المصفوفة اللي انكلمنا عنها قبل كدة ، وابعادها  $n \times n$
- كمان ان  $|\Sigma|$  معناها قيمة المصفوفة و بتعمل 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

PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

Activities Firefox Web Browser

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

file:///home/abdelrahman/Desktop/desktopopp/new\_steps/stanford machine learning/

25 of 31 Automatic Zoom

## 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

PowerPoint Presentation - H 12 13.pdf - Mozilla Firefox

Activities Firefox Web Browser

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

file:///home/abdelrahman/Desktop/desktopopp/new\_steps/stanford machine learning/

26 of 31 Automatic Zoom

## استخدام أكثر من متغير

- ايجاد قيمة  $P$  عن طريق الفانون الخاص بيها

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

- في حالة عدم تواجد ابسلون ، تحديد قيمتها عبر تكرار فرض قيم لها و اختيار اعلي قيمة لـ **F1 Score**
- تحديد اي نقطة يكون قيمة  $P$  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

- متتساش ان في حالة متغير واحد ، القيمة النهائية لـ  $P$  ه تكون حاصل ضرب قيم  $p$  الصغيرة

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

- بينما في حالة اكتر من متغير

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

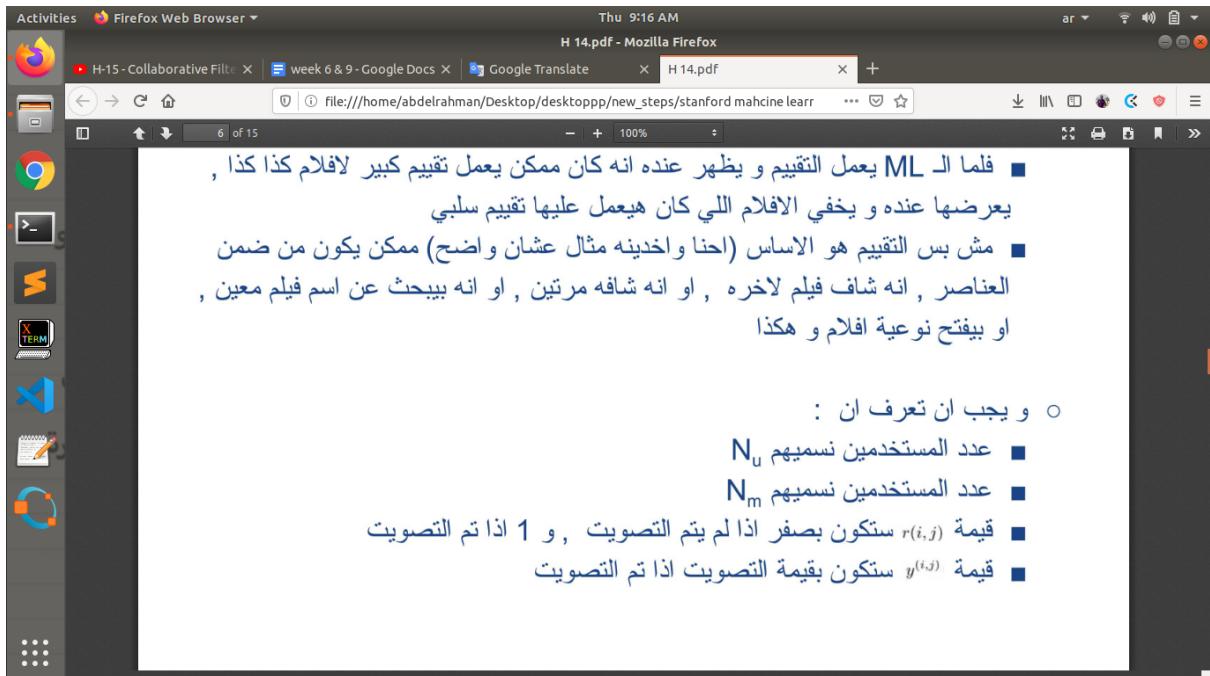
<https://drive.google.com/drive/folders/10pUUwl7sr4mQhdJM8evvKqxX9KGC1gTU>

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 التي انت بتاخدها ببتدى ارشح لك حاجة معينة زى لما تيجى تخش على اليوتيوب هنا اليوتيوب فى الاول خالص لو انتا لسه عامل اكونت جديد مبيكونش عارف حاجه عندك فيبتدى يرشح لك حاجات randomly بعد فتره من استخدام اليوتيوب لو انتا بتسرش 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 ببتدى يعمالك بقا recommendation لحالات قريبة جدا من الى انتا بشوفها او بتدور عليها وده بيساعد جدا فى انه النظام يخليك تستخدeme اطول وقت ممكن ومتقرون تستغنى عنه زى الفيسبوك.

وفي نفس الوقت تقييمات الأشخاص مثلا للأفلام ممكن حد يكون دايما بيحب يتفرج على أفلام أكشن وتلاقيه بيعمل rate الى لبعض الأفلام الأكشن الى شافها وعجبته بينما الرومانسى ممكن يعمل rate ضعيف جدا ولو اتفرج على الفيلم ممكن يقدر يقدم الفيلم او يشوف النهاية عطلو فيها حتى لو معملش system rate او معملش يسنتج انك مش بتحب النوعية دي من الأفلام او بتحبها بناء على انك شوفت الفيلم كله ولا لا طيب هل قدمت كتير من الفيلم عشان تشووف اخره وخلاص على كده هل عملت rate هنا سواء انتا عملت rate او معملتش بس اتفرجت على نوعية معينة من الأفلام الدا system ببتدى يفهم ميك ناحية ايه بربط حتى الحاجات الى بتتميل ليها فيه منها بيفقا كتير جدا بس انتا لسه مشفتهوش فيبدأ يعملها rate حتى لو انتا لسه مشفتهاش بالتقريب مع نفس الحاجات الى انتا شوفتها اصلا وحبيتها من نفس النوعية ويملئ القيم الفارغة فيبتدى يحصلك recommendation بالحالات الى لسه مشفتهاش من نفس النوعية ويبعد عن الحاجات الى مش بتحبها.



Thu 9:17 AM

H 14.pdf - Mozilla Firefox

file:///home/abdelrahman/Desktop/desktopopp/new\_steps/stanford machine learn

- 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 يكون 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 الأفلام الخمسة ، فيه مقدار كذا من  $X_1$  و مقدار كذا من  $X_2$

Thu 9:17 AM

H 14.pdf - Mozilla Firefox

file:///home/abdelrahman/Desktop/desktopopp/new\_steps/stanford machine learn

| Movie                | Alice (1) | Bob (2) | Carol (3) | Dave (4) | $\downarrow$<br>$x_1$<br>(romance) | $\downarrow$<br>$x_2$<br>(action) |
|----------------------|-----------|---------|-----------|----------|------------------------------------|-----------------------------------|
| 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 chases   | 0         | 0       | 5         | 4        | 0.1                                | 1.0                               |
| Swords vs. karate    | 0         | 0       | 5         | ?        | 0                                  | 0.9                               |

- نبدا نجمع الـ features الخاصة بكل فيلم ، هنقول ان  $X^1$  (يعني فيتشرز الفيلم الاول) هتكون

$$\begin{bmatrix} 1 \\ 0.9 \\ 0 \end{bmatrix}$$

وهنا ده بيعتمد على ال linear regression algorithm الخاص بال

Activities Firefox Web Browser Thu 9:21 AM H 14.pdf - Mozilla Firefox

H 14.pdf Mozilla Firefox

file:///home/abdelrahman/Desktop/desktoppp/new\_steps/stanford mahcine learr

13 of 15

حيث :

- $\theta^{(j)}$  هي معاملات كل مستخدم
- $x^{(i)}$  قيم الـ features للفيلم
- $r(i,j)$  بتكون واحد لو المستخدم عمل تصويت للفيلم ، وصفر لو لم يعط تصويت
- $y^{(i,j)}$  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 عايزين نقل القيمة فاختفائها مش هيعمل مشكلة
- متنساش اتنا اضفنا في الآخر قيمة الـ regulation عشان الـ OF

معنى اي  $x$  لو قلنا الفيلم الخامس الى هو مثلا  $x_5$  هنا الفيلم نفسه بالنسبة لمستخدمي النظام كل واحد منهم لديه  $theta$ s معينة خاصة به هو فمعنى كده ان كل  $i$  ب يكون  $x$  of  $theta$ s بالنسبة لكل مستخدم على النظام.

Activities Firefox Web Browser Thu 9:24 AM H 14.pdf - Mozilla Firefox

file:///home/abdelrahman/Desktop/desktoppp/new\_steps/stanford mahcine learr

14 of 15

ولو عايزين نجيب قيم كل المستخدمين ، هنعمل قانون اعم

$$\min_{\theta^{(1)}, \dots, \theta^{(n_u)}} \frac{1}{2} \sum_{j=1}^{n_u} \sum_{i:r(i,j)=1} \left( (\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^n (\theta_k^{(j)})^2$$

ولما اجي اعملها gradient descent عن طريق التفاضل الجزئي ، هنكون كده

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \sum_{i:r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) x_k^{(i)} \quad (\text{for } k = 0)$$

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \left( \sum_{i:r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) x_k^{(i)} + \lambda \theta_k^{(j)} \right) \quad (\text{for } k \neq 0)$$

# Collaborative Filters

طيب فيه بعض الحاجات مبيكونش فيها انك تشوف الحاجه او المحتوى مثلًا زى موقع البيع والشراء بمعنى انى ممكن اخشن على المنتج لكن مقدرش فقيس تغيرات ال user زى الفيلم ده مثلًا شافه كله ولا عمال يقدم وهو بيترج وجاب نهاية الفيلم علطول ففقدر هنا اخد value من كل الحاجات ديه لكن حاجات زى ال online stores وكده بيبتدى يظهر مفهوم تانى اسمه .collaborative filters

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

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