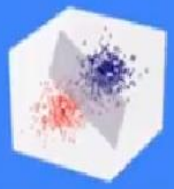
Lec. 10

Linear Classifiers

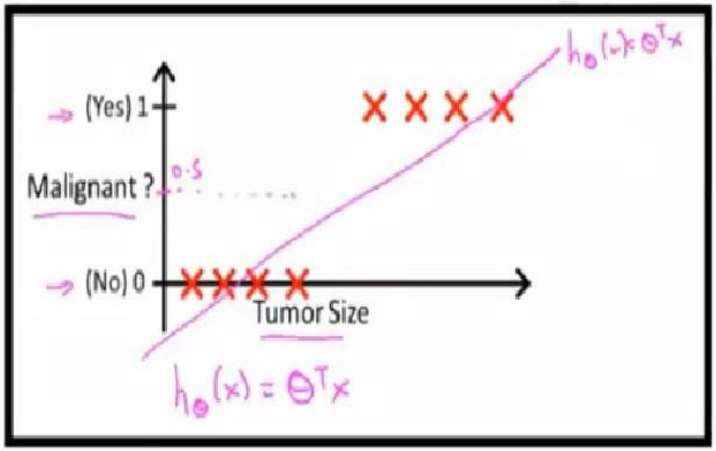
Assist. Prof. Dr. Saad Albawi

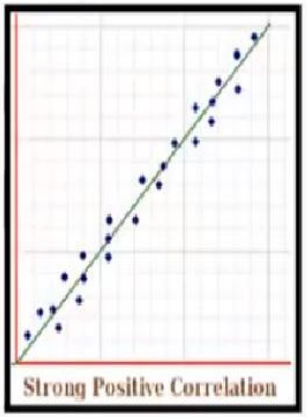


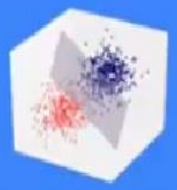
أساسيات التصنيف

التصنيف و الخط الملائم :

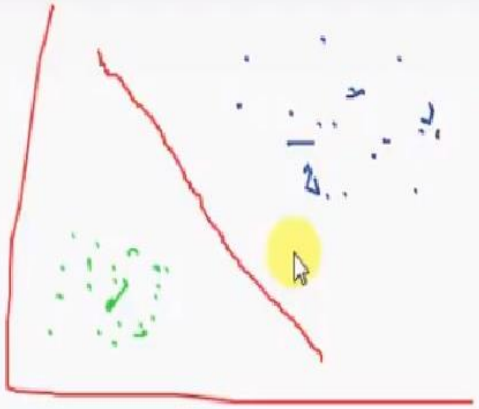
- فى التنبؤ قمنا بعمل الخط الملائم بسهولة
- يصعب فى التصنيف عمل خط يلائم النقط العليا و الدنيا معا
- يقوم الخوارزم بعمل خط يفصل بينهم او curve يمشى عليهم





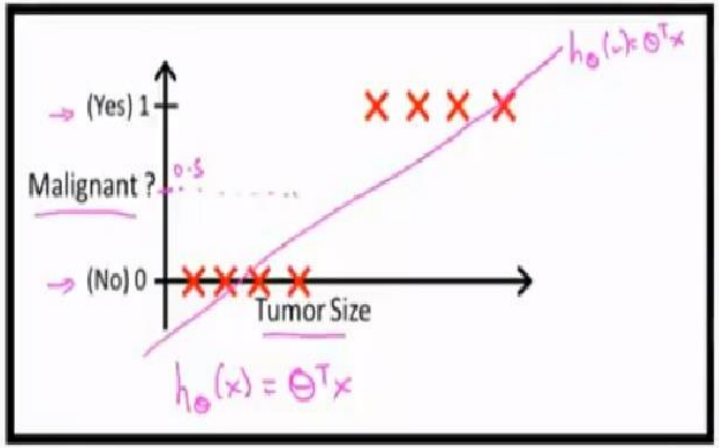


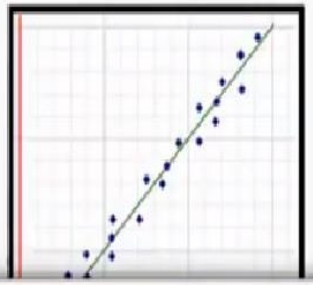
أساسيات التصنيف

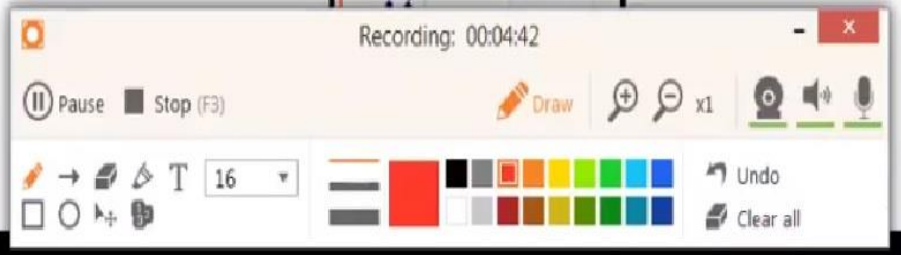


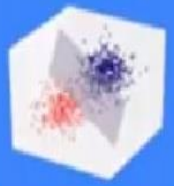
التصنيف و الخط الملائم :

- فى التنبؤ قمنا بعمل الخط الملائم بسهولة
- يصعب فى التصنيف عمل خط يلائم النقط العليا و الدنيا معا
- يقوم الخوارزم بعمل خط يفصل بينهم او curve يمشى عليهم





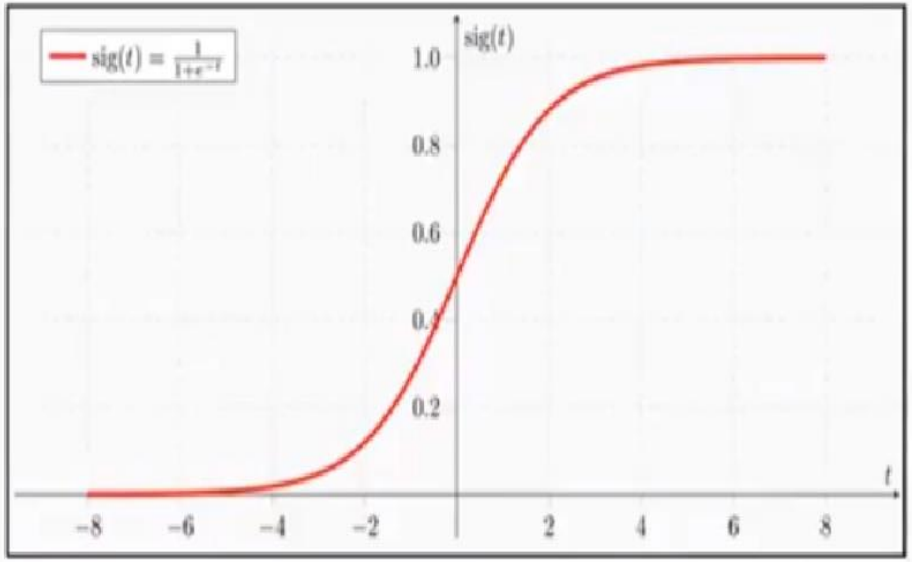




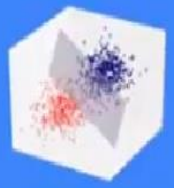
أساسيات التصنيف

الفكرة :

- لازم قيمة الاتش فى النهاية تكون بين الصفر و الواحد , مينفعش اكبر من 1 او اقل من صفر
- لا تنس أن الـ H هى القيمة المتوقعة , وأن الـ Y هى القيمة الحقيقية ,, وأن الـ X هو المدخلات
- يتم هذا عبر معادلة تسمى السيجمويد



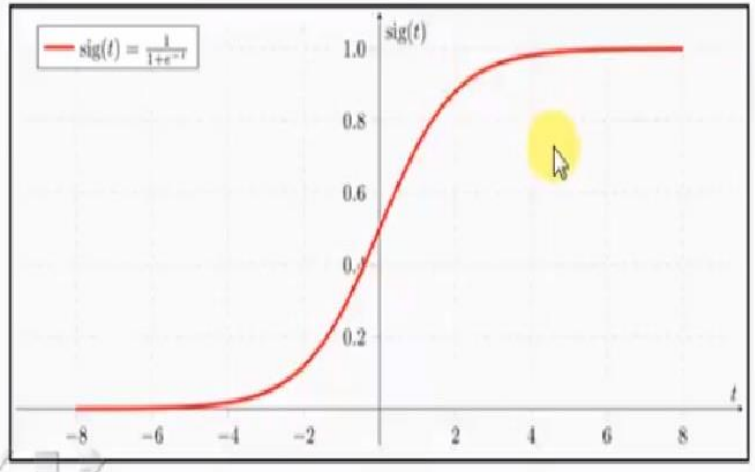
$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$



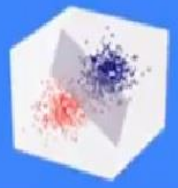
أساسيات التصنيف

السيجمويد :

- حينما تكون قيمة X بالسالب , تصبح الـ e لها أس موجب , وبالتالى تزداد قيمتها , فتصبح قيمة h قليلة
- حينما تقل جدا قيمة X مثلا سالب الف , يكون للـ eأس موجب ضخم , يعنى رقم كبير , فتكون h تساوي 0
- حينما يكون لـ Xقيمة موجبة , يكون للـ e أس سالب , اي قيمة صغيرة , وبالقسمة عليها تزداد قيمة h
- حينما تزداد قيمة X جدا , يكون للـ e أس سالب كبير , أي تساوي صفر , فتكون h تساوي 1



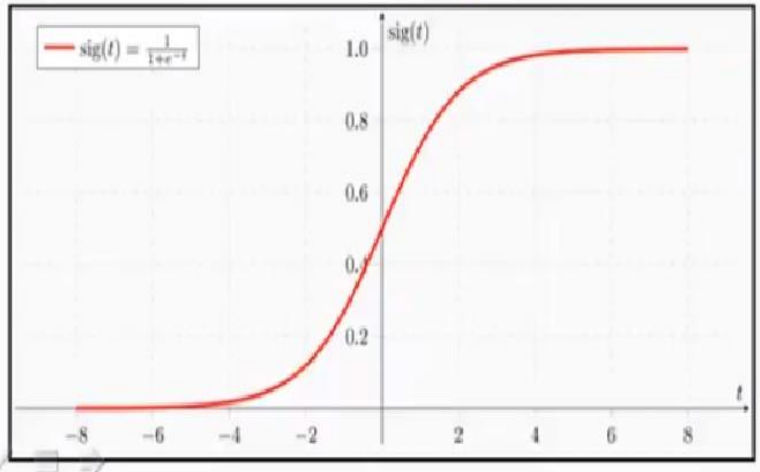
$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$



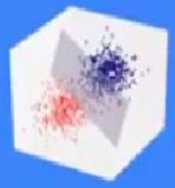
أساسيات التصنيف

السيجمويد :

- أي أن قيم h وهى القيمة المتوقعة , ستتراوي بين 0 و 1 , وهو ما يجعل المخرج احد صنفين
- لو كان هناك قيم بين الصفر والواحد , , فإما يتم تقريبها (0.7) نعتبرها 1 مثلا , او أن يكون باسلوب الاحتمالا
- وبهذا يتم تصنيف البيانات لدينا لقسمين مثل المطلوب (ورم حميد او خبيث , طالب مقبول او مرفوض . .)



$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

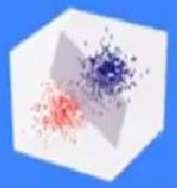


أساسيات التصنيف

المعادلة :

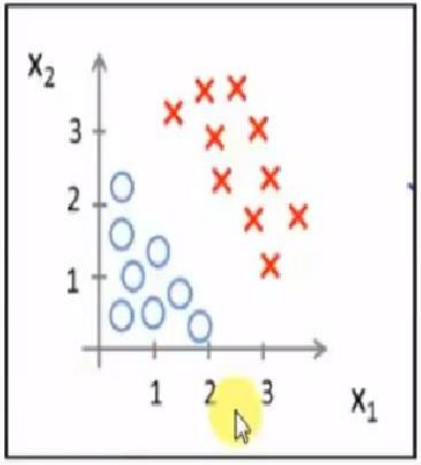
- المقصود بال $\theta^T x$ معادلة حاصل ضرب الثيتات فى الاكسات مثلما فعلنا فى التنبؤ
- قد تكون معادلة من الدرجة الاولى او الثانية او اكثر , حسب كمية التعقيد فى المسألة

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$



أساسيات التصنيف

مثال عملى :



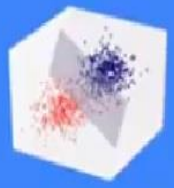
- لدينا عدد من المنتجات ونريد تقسيمها لمنتجات مقبولة باللون الازرق ومنتجات مرفوضة باللون الأحمر , اعتمادا على معلومتين X2 , X1

- معادلة h وهى القيمة المتوقعة , ستكون من الدرجة الأولى , وبالتالى ستعتمد على , X2 , X1, فنحتاج لثلاث ثيتات كالتالى :

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

- بفرض ثيتا صفر تساوي سالب 3 , وكلا من ثيتا 1 و 2 تساوي 1 تكون المعادلة :

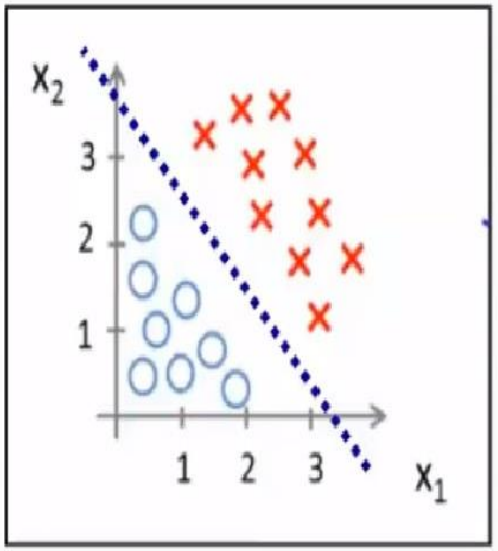
$$-3 + x_1 + x_2 \geq 0$$



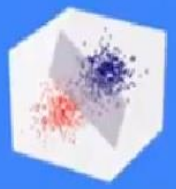
أساسيات التصنيف

مثال عملى :

- هذه المتباينة , تقوم بعمل خط مستقيم , يقسم المساحة لقسمين , أحد الجانبين يحتوي على المقبول و الاخر المرفوض



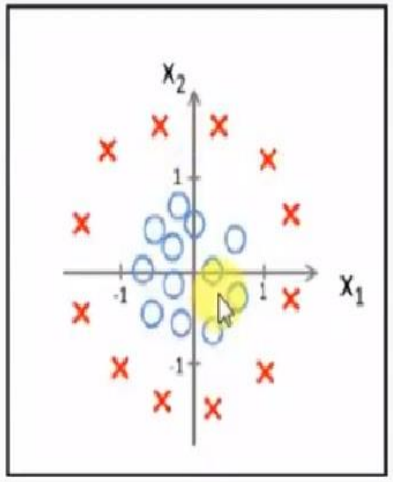
$$-3 + x_1 + x_2 \geq 0$$



أساسيات التصنيف

مثال آخر :

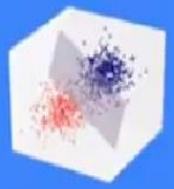
- المنتجات المقبولة فى المنتصف والمرفوضة تحيط بها
- استخدام خط للفصل بينهم لن يكون ممكنا , وبالتالى المعادلة بالدرجة الاولى غير متاح
- يتم استخدام معادلة بالدرجة الثانية كالتالى :



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

- بفرض ثيتا صفر تساوي سالب 1 , , وكلا من ثيتا 1 و 2 تساوي صفر , وكلا من ثيتا 3 و 4 تساوي 1 , تكون :

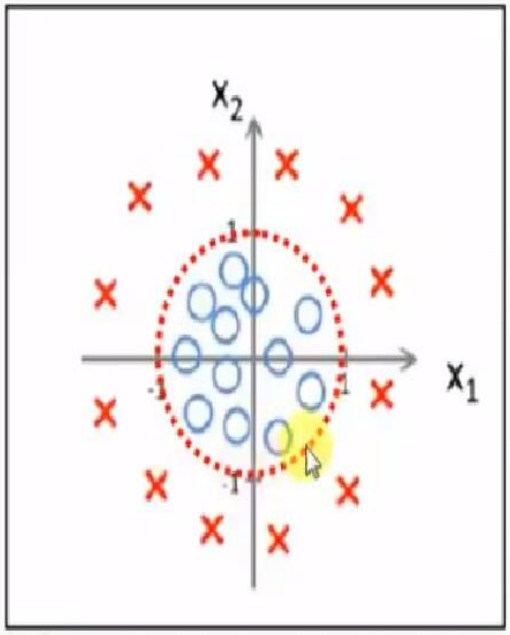
$$x_1^2 + x_2^2 \geq 1$$



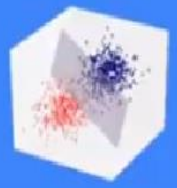
أساسيات التصنيف

مثال آخر :

- وهى معادلة دائرة , مركزها نقطة الأصل , ونصف قطرها 1 , أي أن كل ما داخل الدائرة مقبول , وكل ما خارجها مرفوض



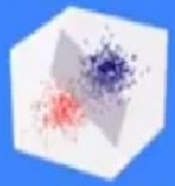
$$x_1^2 + x_2^2 \geq 1$$



أساسيات التصنيف

وبالتالى :

- نجحنا بعمل فاصل (خط مستقيم او دائرة او منحنى) يقوم بالفصل بين الصنفين المطلوبين لدينا , المقبول و المرفوض , السليم و التالف , الخبيث والحميد
- يمكن للمعادلات ان تكون من الدرجات العليا حسب مدي تعقيدها
- يتم فرض قيمة ثيتا مؤقتا , وسنتعلم لاحقا كيفية تحديدها
- عبر ادخال الاكسات (البيانات) يقوم الخوارزم بتحديد هل اي القسمين ستكون هى
- قد تكون الاشكال الفاصلة أشكال أكثر تعقيد من مجرد دائرة او خط



المعادلة المستخدمة للتصنيف

ما هى المعادلة المستخدمة :

- أي الصيغة الرياضية التى يجب استخدامها , لتمكننا من عمل تصنيف سليم للمنتجات كما شرحنا
- فى التنبؤ استخدمنا المعادلة :

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

- لكنها لن تكون صالحة فى تقسيم البيانات Classification , وعلينا البحث عن صيغة جديدة

المعادلة المستخدمة للتصنيف

المعادلة المستخدمة :

- تسمى معادلة الخطأ Cost و التى نستخدم لها الرمز J , والهدف تقليلها بقدر الإمكان
- المعادلة ليست واحدة , بل اثنين , واحدة فى حالة y تساوي 0 وواحدة فى حالة y تساوي 1
- المعادلة ببساطة هى مجموع سالب قيم لوغاريتم h او 1 ناقص h , ويتم قسمتها على m وهو عدد العناصر
- سيتم جمع المعادلتين معا كما سنري بعد قليل

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

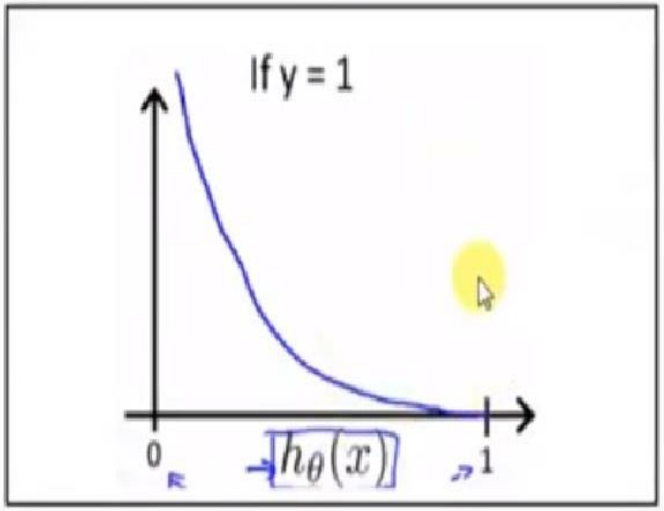
$$\text{Cost}\left(h_\theta(x), y\right) = -\log\left(h_\theta(x)\right) \qquad \text{if y} = 1$$

$$\text{Cost}\left(h_\theta(x), y\right) = -\log\left(1 - h_\theta(x)\right) \qquad \text{if y} = 0$$

المعادلة المستخدمة للتصنيف

الحالة الأولى : y = 1

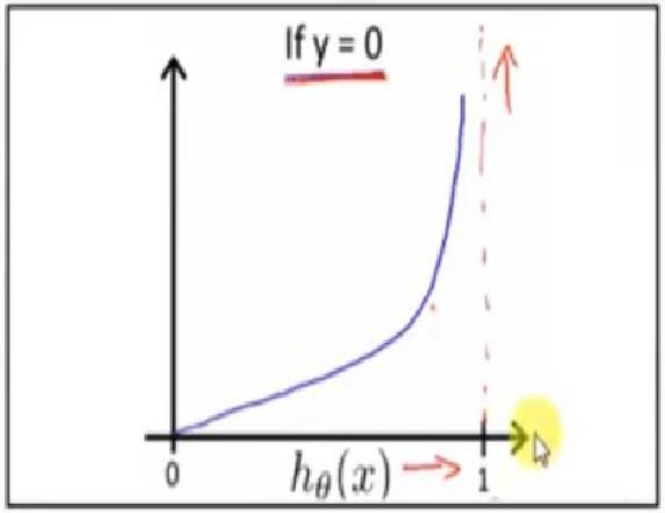
- المحور الافقى به قيمة الـ h و الرأسى فيه قيمة لـ
- حينما تقل قيمة h لصفر , تكون قيمة الخطا كبيرة , لانها بعيدة عن القيمة الحقيقية 1
- حينما تقترب قيمة h من 1 , يكون الخطأ قليل و يصل لصفر حينما تساوي 1 , لان القيمة الحقيقية تساوي 1



المعادلة المستخدمة للتصنيف

الحالة الثانية : y = 0

- المحور الافقى به قيمة الـ h و الرأسى فيه قيمة لـ
- حينما تقل قيمة h لصفر , تكون قيمة الخطا صغيرة , لانها قريبة عن القيمة الحقيقية 1
- حينما تقترب قيمة h من 1 , يكون الخطأ كبير , لان القيمة الحقيقية تساوي 1



المعادلة المستخدمة للتصنيف

المعادلة المجمعة :

- يتم تجميع المعاجلتين معا , فى كلتا حالتى y , حينما تكون بصفر وبواحد , بالصيغة التالية :

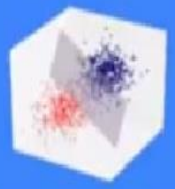
$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))]$$

- حينما تكون y تساوي صفر , يختفى الجزء الاول من المعادلة و تصير هكذا :

$$J(\theta) = -\frac{1}{m}[(1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))]$$

- حينما تكون y تساوي 1 , يختفى الجزء الثانى من المعادلة و تكون هكذا

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)})]$$



المعادلة المستخدمة للتصنيف

ايجاد الثيتا :

- كى لا ننسى , الثيتا هى معاملات الإكسات , التى يتم ايجاجها , وذلك للحصول على معادلة الكيرف الاكثر ملائمة best fit curve

- يتم ايجاد الثيتا بهذه بالمعادلة , حيث :

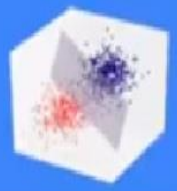
- الرمز j يشير لرقم الثيتا المطلوبة (. . . , 2 , 1 , 0) , أي انه سيتم تكرارها لكل الثيتات المطلوبة

- الفا هى معامل يشير لمقدار خطوة الحساب (مثلما فعلنا فى التنبؤ)

- المعادلة بعد الفا هى تفاضل المعادلة الكبيرة السابقة (مش لازم تعرف اثباتها)

- يتم وضع اكس 0 لثيتا صفر و اكس 1 لثيتا 1 و هكذا

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

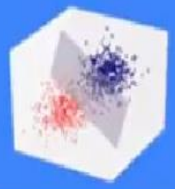


مثال عملي للتصنيف

ايجاد الثيتا :

- كى لا ننسى , الثيتا هى معاملات الإكسات , التى يتم ايجاجها , وذلك للحصول على معادلة الكيرف الاكثر ملائمة best fit curve
- يتم ايجاد الثيتا بهذه بالمعادلة , حيث :
  - الرمز j يشير لرقم الثيتا المطلوبة (. . . , 2 , 1 , 0) , أي انه سيتم تكرارها لكل الثيتات المطلوبة
  - الفا هى معامل يشير لمقدار خطوة الحساب (مثلما فعلنا فى التنبؤ)
  - المعادلة بعد الفا هى تفاضل المعادلة الكبيرة السابقة (مش لازم تعرف اثباتها)
  - يتم وضع اكس 0 لثيتا صفر و اكس 1 لثيتا 1 و هكذا

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$



مثال عملي للتصنيف

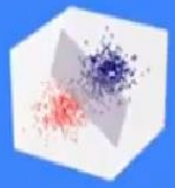
المعادلة كمصفوفات :

- يجب ان نقوم بتحويلها لمصفوفة للتعامل معها باي لغة برمجة , فستكون المعادلة كالتالى , حيث :

(n+1) x 1

- ثيتا هنا هى مصفوفة عمود واحد , وفيها صفوف بعدد الثيتات (nX1)
- الفا و m هى نفس الرموز السابق ذكرها
- اكس ترانسبوز , هى تدوير مصفوفة أكس الكبيرة , كانت (mXn) , وستصير (nXm)
- اكس فى ثيتا , المقصود بها ضرب مصفوفة اكسات فى مصفوفة ثيتات
- الواي هنا هى مصفوفة قيم واي (اصفار ووحايد) يتم طرحها منها (mX1)

$$\theta := \theta - \frac{\alpha}{m} X^T \left(g\left(X\theta\right) - \vec{y} \right)$$

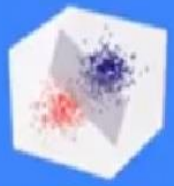


مثال عملي للتصنيف

مثال عملى :

الاسم	الوزن	السن	فقر الدم	مريض
أحمد	70	18	1	1
حامد	88	22	0	1
منى	91	38	0	1
سيد	65	21	1	0
لبنى	79	25	1	0

- لدينا بيانات افراد , و هل مصابين بالانفلونزا او لا
- عدد العينة m هو 5
- عدد العناصر features و رمزها n هو 3
- الارقام الخضراء هى x
- الارقام الحمراء هى y
- عدد الثيتات سيكون 4 , لانه يكون عدد الـ features زائد 1
- الاسماء لن تستخدم لانها غير مؤثرة فى الحساب



مثال عملي للتصنيف

المعادلة كمصفوفات :

- يجب ان نقوم بتحويلها لمصفوفة للتعامل معها باي لغة برمجة , فستكون المعادلة كالتالى , حيث :
  - ثيتا هنا هى مصفوفة عمود واحد , وفيها صفوف بعدد الثيتات (nX1)
  - الفا و m هى نفس الرموز السابق ذكرها
  - اكس ترانسبوز , هى تدوير مصفوفة أكس الكبيرة , كانت (mXn) , وستصير (nXm)
  - اكس فى ثيتا , المقصود بها ضرب مصفوفة اكسات فى مصفوفة ثيتات
  - الواي هنا هى مصفوفة قيم الواي (اصفار ووحايد) يتم طرحها منها (mX1)

$$\theta := \theta - \frac{\alpha}{m} X^T \left(g\left(X\theta\right) - \vec{y} \right)$$

مثال عملي للتصنيف

المصفوفات :

مصفوفة ثيتا 4X1

1
1
1
1
1

مصفوفة y 5X1

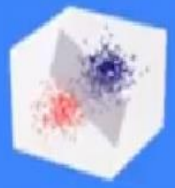
1
1
1
0
0

مصفوفة X 5X4

1	1	18	70
1	0	22	88
1	0	38	91
1	1	21	65
1	1	25	79

تدوير مصفوفة X 4X5

1	1	1	1	1
1	0	0	1	1
18	22	38	21	25
70	88	91	65	79



مثال عملي للتصنيف

عملية ضرب X في ثيتا :

- حاصل ضرب مصفوفة 5X4 في مصفوفة 4X1 ستكون مصفوفة 5X1
- ستكون كالتالى :

90

111

130

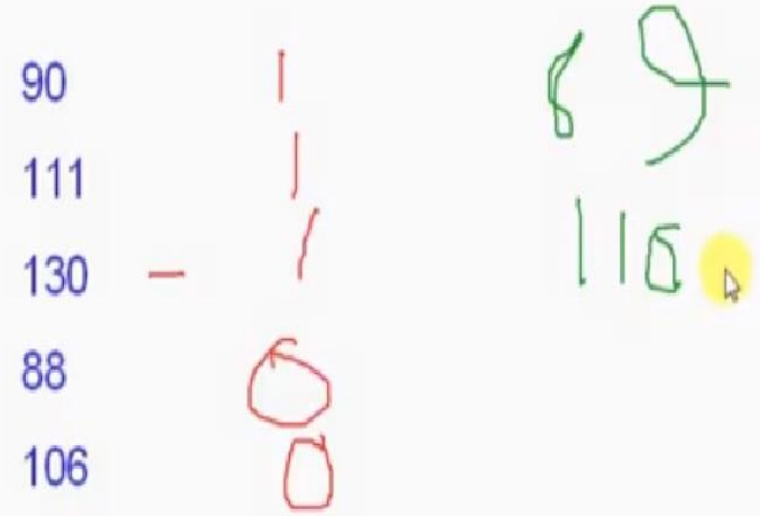
88

106

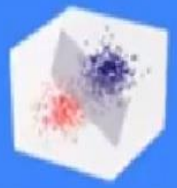
مثال عملي للتصنيف

عملية ضرب X في ثيتا :

- حاصل ضرب مصفوفة 5X4 في مصفوفة 4X1 ستكون مصفوفة 5X1
- ستكون كالتالى :



سيظهر إعلان بعد 4.



مثال عملي للتصنيف

الطرح من y :

- كلا المصفوفتين هى 5X1 , وبالتالى الطرح يمكن أن يتم و ستكون أيضا 5X1
- ستكون كالتالى :

89

110

129

88

106

مثال عملي للتصنيف

الطرح من y :

- كلا المصفوفتين هى 5X1 ، وبالتالى الطرح يمكن أن يتم و ستكون أيضا 5X1
- ستكون كالتالى :

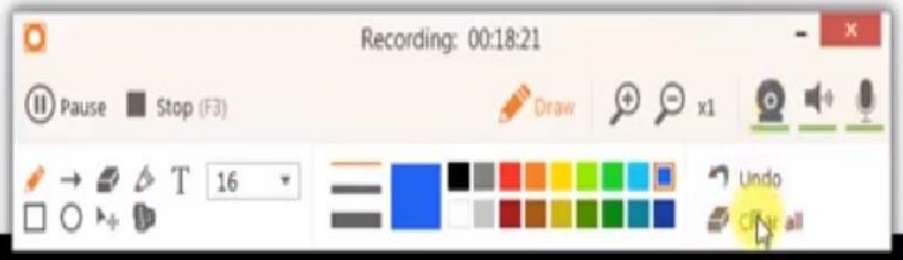
89

110

129

88

106



مثال عملي للتصنيف

الضرب فى تدوير إكس :

- تدوير إكس هو 4X5 و المصفوفة الموجودة هى 5X1 فالضرب ممكن . وستكون 4X1

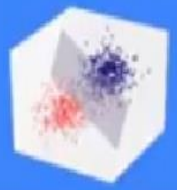
1	1	1	1	1
1	0	0	1	1
18	22	38	21	25
70	88	91	65	79

X

89
110
129
88
106

=

522
283
13422
41743

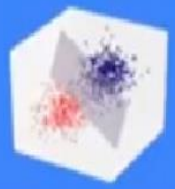


مثال عملي للتصنيف

الضرب فى الفا على m :

- بفرض الفا تساوي 0.5 نقسمها على m التى تساوي 5 تكون 0.1
- نضرب 0.1 فى القيم الاربعة تكون

52.2
28.3
1342.2
4174.3



مثال عملي للتصنيف

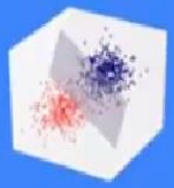
اخيرا نطرح قيم ثيتا الاصلية منها تكون :

1		52.2		-51.2
1	-	28.3	=	-27.3
1		1342.2		-1341.2
1		4174.3		- 4173.3

مثال عملي للتصنيف

التكرار :

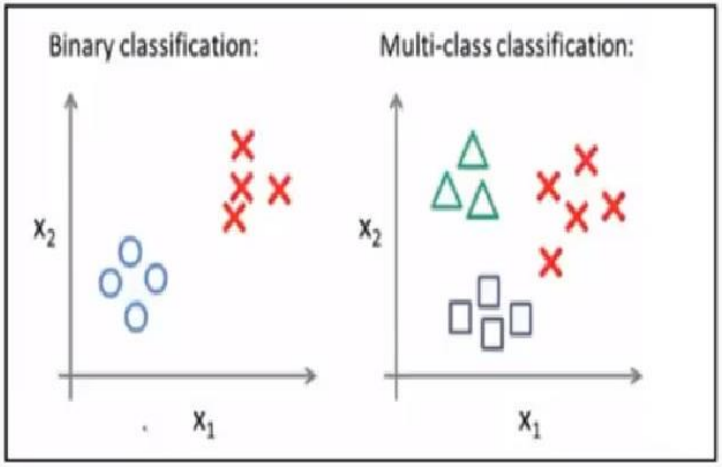
- يتم تكرار العملية مرات عديدة , وكل مرة يتم ادخال قيم ثيتا التم تم اخراجها , حتى نصل لقيم ثيتا المثالية
- قيم ثيتا هذه , هى التى ستقوم برسم الخط او الدائرة او الكيرف الذي سيقوم بالفصل بين الصنفين , المريض و غير المريض
- اذا جائت الينا بيانات فرد جديد , ونريد ان نعلم هل هو مريض ام غير مريض , فنقوم بضرب قيم اكسات (features) فى الثيتات المثالية , وسيكون الرقم اما صفر او واحد , وبهذا نقوم قد قمنا بعمل تصنيف سليم



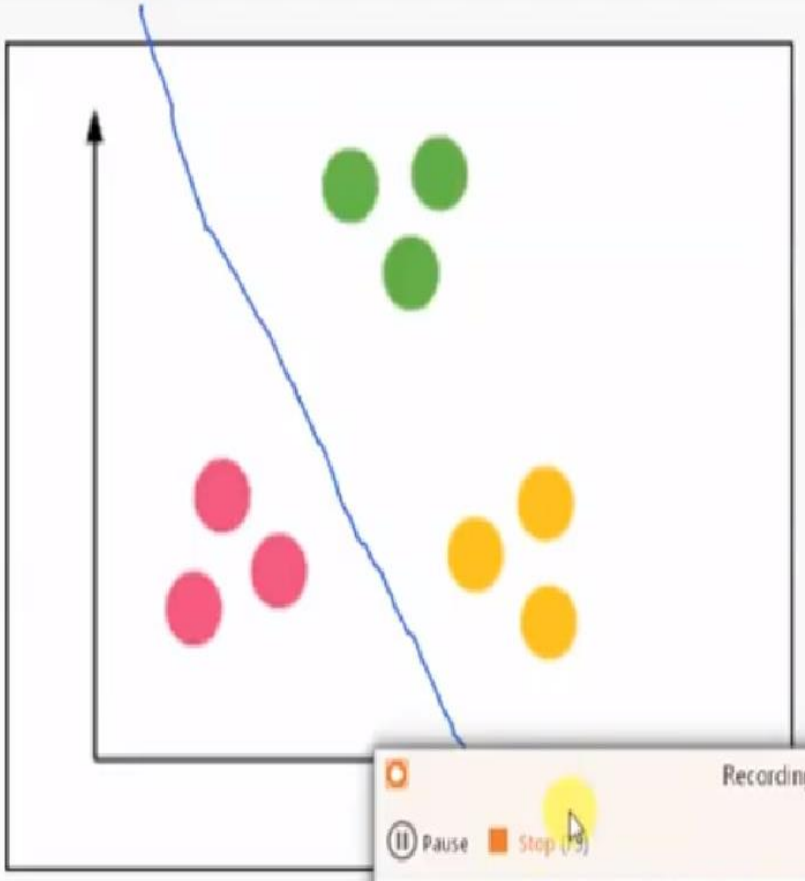
التصنيف المتعدد MultiClassification

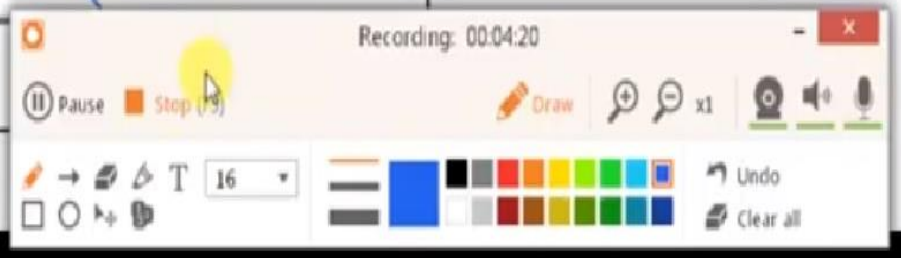
معناه :

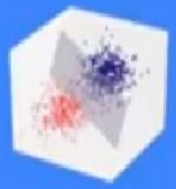
- أن التصنيف ليس فقط لعنصرين او إجابتين , ولكن لأكثر من إجابة
- بدلا من أن يكون الشخص مريض أو غير مريض , يكون مريض جدا , مريض عادي , متماثل للشفاء , غير مريض
- بدلا من المنتج مقبول او مرفوض : منتج متميز , منتج مقبول , مرفوض لكن يمكن اصلاحه , مرفوض تماما
- بدلا من استخدام 1 و 0 للتصنيف , سيكون هناك ارقام عديدة (0 , 1 , 2 , 3 , . . .)



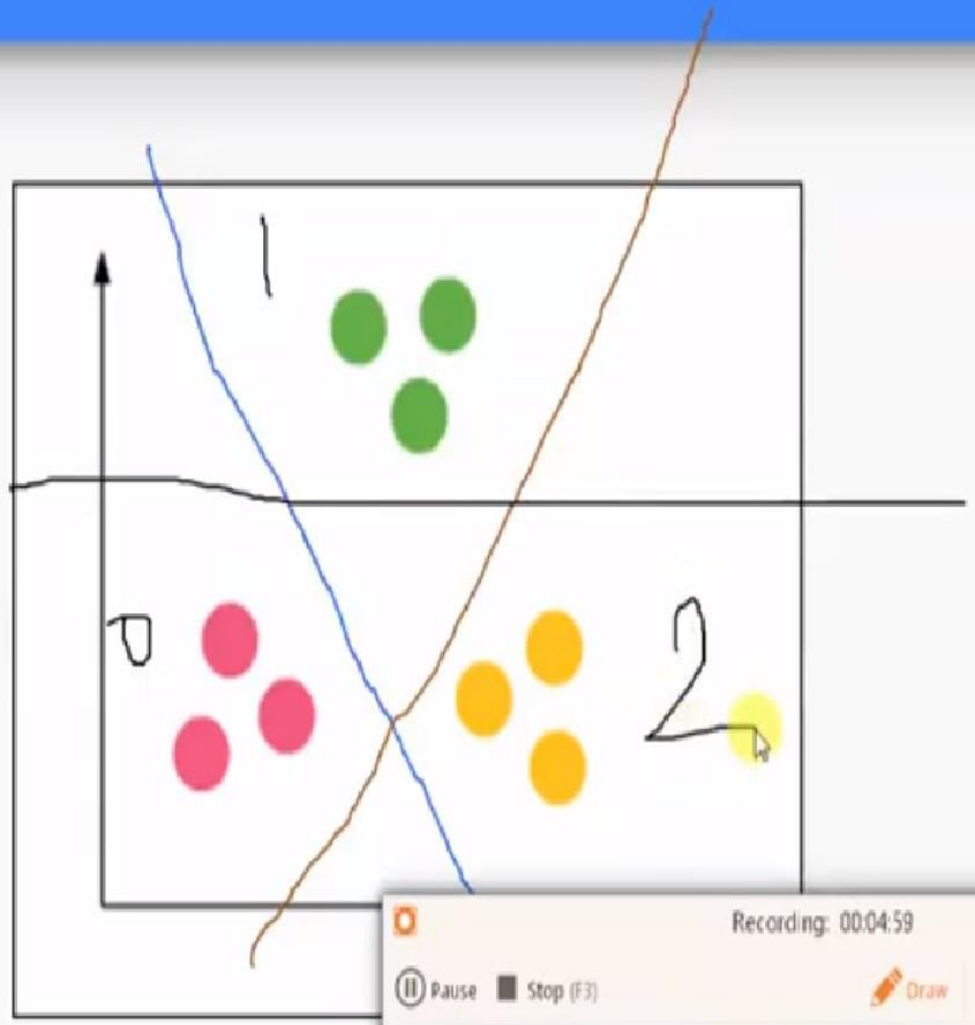
MultiClassification التصنيف المتعدد



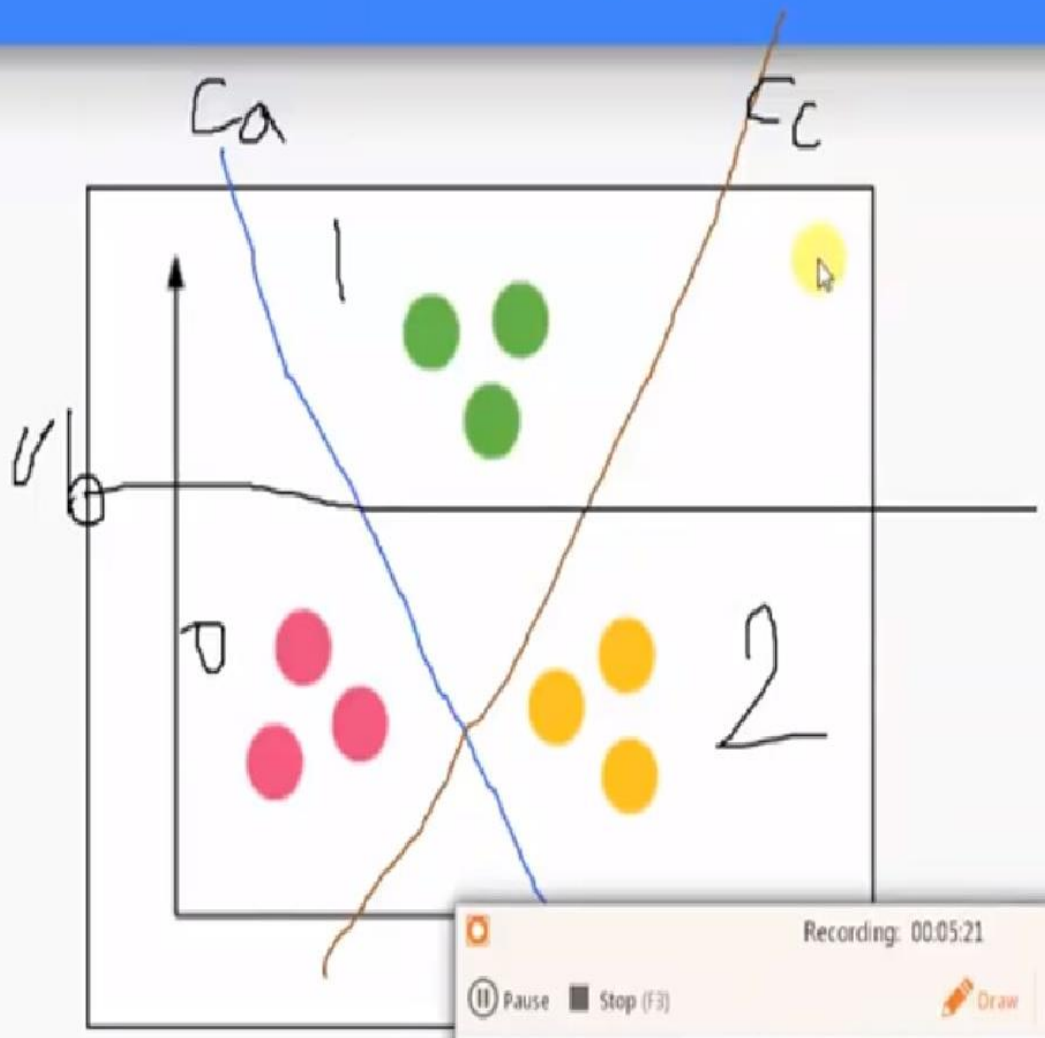




MultiClassification التصنيف المتعدد

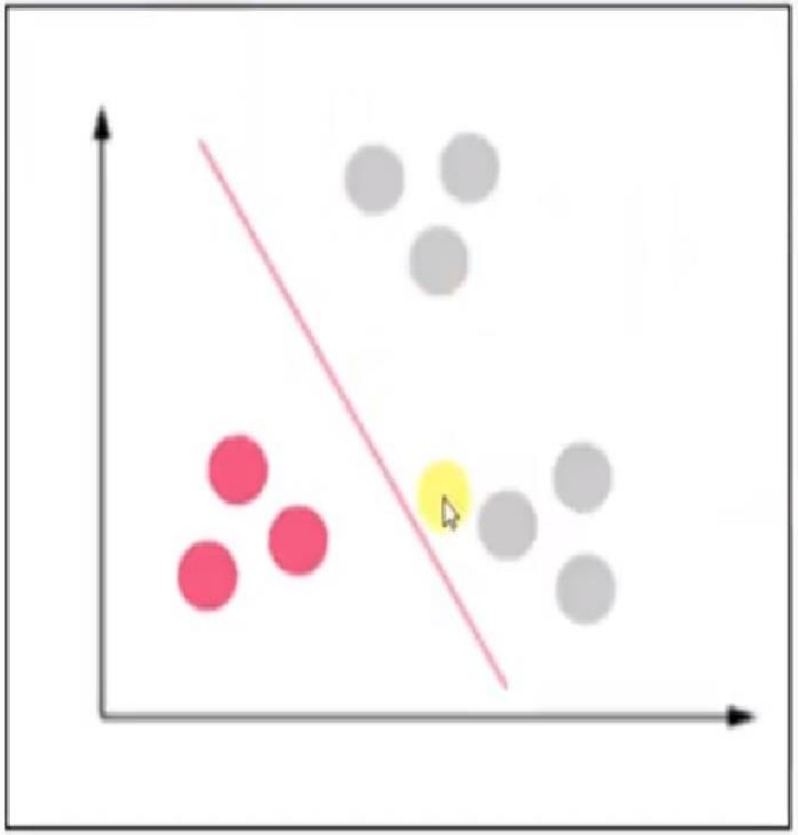


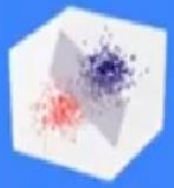
MultiClassification التصنيف المتعدد



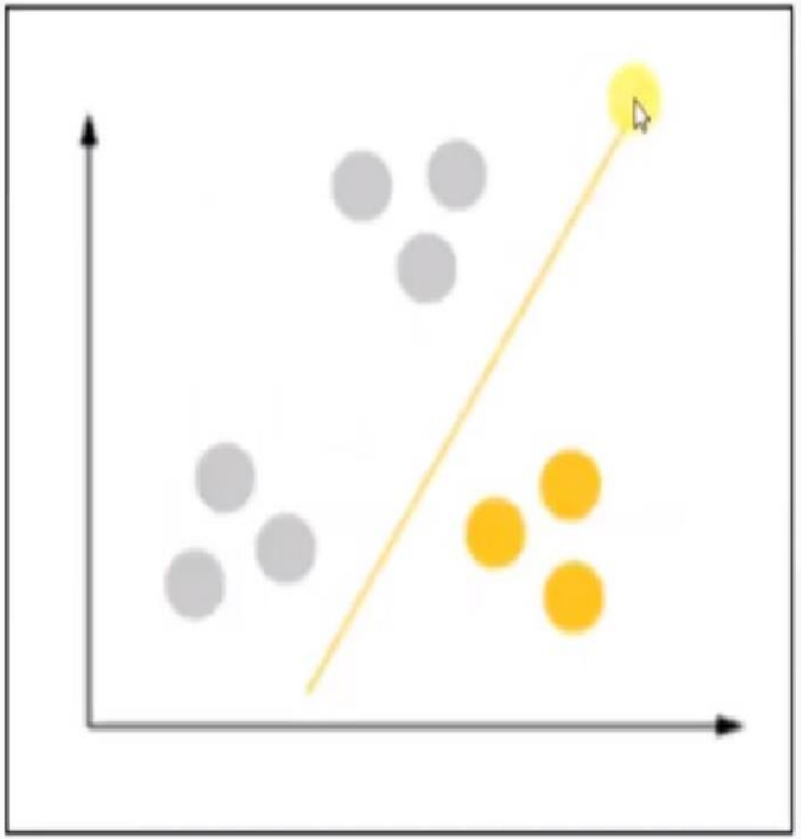
Recording: 00:05:21
Pause
Stop (F3)
Draw
x1
16
Undo
Clear all

التصنيف المتعدد MultiClassification

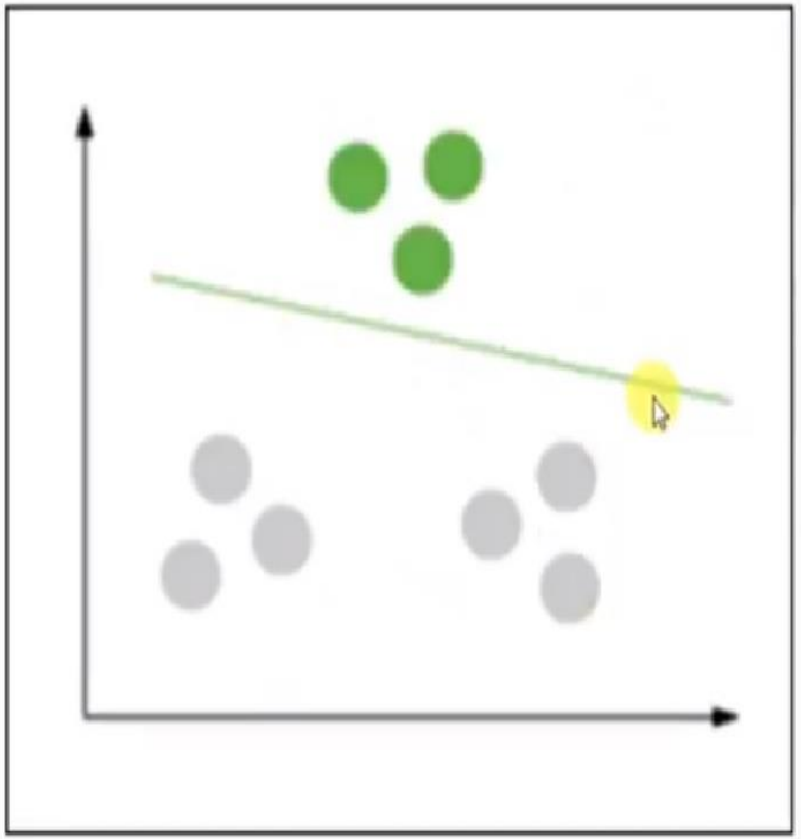


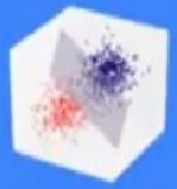


التصنيف المتعدد MultiClassification



التصنيف المتعدد MultiClassification

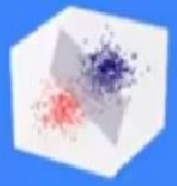




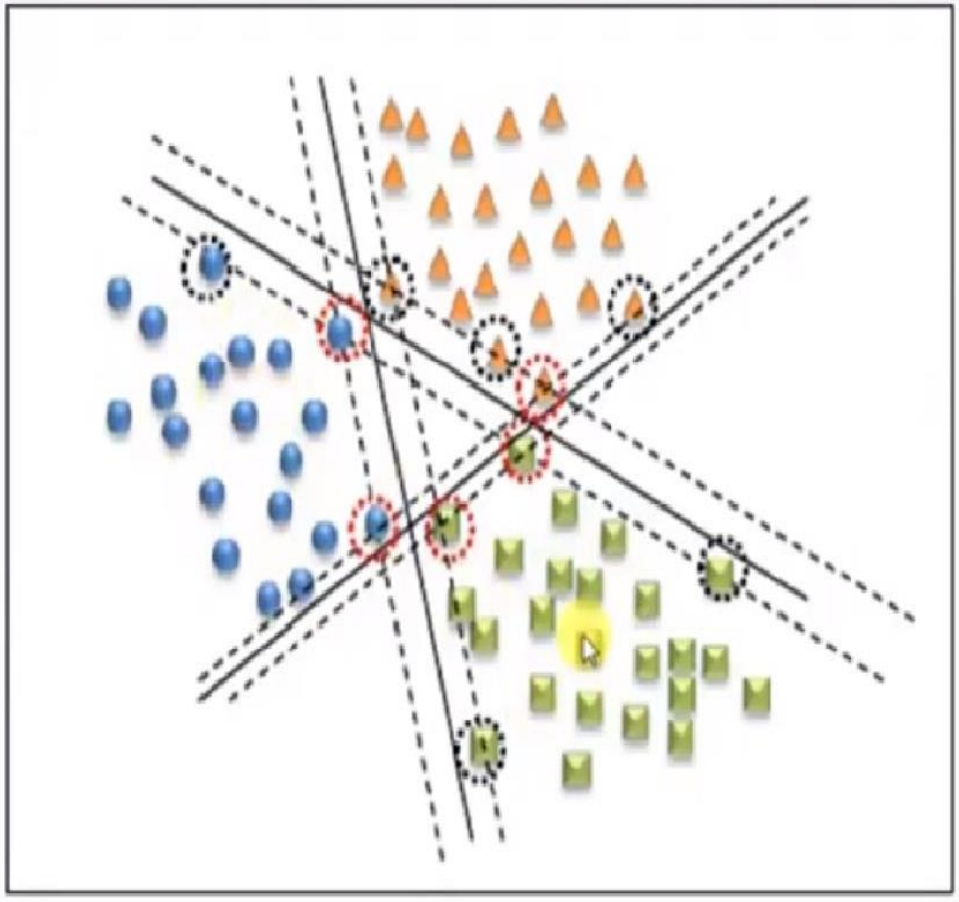
التصنيف المتعدد MultiClassification

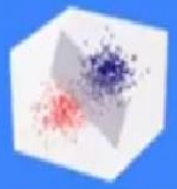
الواحد للجميع , والجميع للجميع :

OVA vs AVA

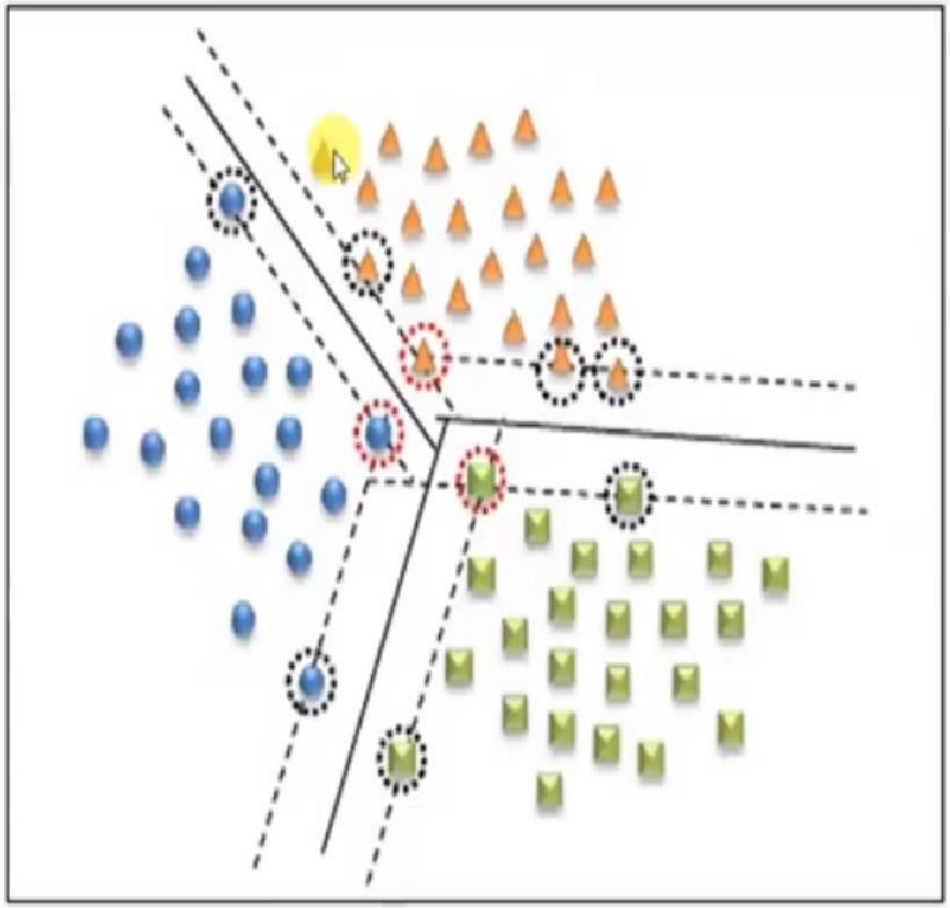


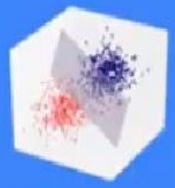
MultiClassification التصنيف المتعدد





MultiClassification التصنيف المتعدد

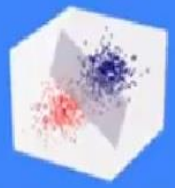




الضبط الزائد و الناقص UF , OF

المعنى :

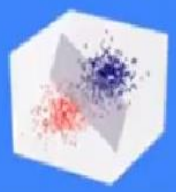
- يقصد بالضبط الناقص Underfitting حينما تكون المعادلة الرياضية المستخدمة أضعف وأقل من تغطية جميع النقاط الموجودة
- يقصد بالظبط الزائد Overfitting حينما تقوم المعادلة الرياضية المستخدمة بضبط جميع النقاط بوجه كامل , مما يسبب أخطاء فى التوقع



الضبط الزائد و الناقص UF , OF

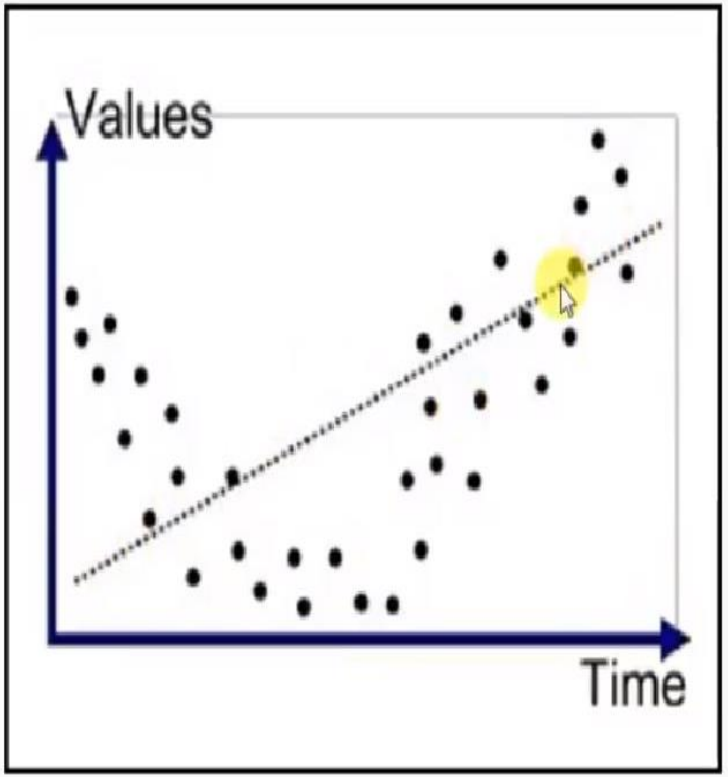
المعنى :

- يقصد بالضبط الناقص Underfitting حينما تكون المعادلة الرياضية المستخدمة أضعف وأقل من تغطية جميع النقاط الموجودة
- يقصد بالظبط الزائد Overfitting حينما تقوم المعادلة الرياضية المستخدمة بضبط جميع النقاط بوجه كامل , مما يسبب أخطاء فى التوقع



الضبط الزائد و الناقص UF , OF

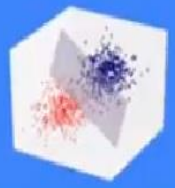
الضبط الناقص :



- تكون المعادلة بسيطة , وصغيرة , وغير قادرة على الالمام و المرور بأغلب النقاط الموجودة

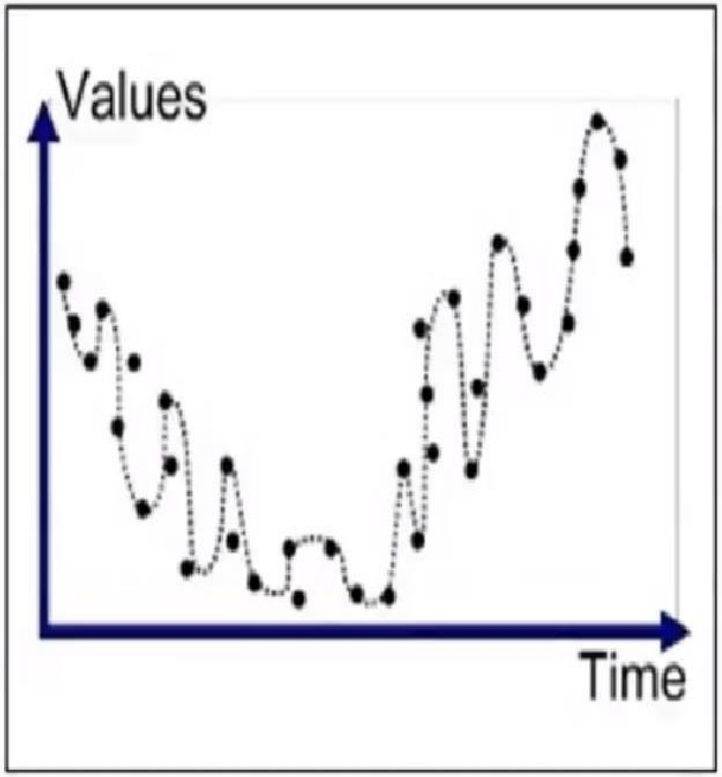
- يكون لها مشكلة كبيرة فى توقع القيم القادمة , وتكون نسبة الخطأ كبيرة

- غالبا تأتى بسبب نقط البيانات (إما عدد العينات m او عدد العناصر n), أو اختيار المعادلة غير المناسبة



الضبط الزائد و الناقص UF , OF

الضبط الزائد :

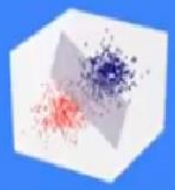


- حينما تقوم المعادلة بالمرور على كل النقاط بالظبط , وبشكل مبالغ فيه

- الشكل يبدو جميل , لكن المشكلة فى توقع اي قيمة جديدة

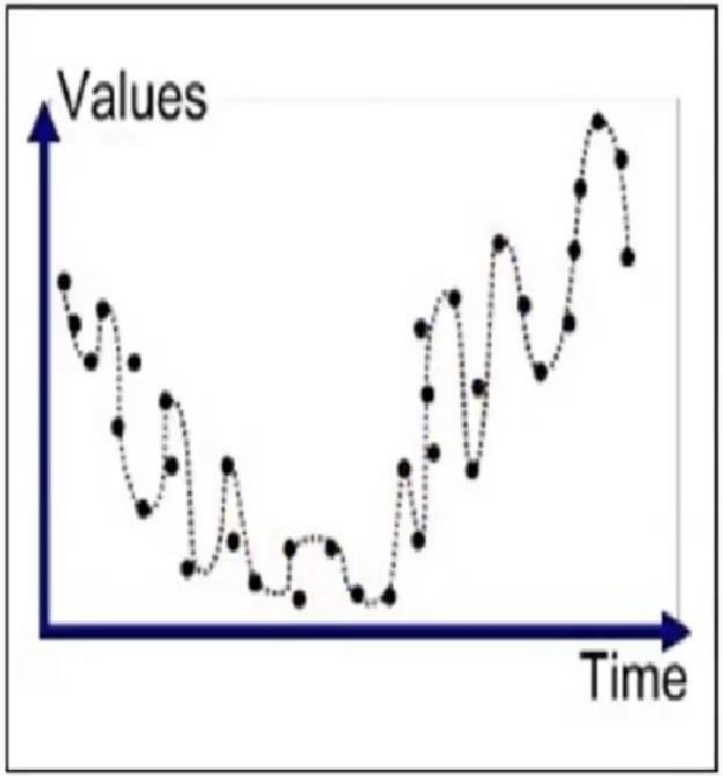
- غالبا تأتى بسبب الزيادة الغير مفيدة للبيانات (إما عدد العينات mاو عدد العناصر n), أو اختيار المعادلة غير المناسبة

الضبط الزائد و الناقص UF , OF



الضبط الزائد و الناقص UF , OF

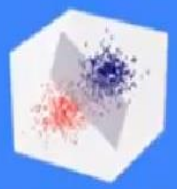
الضبط الزائد :



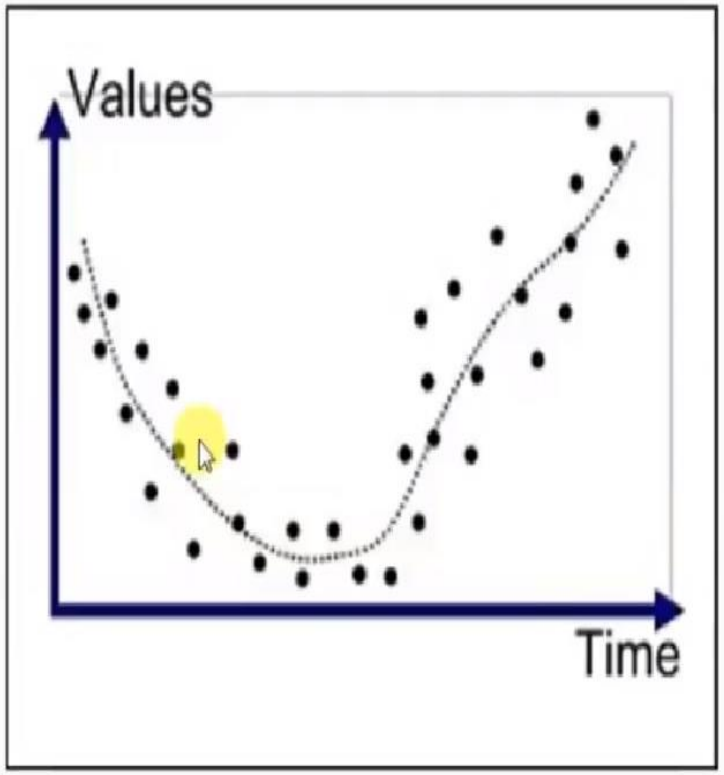
- حينما تقوم المعادلة بالمرور على كل النقاط بالظبط , وبشكل مبالغ فيه

- الشكل يبدو جميل , لكن المشكلة فى توقع اي قيمة جديدة

- غالبا تأتى بسبب الزيادة الغير مفيدة للبيانات (إما عدد العينات mاو عدد العناصر n), أو اختيار المعادلة غير المناسبة



الضبط الزائد و الناقص UF , OF

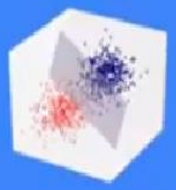


الضبط المناسب :

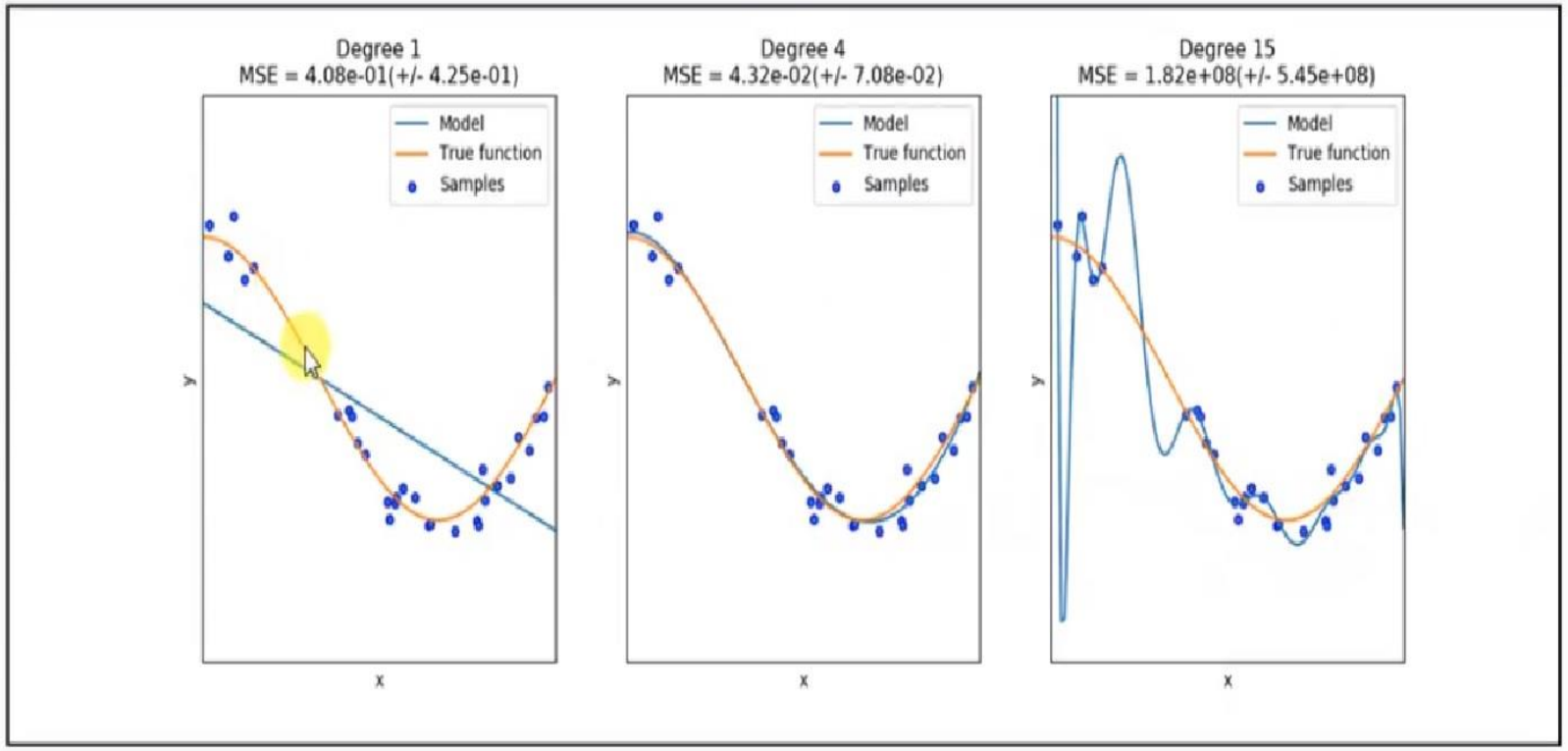
- حينما تقوم المعادلة بالمرور على أغلب النقاط بالظبط , وبشكل مناسب

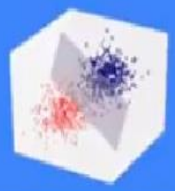
- الشكل قد لا يبدو تاما , لكن التوقع يكون قريب جدا من الصحيح

- يتم ظبطها باختيار عدد مناسب من البيانات , والمعادلة السليمة و استخدام اسلوب التنعيم (Regularization)

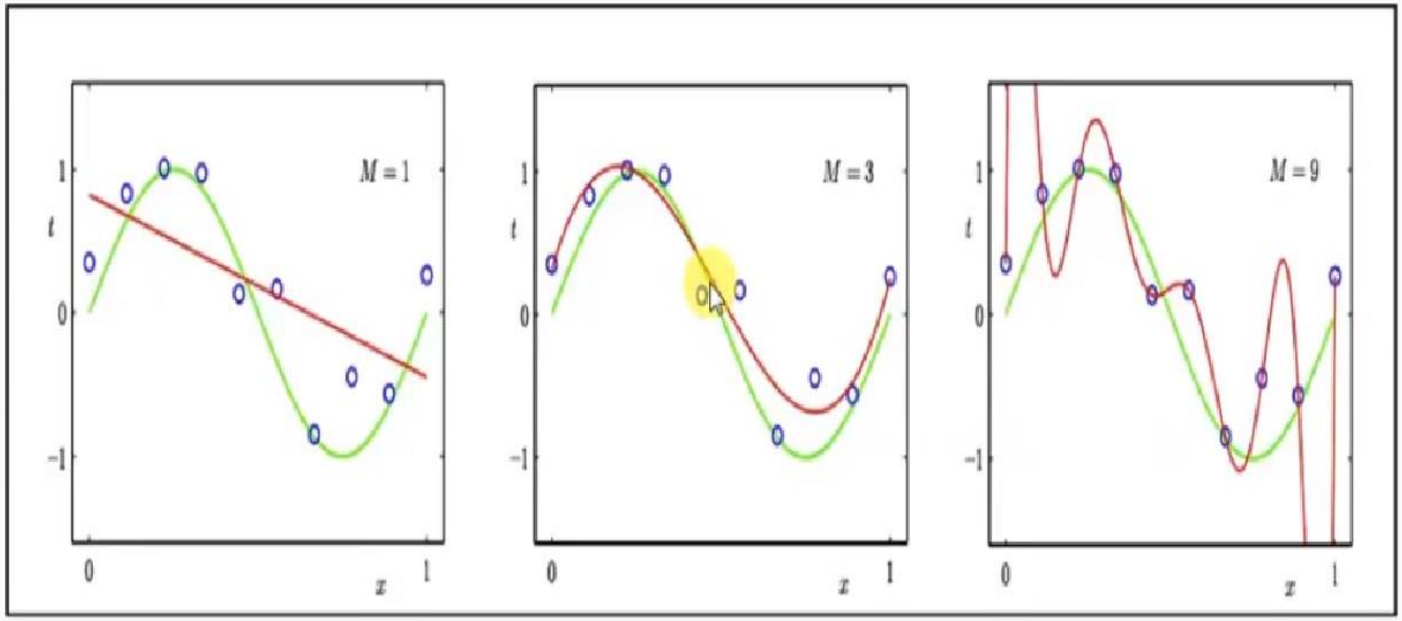


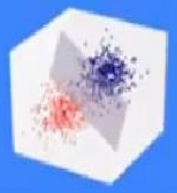
الضبط الزائد و الناقص UF , OF



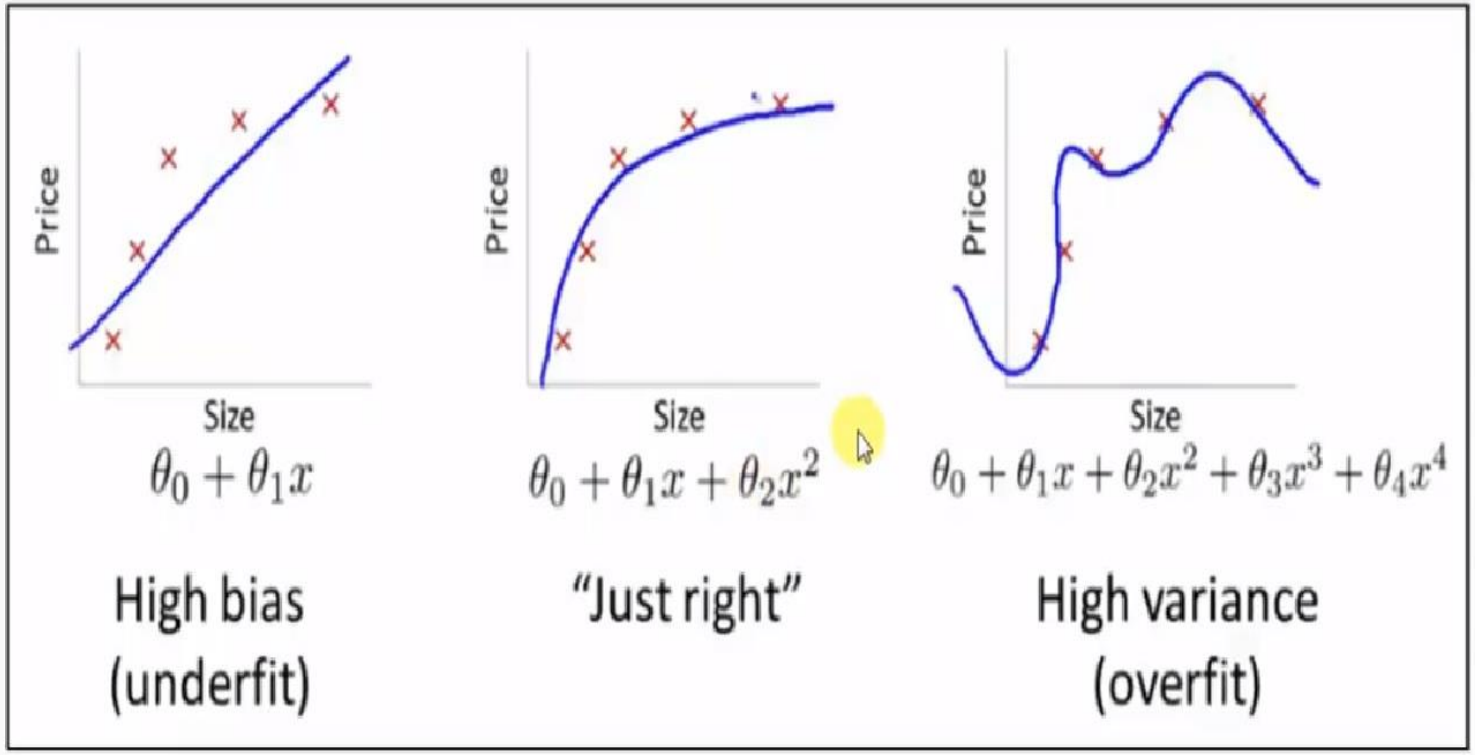


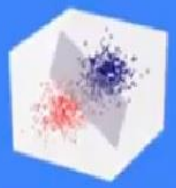
الضبط الزائد و الناقص UF , OF



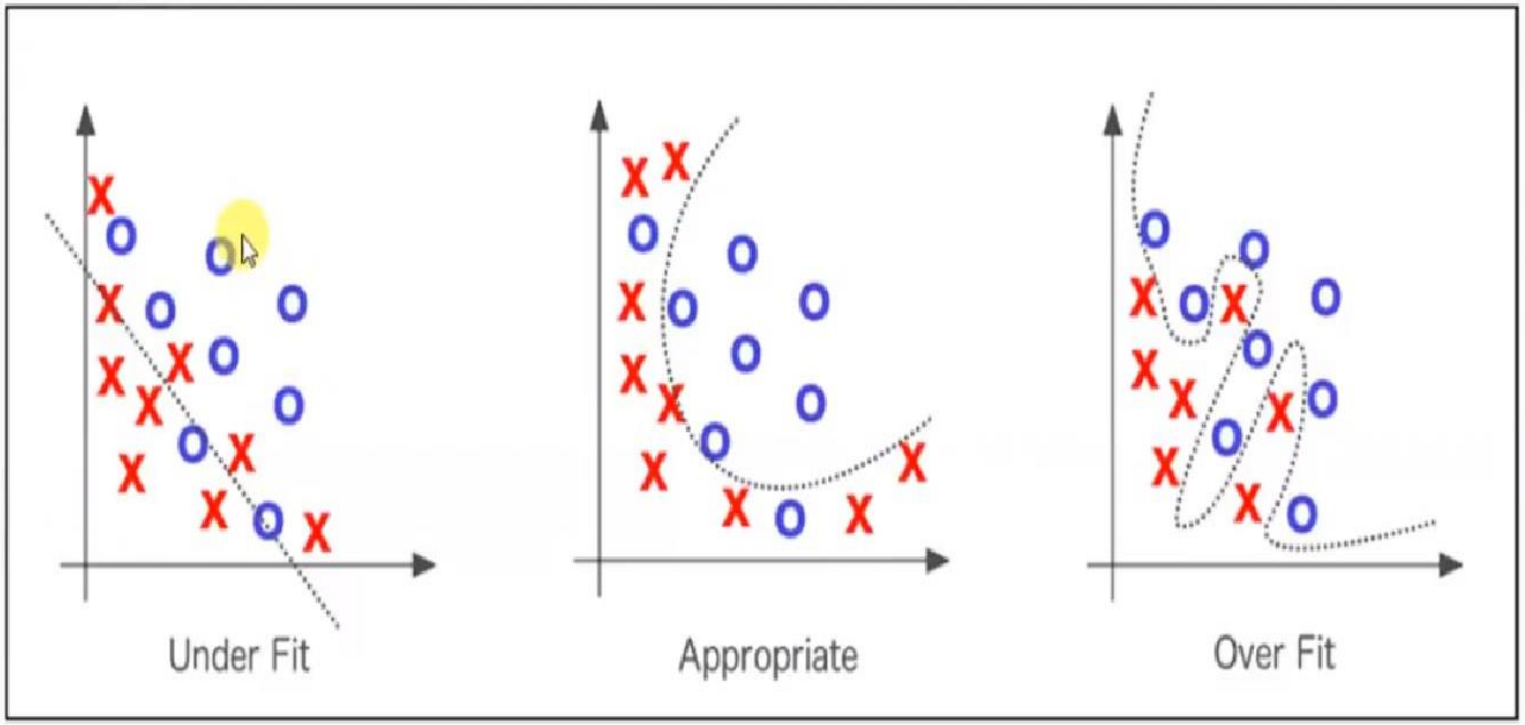


الضبط الزائد و الناقص UF , OF





الضبط الزائد و الناقص UF , OF



تنعيم البيانات Regularization

المفهوم :

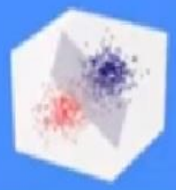
- عملية تنقية البيانات بهدف التخلص من الضبط الزائد OF

- يتم عبر 3 طرق هامة, قد نقوم باحداها فقط او باثنين فقط او بالثلاثة معا :
  - انتقاء البيانات
  - تغيير المعاملات
  - اضافة (المدا)

تنعيم البيانات Regularization

انتقاء البيانات :

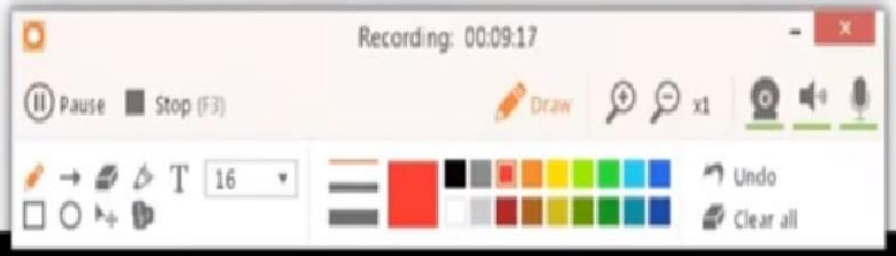
- يقصد بها انتقاء العناصر features التى يتم استخدامها فى اي خوارزم
- وجود عناصر غير مؤثرة فى الناتج المطلوب سيقوم بعمل تضليل mislead للخوارزم
- فى هذه الخطوة لابد من الاستعانة بمتخصصين فى المجال , ولن يعتمد هذا على معلوماتك الخاصة
- ما الذي يؤثر فى جودة ثمرة البطاطس (الطول , الوزن , اللون , نوع التربة , كمية الماء , السماد , الشمس , درجة الحرارة , الرياح , الشهر الذي تم زراعتها فيه . . .)
- نزع العناصر غير المؤثرة يؤدي لتطوير الخوارزم الخاص بك بشكل كبير , و دقة البيانات الخارجة منه
- كما أن اختيار العينات اساسا (الصفوف m) له دور مهم فى صحة البيانات ,فيجب ان تكون متجانسة و ممثلة لكل اجزاء المجموعة

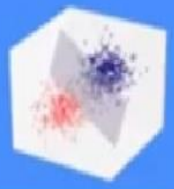


تنعيم البيانات Regularization

انتقاء البيانات :

- يقصد بها انتقاء العناصر features التى يتم استخدامها فى اي خوارزم
- وجود عناصر غير مؤثرة فى الناتج المطلوب سيقوم بعمل تضليل mislead للخوارزم
- فى هذه الخطوة لابد من الاستعانة بمتخصصين فى المجال , ولن يعتمد هذا على معلوماتك الخاصة
- ما الذي يؤثر فى جودة ثمرة البطاطس (الطول , الوزن , اللون , نوع التربة , كمية الماء , السماد , الشمس , درجة الحرارة , الرياح , الشهر الذي تم زراعتها فيه . . .)
- نزع العناصر غير المؤثرة يؤدي لتطوير الخوارزم الخاص بك بشكل كبير , و دقة البيانات الخارجة منه
- كما أن اختيار العينات اساسا (الصفوف m) له دور مهم فى صحة البيانات ,فيجب ان تكون متجانسة و ممثلة لكل اجزاء المجموعة





تنعيم البيانات Regularization

تغيير المعاملات :

- يقصد بها تغيير معاملات العناصر التى تؤثر فى المخرج لكن ليس بشكل كبير
- فمثلا درجة حرارة المدينة التى يعيش فيها الطالب , تؤثر بشكل ما فى تفوقه و درجاته , ولكن بمقدار بسيط
- فلو كانت لدي معادلة بها اربع عوامل مؤثرة هى إكس 1 و 2 و 3 و 4 كالتالى :

$$\theta_0 + \theta_1 x + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

- فنرغب فى تقليل معاملات (ثيتا) للعناصر غير المهمة (اكس 3 و 4) و بقاء معاملات العناصر الهامة (اكس 1 و 2) كما هى

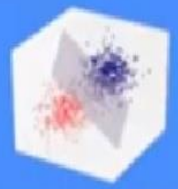
تنعيم البيانات Regularization

تغيير المعاملات :

- ويتم هذا عبر اضافة ارقام ضخمة للمعاملات المطلوب تقليلها لمعادلة التكلفة Cost Function

$$min_\theta \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + 1000 \cdot \theta_3^2 + 1000 \cdot \theta_4^2$$

- إضافة أرقام كبيرة جانب الثيتات المطلوب تقليلها , سيجبر المعادلة لتقليل قيمها جدا ليكون حاصل ضربهم رقم معقول



تنعيم البيانات Regularization

إضافة لمدا λ :

- واحيانا يسمى معامل التنعيم Regularization factor
- ونقوم بضربه فى مجموع مربعات جميع الثيتات (باستثناء ثيتا صفر) , ثم إضافته لمعادلة الكوست

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

- باختيار قيمة معينة للمدا , و بدأ ادخال المعادلة المحدثة فى الخوارزم , نري أن الـ OF سيختفى
- اذا ظل الـ OF موجود , نغير قليلا من قيمة لمدا و نجرب مرة اخري

تنعيم البيانات Regularization

إضافة لمدا λ :

- الصيغة المعدلة لمعادلة الكوست ستكون

Repeat {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha \left[\left(\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \right] \qquad j \in \{1, 2 \ldots n\}$$

}

- وهى جزئين لان الثيتا صفر لا يتم ضربها فى اللمدا , والثانية لباقى الثيتات
- يتم تكرار العملية للوصول لافضل قيمة للثيتات , ويتم تغيير قيمة اللمدا اذا ما ظل الـ OF موجود