بسم الله الرحمن الرحيم

Linear regression

أول حاجه هنرسم صوره كدا بتوضح احنا بنعمل ايه في ال linear regression ... عندك داتا سيت .. شوية نقط .. و بتحاول تشوف أحسن خط يحاول يجيب أحسن فيت للنقط ديه .. احنا هنا هندور علي خط optimal .. عاوزين ن formalize اللي بيحصل الاول .. الداتا نفسها بتتكون من شوية tuples .. إنبوت فيتشرز و تارجت أوبتوت .. ال x لما بتبقا bold ده عباره عن فيكتور أوف فيتشرز .. زي مثلاً عدد أوض النوم عدد المطابخ عدد الحمامات ... التارجت هيبقا numerical value .. يعني مثلا زي ال house prices .. أي حاجه bold هي فيكتور ... أي حاجه مش bold هي سكيلر عادي ... التارجت هنا اسمو تارجت عشان هو ال ground truth

احنا هدفنا اننا نلاقي hypothesis h اللي تقدر ت x map ل t ..

Problem

- Data: $\{(\boldsymbol{x_1}, t_1), (\boldsymbol{x_2}, t_2), \dots, (\boldsymbol{x_N}, t_N)\}$
  - $\boldsymbol{x} = < x_1, x_2, \dots, x_M >$: input vector
  - t: target (continuous value)
- Problem: find hypothesis h that maps $\boldsymbol{x}$ to t
  - Assume that h is linear:

$$y(\boldsymbol{x}, \boldsymbol{w}) = w_0 + w_1 x_1 + \cdots + w_M x_M = \boldsymbol{w^T} \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}$$

column vector

- Objective: minimize some loss function
  - Euclidean loss: $L_2(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}(y(\boldsymbol{x_n}, \boldsymbol{w}) - t_n)^2$

prediction

دلوقت احنا عاوزين نحل المشكله عن طريق optimization .. هو بيتكلم علي convex optimization .. ايه هي ال convex .. هي اللي بتحدد هل المشكله سهله ولا لا .. non convex ديه بتبقا صعبه ...

Optimization

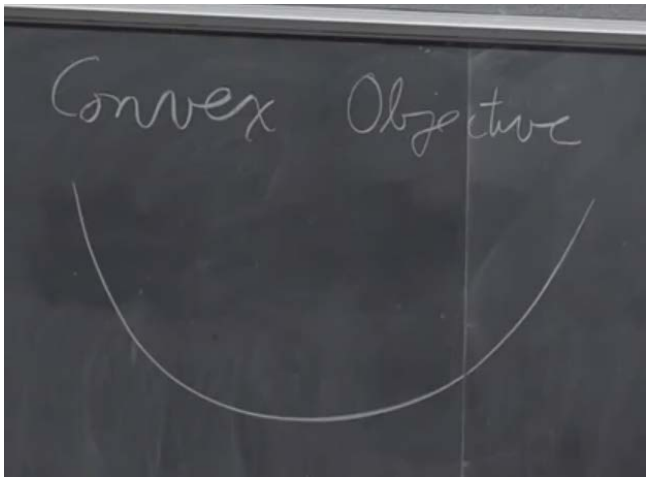
- Find best w that minimizes Euclidean loss

$$\boldsymbol{w}^* = argmin_{\boldsymbol{w}} \frac{1}{2}\sum_{n=1}^{N}\left(t_n - \boldsymbol{w^T}\begin{pmatrix} 1 \\ \boldsymbol{x_n} \end{pmatrix}\right)^2$$

- Convex optimization problem

  $\Longrightarrow$ unique optimum (global)

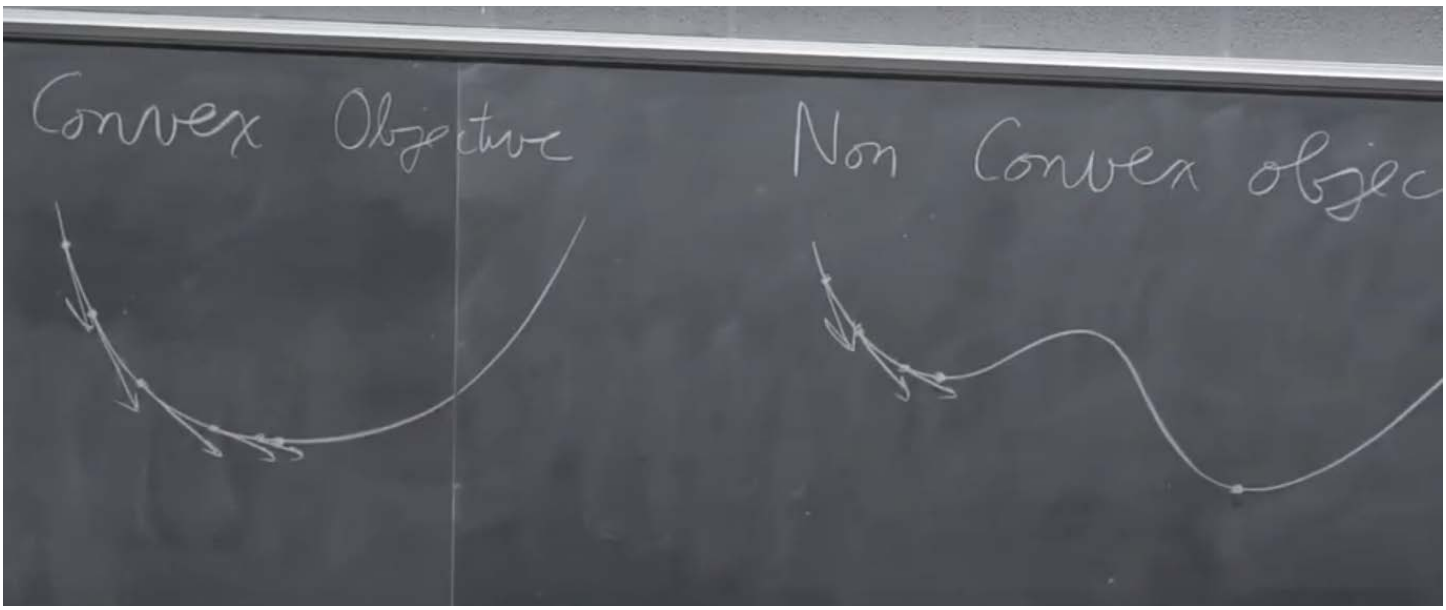
الدكتور رسم علي الصبوره بالطباشير .. هه . رسم convex objective .. وهو عاوز يجيب ال minimum of this function .. وبيقولك الموضوع سهل عشان انت عندك نقطه واحده منيمم .. و منين ما تبدأ في الكيرف بتاعك هتعرف تنزل تحت عكس ال gradient .. و توصل للمنيمم .. جامد ..



حاجه مثلاً gradient descent .. بتنزل علي الكيرف لحد ما توصل ..

هنيجي نبص علي Non convex objective

آدي كيرف عندو 2 منيمم ... واحده لوكال منيمام و التانيه جلوبال منيمام .. المشكله انك هتوصل للوكال منيمم تمام .. بس ال gradient = 0 وانا مش عارف اتحرك خلاص ... فبالتالي gradient decsent مش هتعرف ت converge علي الجلوبال منيمام ..الا يعني لو بدأت من نقطه معينه



لل non convex optimization problems معندناش reliable techniqurs عشان نقدر ن converge بسرعه يعني ..

المشكله اللي احنا بنعتبرها في المحاضره هي convex .. و الهدف بتاعنا هو quadratic objective .. زي الصوره اللي علي اليمين فوق ... نرجع تاني للسلايدز ...

الدكتور قال اننا ممكن نلاقي ال gradient و ننزل خطوات وهكذا .. في حل احسن .. ان تقدر تحسب ال gradient عادي .. بس في شرط ان ال gradient = 0 انت وقفت خلاص غير كدا افضل اتحرك .. وهي ديه النقطه الوحيده اللي هي جلوبال منيمم ... احنا عاوزين نلاقي ال W اللي بت satisify the equation .. اللي فيها ال Summation .. خد بالك ان ال W هي اللي مش عارفنها .. فهي ال variable .. احنا هنعوز نفجيب بقا ال W .. فقالك هتجيبها ازاي ..

$$Aw = b$$

Solution

- Let $\bar{\boldsymbol{x}} = \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}$ then $\min_{\boldsymbol{w}} \frac{1}{2} \sum_{n=1}^{N} (t_n - \boldsymbol{w}^T \bar{\boldsymbol{x}}_n)^2$
- Find $\boldsymbol{w}^*$ by setting the derivative to 0

$$\frac{\partial L_2}{\partial_{w_j}} = \sum_{n=1}^{N} (t_n - \boldsymbol{w}^T \bar{\boldsymbol{x}}_n) \bar{x}_{nj} = 0 \quad \forall j$$

$$\Longrightarrow \sum_{n=1}^{N} (t_n - \boldsymbol{w}^T \bar{\boldsymbol{x}}_n) \bar{\boldsymbol{x}}_n = 0$$

- This is a linear system in $\boldsymbol{w}$, therefore we rewrite it as $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}$

  where $\boldsymbol{A} = \sum_{n=1}^{N} \bar{\boldsymbol{x}}_n \bar{\boldsymbol{x}}_n^T$ and $\boldsymbol{b} = \sum_{n=1}^{N} t_n \bar{\boldsymbol{x}}_n$

Solution

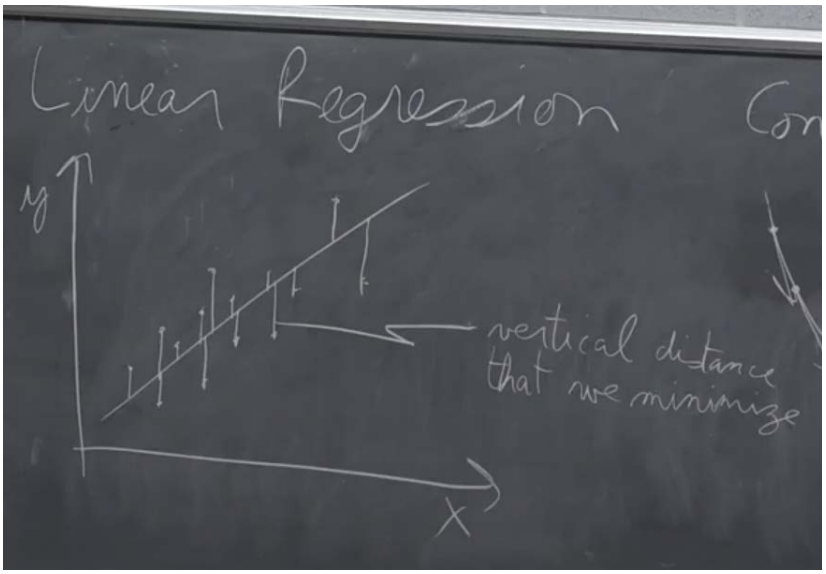
- If training instances span $\Re^{M+1}$ then $\boldsymbol{A}$ is invertible:

$$\boldsymbol{w} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

- In practice it is faster to solve the linear system $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}$ directly instead of inverting $\boldsymbol{A}$
  - Gaussian elimination
  - Conjugate gradient
  - Iterative methods

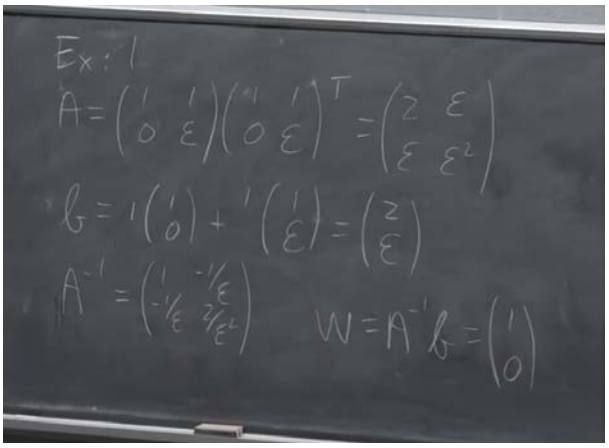
الدكتور هيرسم ع الصبوره :

احنا مثلا لاقينا ال line بتاع ال linear regression اللي محتاجينو .. لما تلاقي الخط اللي انت محتاجو .. احنا هدفنا نقلل ال Euclidean distance with respect to y .. فاحنا هدفنا نقلل ال vertical distance ما بين ال lines .. المهم خد بالك ان كل نقطه هتعوز تشد الخط في ناحيه .. ففي الآخر كلو هي stabilize عند ال minimum شد في الخط

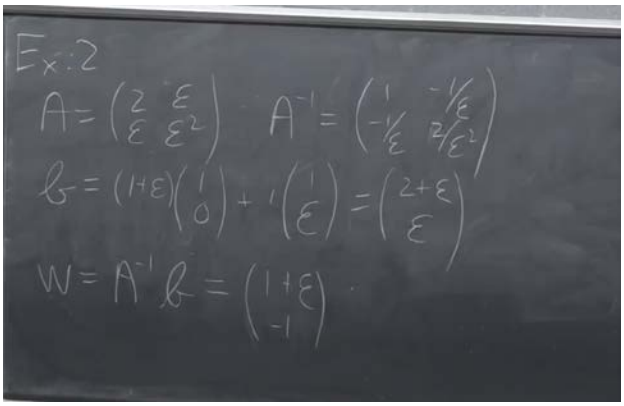


لما بنلاقي الحل بناءً علي ال least square .. الحل مش بيبقي robust .. عشان ممكن يحصل perturbation للانبوت .. فده هيغير ال line هيبقا فين .. و انت عاوز تشوف لو حصل تغيير بسيط .. الخط هيتغير تغيير بسيط .. بس انت ممكن تلاقي ان لو حصل تغيير في الانبوت .. هيحصل تغيير كبير جداً في الخط .. فعشان كدا هو unstable .. فده كدا overfitting .. ومشكلة ال least square انو too sensitive to noise ... هنعمل مثال بسيط:

Ex:1هنستخدم ال solution سلايد اللي في السلايدز بتاعت الدكتور



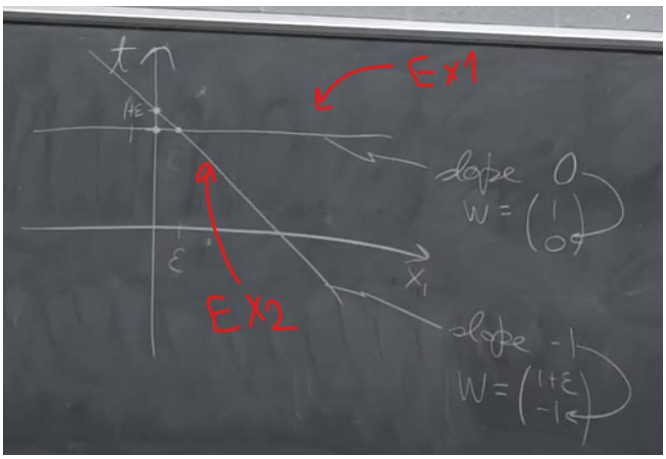
تعال نشوف تاني مثال .. الفرق هنا اننا حطينا تارجت 1+epsilon بدل ما كان 1 .. تغيير بسيط .. تعال نشوف التأثير عامل ازاي ..



قيم ال W اتغيرت .. اول رقم بدل 1 بقي 1+epsilon .. انما تاني رقم راح من 0 ل 1- .. تغيير بسيط .. هوب قام ضاربلك الماجنتيود ... فديه مشكله مينفعش انها تحصل .. الدكتور هيرسم صوره تانيه توضح ليه ده بيحصل .. اول حاجه هنرسم x1 علي ال x axis و ال y axis هو التارجت .. خد ال epsilon مع ال 1 و الصفر مع ال 1+epsilon .. فهنرسم خط لل solution ... في الحل الي اترسم ده انت عندك slope قيمتو 1- .. فبالتالي الحل

W = (1+epsilon

-1)



ازاي نتعامل مع المشكله ديه .. هو ال regularization .. هنحط penalty term هيحاول ي constrain نوع الحل اللي احنا ناخدو في الاعتبار .. الهدف الاساسي ان كنا ن minimize the Euclidean distance ما بين الأوتبوت وبين التارجت .. بس دلوقت كمان احنا هنقلل ال magnitude بتاع ال weight ... ماهو لو انت عندك weights كبيره مقارنة بأرقام تانيه .. كدا انت رايح ف داهيه .. فكرة ال regularization هتوضح المحاضره الجايه... فمتستعجلش

Regularization

- Idea: favor smaller values
- Tikhonov regularization: add $||\boldsymbol{w}||_2^2$ as a penalty term
- Ridge regression:

$$\boldsymbol{w}^* = argmin_{\boldsymbol{w}} \frac{1}{2}\sum_{n=1}^{N} \left(t_n - \boldsymbol{w}^T \bar{\boldsymbol{x}}_{\boldsymbol{n}}\right)^2 + \frac{\lambda}{2}||\boldsymbol{w}||_2^2$$

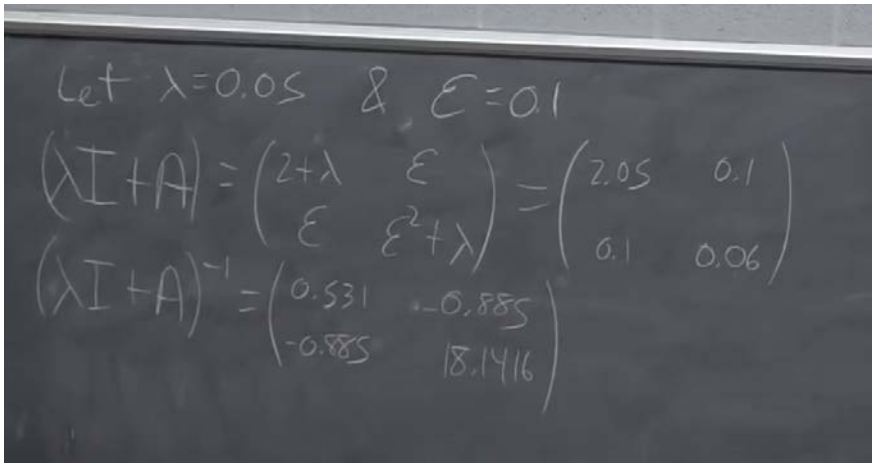
where λ is a weight to adjust the importance of the penalty

||W||^2 .. السكوير بتاع ال Euclidian norm .. واللمضه ديه هايبر بارمتر تعملو أوبتيمايزيشن .. ايه اللي هيحصل لما نضيف ال penalty term ... السستم لسه linear .. بس ال solution اتغير ..

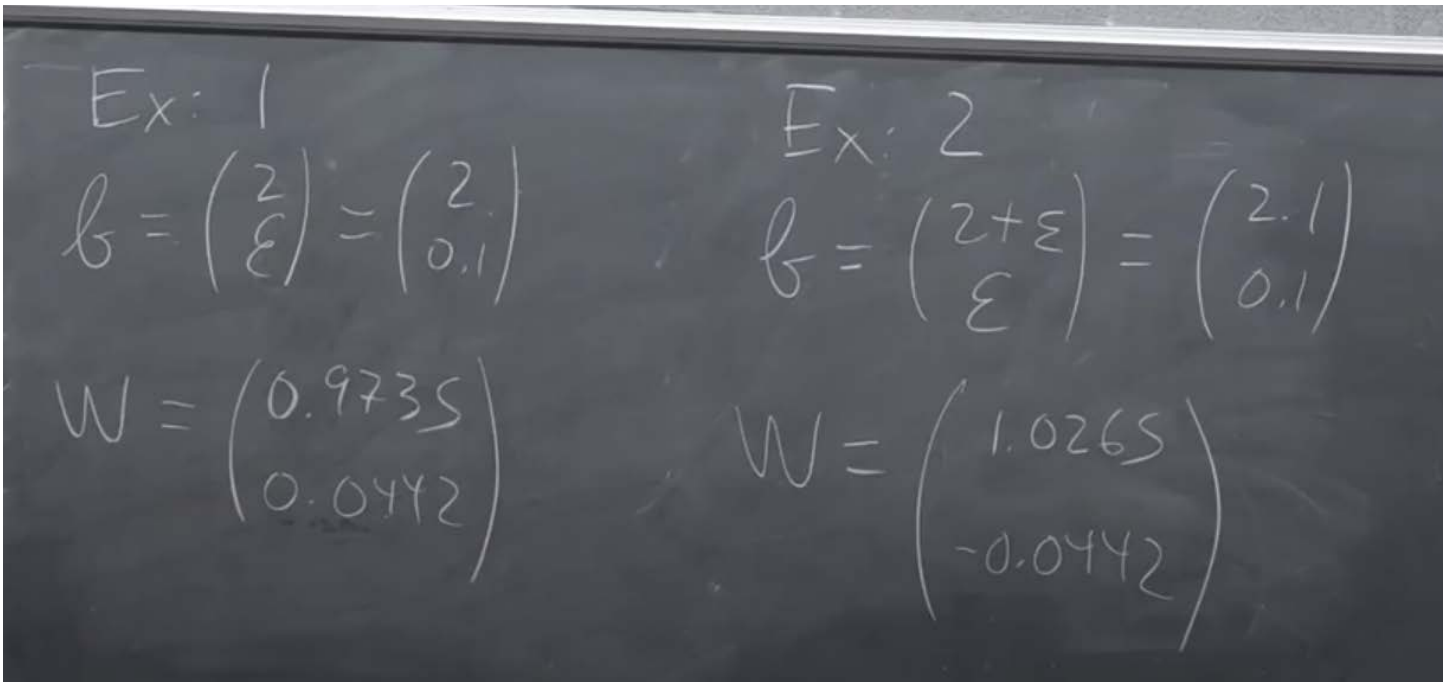
Regularization

- Solution: $(\lambda \boldsymbol{I} + \boldsymbol{A})\boldsymbol{w} = \boldsymbol{b}$
- Notes
  - Without regularization: eigenvalues of linear system may be arbitrarily close to 0 and the inverse may have arbitrarily large eigenvalues.
  - With Tikhonov regularization, eigenvalues of linear system are $\geq \lambda$ and therefore bounded away from 0. Similarly, eigenvalues of inverse are bounded above by $1/\lambda$.

الدكتور رسم ع الصبوره بالطباشير تاني ...



لو رجعنا تاني لمثال رقم 1 و 2



كدا الارقام بقت قريبه من بعضها