

# Lec3 - Overfitting management

Monday, 17 June 2024 16:10

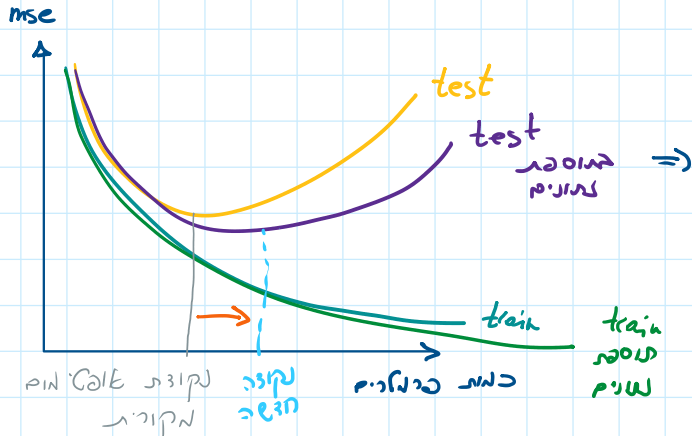
למה: התאוצות עם גודל  $\lambda$   $\rightarrow$  overfitting

רקע: רוב המודלים מוצגים הם לכוון  $\rightarrow$  כוונת המודל

overfitting  $\leftarrow$  מספר גדול יחסית של

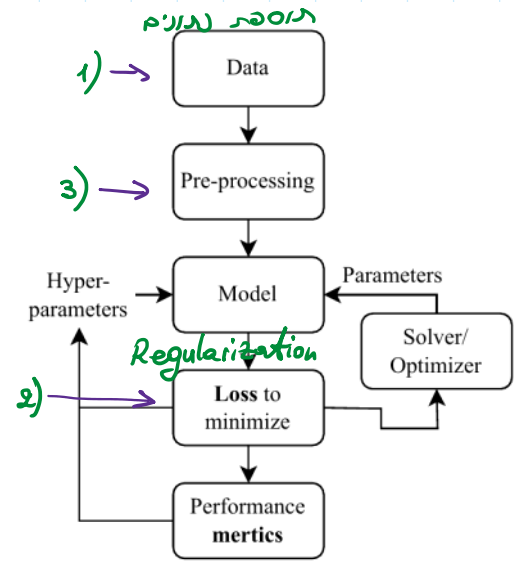
פרמטרים או נתונים (מספרים) (גודל של  $\mu$ )

פתרון מס' 1: הוספת נתונים



נצטרף יותר  
מא/יצי  
של פרמטרים  
ייעוצים

חסכון משמעותי:  
נתונים הם יקרים  
על מאוד.



## Regularization פתרון מס' 2

הגדרה Regularization: Penalty to the loss function

$$\mathcal{L}_{\text{new}} = \mathcal{L} + \lambda g(\mathbf{w})$$

$\lambda$  is termed regularization parameter.  
hyper-parameter

$L_1$ -regularization: Special case of  $g(\cdot)$ , where

$$g(\mathbf{w}) = \frac{\lambda}{2M} \|\mathbf{w}\|_1 = \frac{\lambda}{2M} \sum_{i=1}^N |w_i|$$

$L_2$ -regularization: Special case of  $g(\mathbf{w}) = \frac{1}{2M} \|\mathbf{w}\|^2$ ,

$$\mathcal{L}_{\text{new}} = \mathcal{L} + \lambda \mathbf{w}^T \mathbf{w} \sum_{i=1}^N w_i^2 \quad (4.3)$$

הסרה מודל ענאי/פולנומל  
לא כולל  $w_0$  (משקל = bias)

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2M} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 + \frac{\lambda}{2M} \underbrace{\|\mathbf{w}\|^2}_{\sum_{i=1}^N w_i^2} \quad (4.4)$$

הרצאה קודמת

Ridge Regression = חיפוי ענאי +  $L_2$ -reg.

מצאת משקל  $\mathbf{w}$

$$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}) = \frac{1}{M} (-\mathbf{X}^T (\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \mathbf{w}) = 0$$

$$\mathbf{w} = (\lambda \mathbf{I} + \mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

overfitting  $\lambda = 0$   $\rightarrow$  תוצאה שבה  $\lambda = 0$   $\rightarrow$  תוצאה שבה  $\lambda = 0$   $\rightarrow$  תוצאה שבה  $\lambda = 0$

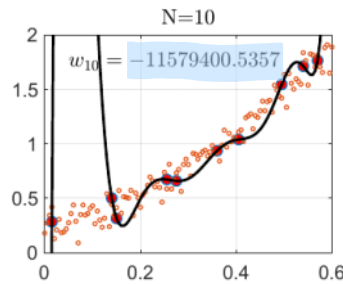
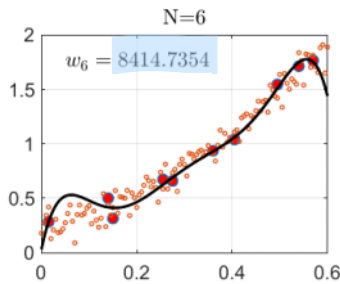
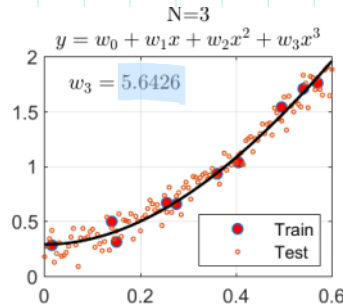
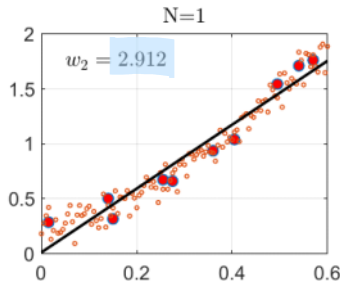
תוספת  
של reg.

תוספת  
reg.

overfitting  $\lambda = 0$  תוצאה שהיך לקצוות  
underfitting  $\lambda \rightarrow \infty$

0 <  $\lambda$  < 3 ערכים אחרים

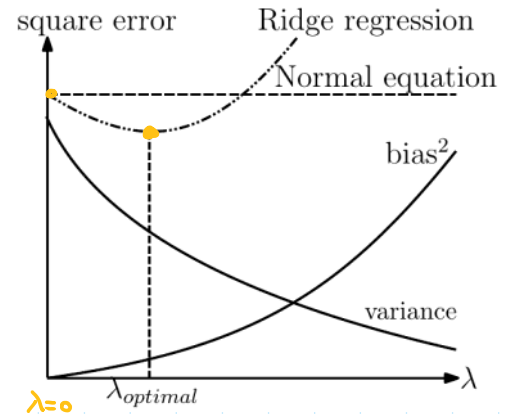
poly. תוצאה



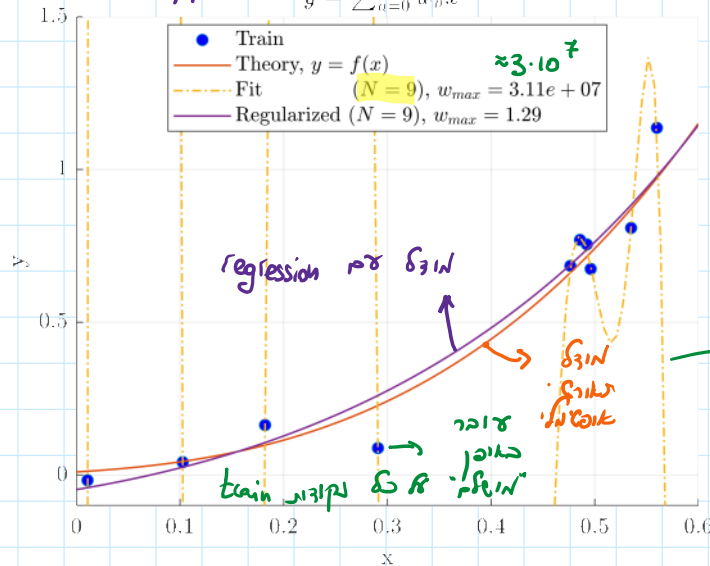
השגיאה:

תוצאה של

reg.  $\leftarrow$  שגיאה יותר נמוכה  
 $\leftarrow$  השגיאה בחומר  
למשוואה של ביטויים



צמצום מספרים: מונע overfit עם מונע  
והקטן ע"י שיטת פונקציה



הבעיה:  
\* סדר כולל  $\leftarrow$  היפר-כר  
של מונע (מספר שלם)

\* regularized  $\leftarrow$  היפר-כר  
של פונקציה מתוך (מספר רצף)  
מונע מקורי  
overfit

ערכים מקדמים של מונעים

השגיאה  
א יכולה להשפיע בצורה מפורטת (או פחות להשפיע)  
על המונע

\* יציבות מספרים: חישובים עם מספרים גשויים  
של כמה סדר גודל יכול להשפיע על ציור

רקע:  
"בכח מציב"  
שבו מספרים בעצמם  
מסומנים 'היו הרבה'  
יותר בצדדים/קטנים מה  
השאר

## Standardization

שיטה נורמליזציה

מאפיינים:  $0 =$  ממוצע

סטית תקן = שונות = 1

עבור כל המאפיינים  $X$

עבור כל המאפיינים

0 =  $z_{std}$  של הממוצע

1 = שונות

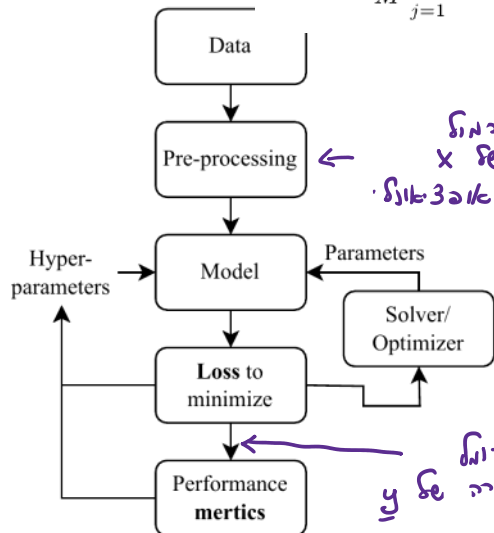
$$z_{std} = \frac{x - \bar{x}}{\sigma_x}$$

לממוצע

$$\bar{x} = \frac{1}{M} \sum_{j=1}^M x_j$$

שונות

$$\sigma_x^2 = \frac{1}{M} \sum_{j=1}^M (x_j - \bar{x})^2$$



Implementation steps:

1. On **train** dataset, evaluate  $\bar{x}$  and  $\sigma_x$ .
2. Apply normalization on **train** dataset, using  $\bar{x}$  and  $\sigma_x$ .
3. Apply normalization on **test** dataset, using **same**  $\bar{x}$  and  $\sigma_x$  (no recalculation).

למאפיינים

לא מחשבים ערכים מחדש

הערך: When normalization is applied to  $y$ , the output of the model is transformed back,  $\hat{y} = \hat{y}_{std} \sigma_y + \bar{y}$ .

חישוב ע"ש שני טבלאות

<https://www.mathworks.com/help/releases/R2024a/stats/zscore.html>

**Normalization:** Mapping all values of a feature to be in the range  $[0, 1]$  by the transformation

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \in [0, 1] \quad (4.9)$$

Implementation steps for normalization are similar to standardization.

הערך:

normalization שונה  
standardization  
כאשר  $0$  והממוצע

train נתונים למודל  
באופן