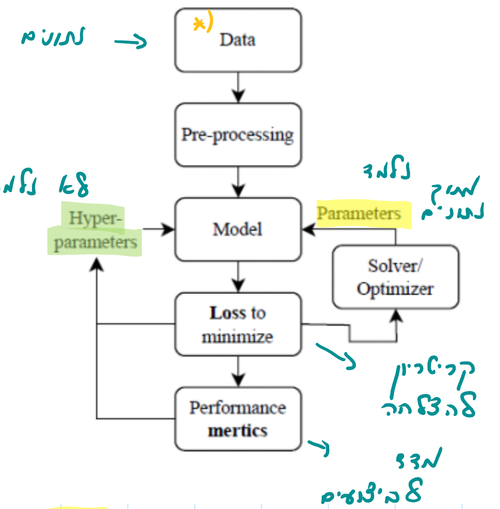
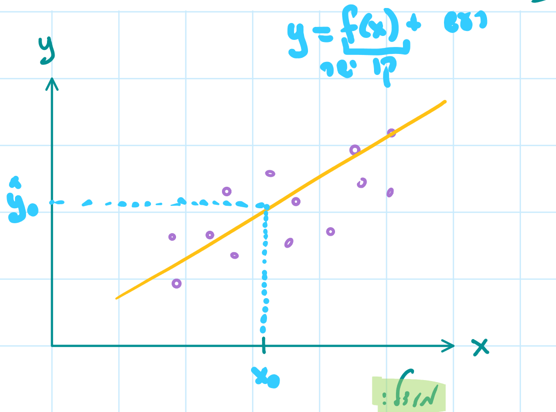
Least-squares and Linear Regression

Goal:
- The goal of the least squares (LS) method is to minimize MSE (or RMSE) between the given data and the parametric model.
- Define and analyze a model that is based on a linear relation between data and the outcome.
- Find the linear model parameters by LS.





נתונים:

M points (or measurements).

$\{x_k, y_k\}_{k=1}^{M}$ dataset

מודל:

$$y = f(x; w_0, w_1) = w_0 + w_1 x$$

משקלים weight

פרמטרים

חיזוי ע"י מודל:

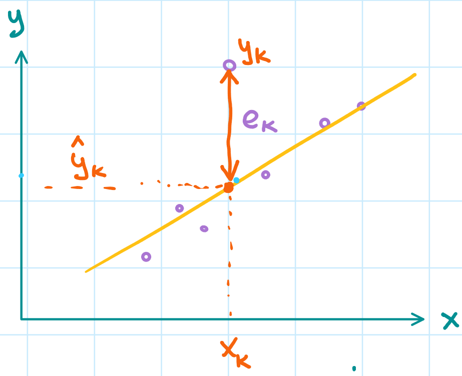
$$\hat{y}_k = f(x_k; w_0, w_1) = w_0 + w_1 x_k$$

מטרה: למצוא פרמטרים w_0, w_1 הכי מתאימים לנתונים

The performance metric is mean-square error (MSE) that is given by

שגיאה ריבועית ממוצעת

$$J_{mse}(w_0, w_1) = \frac{1}{M}\sum_{k=1}^{M}(y_k - \hat{y}_k)^2 = \frac{1}{M}\sum_{k=1}^{M} e_k^2 \quad \text{; sum of squared errors} \tag{2.3}$$



or root-MSE (RMSE)

$$J_{rmse}(w_0, w_1) = \sqrt{J_{mse}(w_0, w_1)}. \tag{2.4}$$

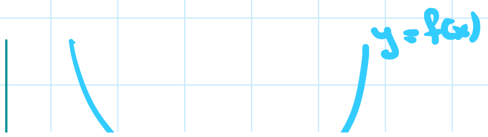
$$= \sqrt{\frac{1}{M}\sum_{k=1}^{M}(y_k - \hat{y}_k)^2}$$

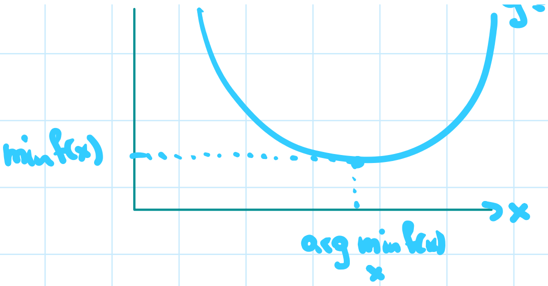
Loss/cost function

$$\mathcal{L}(w_0, w_1) = \sum_{k=1}^{M}(y_k - \hat{y}_k)^2 = \sum_{k=1}^{M}(y_k - w_0 - w_1 x_k)^2$$

$$w_0, w_1 = \arg\min_{w_0, w_1} J_{mse}(w_0, w_1) = \arg\min_{w_0, w_1} J_{rmse}(w_0, w_1) = \arg\min_{w_0, w_1} \mathcal{L}(w_0, w_1)$$

- $\min f(x)$ returns the minimum value of $f(x)$ for all possible values of x
- $\arg\min_x f(x)$ return the value of x, such that if $y = \arg\min_x f(x)$ then $\min f(x) = f(y)$





* חישוב מקדמי w_0, w_1

$$\begin{cases} \dfrac{\partial}{\partial w_0}\mathcal{L}(w_0,w_1)=0 \\ \dfrac{\partial}{\partial w_1}\mathcal{L}(w_0,w_1)=0 \end{cases}$$

$$\begin{cases} w_0 M + w_1 \sum_{k=1}^{M} x_k = \sum_{k=1}^{M} y_k \\ w_0 \sum_{k=1}^{M} x_k + w_1 \sum_{k=1}^{M} x_k^2 = \sum_{k=1}^{M} x_k y_k \end{cases}$$

$$\begin{cases} 2\sum_{k=1}^{M}(y_k-w_0-w_1x_k)\cdot(-1)=0 \\ 2\sum_{k=1}^{M}(y_k-w_0-w_1x_k)\cdot(-x_k)=0 \end{cases}$$

$$\sum_{k=1}^{M}(y_k-w_0-w_1x_k)^2$$

$$\frac{d}{dx} f(g(x)) = f'(g(x))\,g'(x)$$

$f(x) = x^2$ $\quad f'(x) = 2x$

$g(x) = (\ldots)$ $\quad g'(x) = -1 \quad w_0$

$= -x_k \quad w_1$

Vector/Matrix Notation

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}, \quad \mathbf{1}_M = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^M, \ \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \tag{2.14}$$

$\hat{y}_1 = w_0 + w_1 x_1$

$$\underbrace{\begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_M \end{bmatrix}}_{M\times 1} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ \vdots & \\ 1 & x_M \end{bmatrix}}_{X,\ M\times 2} \underbrace{\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}}_{2\times 1}$$

ממדי מטריצה

2) עמודות — N colums

1) שורות — M rows

$$\begin{matrix} \square & \square \\ \square & \square \\ \square & \square \end{matrix} \in \mathcal{R}^{M\times N}$$

$$\hat{\mathbf{y}} = f(\mathbf{X};\mathbf{w}) = \mathbf{1}_M w_0 + \mathbf{x} w_1 = \mathbf{X}\mathbf{w}$$

Loss

$$\mathcal{L}(\mathbf{w}) = (\mathbf{y}-\hat{\mathbf{y}})^T(\mathbf{y}-\hat{\mathbf{y}}) = \|\mathbf{y}-\hat{\mathbf{y}}\|^2$$
$$= (\mathbf{y}-\mathbf{X}\mathbf{w})^T(\mathbf{y}-\mathbf{X}\mathbf{w}) = \|\mathbf{y}-\mathbf{X}\mathbf{w}\|^2$$

$$\|\mathbf{z}\|^2 = z_1^2 + z_2^2 + \cdots + z_N^2$$

$$\nabla_{\mathbf{w}}\mathcal{L}(\cdot) = -\mathbf{X}^T(\mathbf{y}-\mathbf{X}\mathbf{w}) = 0$$

$$\mathbf{X}^T(\mathbf{y}-\mathbf{X}\mathbf{w}) = \mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}\mathbf{w} = 0$$

$$\mathbf{X}^T\mathbf{y} = \left(\mathbf{X}^T\mathbf{X}\right)\mathbf{w}$$

$$\boxed{\underbrace{\mathbf{w}_{opt}}_{2\times 1} = \underbrace{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}_{2\times M\ M\times 2 \ \to\ 2\times 2} \underbrace{\mathbf{X}^T\mathbf{y}}_{2\times M\ M\times 1}}$$

Multivariate LS

N עמודות מונים $\quad$ למונים

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \underbrace{\overbrace{\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_N}}_{\text{עמודות}} \end{bmatrix} \in \mathbb{R}^{M\times(N+1)}$$

דרישה:

* נתונים: $(\mathbf{X}, \mathbf{y})$

מטריצה X היא full rank

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_1 & \cdots & \mathbf{x}_N \end{bmatrix} \in \mathbb{R}^{M\times(N+1)}$$

עמוד

עמודות נוספות

$$\underline{\omega} = \begin{bmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_N \end{bmatrix}$$

פתרון: ללא שינוי

$$\boxed{\mathbf{w}_{opt} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}}$$

Projection matrix *

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$
$$= \mathbf{X}\mathbf{X}^+\mathbf{y} = \mathbf{P}\mathbf{y}$$

$$\mathbf{P} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$$

- Symmetric $\mathbf{P} = \mathbf{P}^T$,
- Idempotent $\mathbf{P} = \mathbf{P}^2$,
- Orthogonality, $\mathbf{P} \perp (\mathbf{I} - \mathbf{P})$
  Proof. $\mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{P} - \mathbf{P}^2 = \mathbf{0}$.
- $\mathbf{I} - \mathbf{P}$ is also projection matrix.

שגיאת חזוי

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y} = (\mathbf{I} - \mathbf{P})\,\mathbf{y}$$

Average error

$$\bar{e} = \frac{1}{M}\sum_{k=1}^{M} e_k$$
$$= \sum_{k=1}^{M} e_k = \mathbf{1}^T\mathbf{e} = 0$$

$$\nabla_{\omega} L = \mathbf{X}^T \underbrace{(\mathbf{y} - \mathbf{X}\mathbf{w})}_{\mathbf{e}} = \underline{0}$$

$$\begin{bmatrix} \mathbf{1}^T \\ \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} \mathbf{e} = \underline{0}$$

$$\mathbf{1}^T\mathbf{e} = 0$$

* שימוש: $(\mathbf{X}, \mathbf{y})$

מניחים $\bar{X}$ היא full rank

← אין עמודות תלויות לינארית

$$x = A^{-1}b \;\cancel{\Leftarrow}\; Ax = b \quad *$$

אף פעם לא מחשבים ישירות

Moore–Penrose inverse (pseudo-inverse) *

$$\mathbf{X}^+ = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$$

$$\mathbf{X}^+\mathbf{X} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{I}$$

הרחבה של מטריצה הופכה עבור מטריצות לא מלבניות

$$\mathbf{w}_{opt} = \mathbf{X}^+\mathbf{y}$$

Error and data orthogonality *

$$\mathbf{e} \perp \mathbf{X} \Rightarrow \underbrace{\mathbf{X}^T}_{(N+1)\times M}\underbrace{\mathbf{e}}_{M\times 1} = \underline{\mathbf{0}}^{(N+1)\times 1}$$

Proof:

$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T\mathbf{y} - \underbrace{\left[\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]}_{I}\mathbf{X}^T\mathbf{y}$$
$$\mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{y} = \mathbf{0}$$

Error and prediction orthogonality

$$\mathbf{e} \perp \hat{\mathbf{y}} \Rightarrow \underbrace{\hat{\mathbf{y}}^T}_{1\times M}\underbrace{\mathbf{e}}_{M\times 1} = \mathbf{e}^T\hat{\mathbf{y}} = 0^{1\times 1}$$

Proof:

$$\hat{\mathbf{y}}^T\mathbf{e} = \mathbf{y}^T\mathbf{P}\,(\mathbf{I} - \mathbf{P})\,\mathbf{y}$$
$$= \mathbf{y}^T\mathbf{P}\mathbf{y} - \mathbf{y}^T\mathbf{P}\mathbf{P}\mathbf{y}$$
$$= \mathbf{y}^T\mathbf{P}\mathbf{y} - \mathbf{y}^T\mathbf{P}\mathbf{y} = 0$$

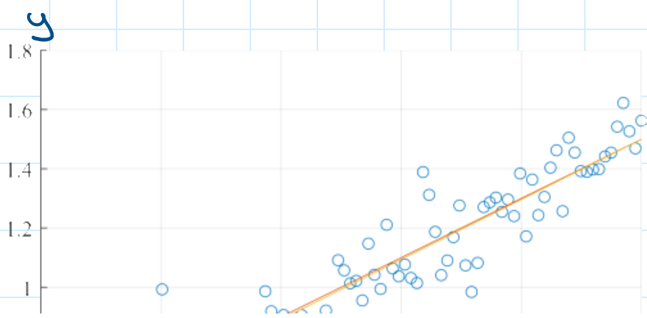
MSE

$$mse_{min} = \sum_{k=1}^{M} y_k^2 - \sum_{j=0}^{N} w_j \underbrace{\underbrace{\mathbf{y}^T}_{1\times M}\underbrace{\mathbf{x}_j}_{M\times 1}}_{1\times 1}$$

Matlab

ידועים מראש

```
w0_theory = 0.5;
w1_theory = 1;

x1 = linspace(0,1)';
y_theory = w0_theory + w1_theory*x1;

sigma = 0.1;
y = y_theory + sigma*randn(length(x1),1);
```



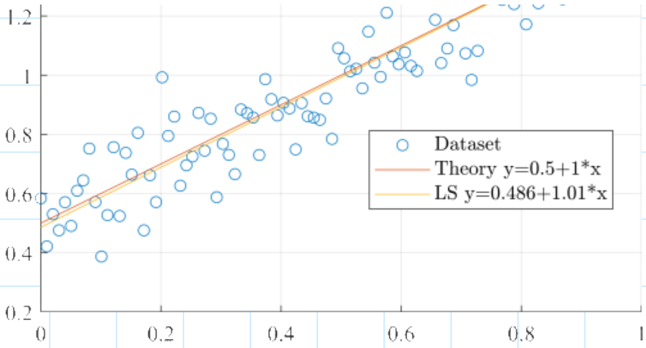
```
y_theory = w0_theory + w1_theory*x1;

sigma = 0.1;
y = y_theory + sigma*randn(length(x1),1);

M = length(y);
X = [ones(M,1) x1];

w_ls = pinv(X)*y;
y_hat = X*w_ls;
```

$y = f(x) + err$ (AWGN)

$M = 100$

$\mathbf{w}_{opt} = \mathbf{X}^{+}\mathbf{y}$

$\hat{\mathbf{y}} = f(\mathbf{X};\mathbf{w}) = \mathbf{1}_M w_0 + \mathbf{x} w_1 = \mathbf{X}\mathbf{w}$



x

$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$

$\mathbf{e} \perp \mathbf{X} \Rightarrow \mathbf{X}^T\mathbf{e} = \mathbf{0}$

$\mathbf{e}^T\hat{\mathbf{y}} = 0$

```
e = y - y_hat;
X'*e
mean(e)
y_hat'*e
```

$\sim 10^{-16}$

2.1 This assignment focuses on understanding the interpretation of weights values. For the multivariate vector of the weights, $\mathbf{w}$:

- What is the meaning of the $+$ or $-$ sign of the each weight w_j?
- What is the influence of the magnitude (relative size) of weights w_j?

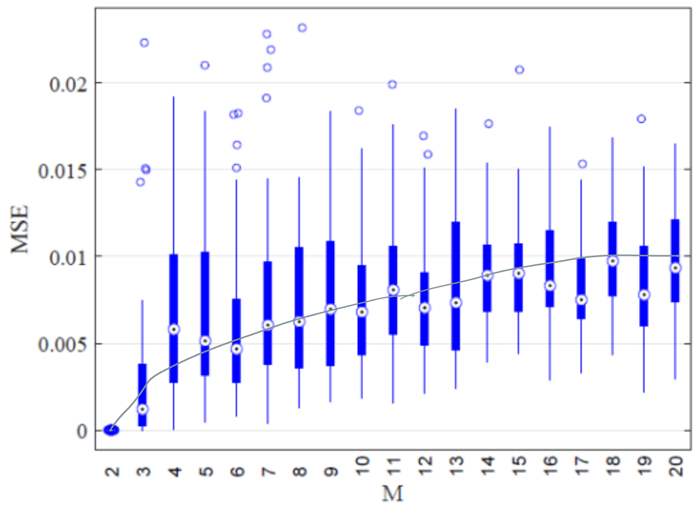
תרגיל בית

2.2 This assignment focuses on understanding the effects of sample size on the estimation accuracy of linear regression parameters using the Least Squares (LS) method. The task involves running simulations to generate linear data with varying numbers of data points, followed by fitting a linear regression model to this data and analyzing the resulting mean squared error (MSE).

- Perform linear regression using the LS method, and analyze the MSE across different sample sizes. גדלי M
- Repeat each MSE evaluation for at least 30 times.
- Summarize the results in `boxplot` as in the plot.
- Does it seem reasonable the the MSE grows with an increase in M?

σ = קבוע

M משתנה



These assignments will help you understand the practical implications of linear regression and the influence of data size on model performance.