CSCI3230 (ESTR3108)
Fundamentals of Artificial Intelligence

Tutorial 1

CHEN, Yueyao

Email: yychen@cse.cuhk.edu.hk
Office: Room 1024, 10/F, SHB

Dept. of Computer Science & Engineering
The Chinese University of Hong Kong

Outline

Part 1. Review linear algebra

Part 2. Least squares

Part 1. Review linear algebra

Vector

- Notation for vectors in $\mathbb{R}^n$:

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{column vector}$$

$$X' = (x_1, \cdots, x_n) \quad \text{row vector}$$

- Transpose: column vector $\leftrightarrow$ row vector

$$X' = X^T, \quad X'^T = X$$

- In the context of vectors, column vectors and row vectors are the same. But when putting them with matrices, they are totally different.

Vectors

- Scaling: $a \in \mathbb{R}$, $X = (x_1, \ldots, x_n) \in \mathbb{R}$, $aX = (ax_1, \ldots, ax_n)$
- Addition: $X = (x_1, \ldots, x_n)$, $Y = (y_1, \ldots, y_n)$,
  $X + Y = (x_1 + y_1, \ldots, x_n + y_n)$

- Suppose we have m vectors: $X_1 \ldots X_m$, linear combination of these vectors: for any m scalars $a_1 \ldots a_m$,

$$a_1 X_1 + \cdots + a_m X_m$$

- Inner product (dot product): $\forall X \in \mathbb{R}^n, Y \in \mathbb{R}^n$

$$X \cdot Y = X^T Y = \sum_{i=1}^{n} x_i y_i$$

- Euclidean norm: $\|X\|_2 = \sqrt{X^T X} = \sqrt{\sum_{i=1}^{n} x_i^2}$

Matrix

- Notation for $m \times n$ matrices:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

  $a_{i,j}$ is the element at the i-th row, j-th column of $\mathbf{A}$.

- Denote the i-th column of $\mathbf{A}$ as $\mathbf{A}_i$ (column vectors), and j-th row of $\mathbf{A}$ as $\mathbf{A}^{(j)}$ (row vectors).
- Transpose: columns $\leftrightarrow$ rows

$$\mathbf{A}^T = \begin{pmatrix} a_{1,1} & \cdots & a_{m,1} \\ \vdots & \ddots & \vdots \\ a_{1,n} & \cdots & a_{m,n} \end{pmatrix} = [{A^{(1)}}^T, \ldots, {A^{(m)}}^T]$$

Matrix

- n-dimensional row vector: $1 \times n$ matrix
- n-dimensional column vector: $n \times 1$ matrix
- Scaling: $c \in \mathbb{R}$, $\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$, $c\mathbf{A} = \begin{pmatrix} ca_{1,1} & \cdots & ca_{1,n} \\ \vdots & \ddots & \vdots \\ ca_{m,1} & \cdots & ca_{m,n} \end{pmatrix}$
- Addition: $\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix}$

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{1,1} + b_{1,1} & \cdots & a_{1,n} + b_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} + b_{m,1} & \cdots & a_{m,n} + b_{m,n} \end{pmatrix}$$

$\mathbf{A}$ and $\mathbf{B}$ have the same shape: $\mathbb{R}^{m \times n}$

Special matrices

- A square matrix (in $\mathbb{R}^{n\times n}$) in which every element except the principal diagonal elements is zero is called a Diagonal Matrix :

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda_n \end{pmatrix}$$

- Identity matrix in $\mathbb{R}^{n\times n}$: $I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$
- Symmetric matrix: $A = A^T$

Matrix-vector multiplication

- Matrix-vector multiplication

$$
\begin{aligned}
\mathbf{A}X &= \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{m,1} \end{pmatrix} x_1 + \begin{pmatrix} a_{1,2} \\ \vdots \\ a_{m,2} \end{pmatrix} x_2 + \cdots \begin{pmatrix} a_{1,n} \\ \vdots \\ a_{m,n} \end{pmatrix} x_n \\
&= \mathbf{A}_1 x_1 + \cdots + \mathbf{A}_n x_n
\end{aligned}
$$

- Notice the shape: $\mathbf{A} \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^n$, $\mathbf{A}X \in \mathbb{R}^m$.
- Matrix-vector multiplication $=$ linear combination of $\mathbf{A}$'s columns.

Matrix-vector multiplication

Another perspective of matrix-vector multiplication:

- Matrix-vector multiplication

$$
\begin{aligned}
\mathbf{A}X &= \begin{pmatrix} \mathbf{A}^{(1)} \\ \vdots \\ \mathbf{A}^{(m)} \end{pmatrix} X \\
&= \begin{pmatrix} \mathbf{A}^{(1)} \cdot X \\ \vdots \\ \mathbf{A}^{(m)} \cdot X \end{pmatrix}
\end{aligned}
$$

  Recall that $\mathbf{A}^{(i)}$ is the i-th row of $\mathbf{A}$.

- Matrix-vector multiplication $=$ dot product of $\mathbf{A}$'s rows and X.

Matrix-matrix multiplication

- Matrix-matrix multiplication

$$\mathbf{AX} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_{1,1} & \cdots & x_{1,l} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,l} \end{pmatrix}$$

- Notice the shape: $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{X} \in \mathbb{R}^{n \times l}$, $\mathbf{AX} \in \mathbb{R}^{m \times l}$.

- Let X_i be the i-th column of $\mathbf{X}$, the i-th column of $\mathbf{AX}$ is $\mathbf{A}X_i$:

$$\mathbf{AX} = [\mathbf{A}X_1, \ldots, \mathbf{A}X_l]$$

Inverse matrix

- For squared matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, if there is a $\mathbf{B} \in \mathbb{R}^{n \times n}$ such that

$$\mathbf{AB} = I = \mathbf{BA}$$

  $\mathbf{A}$ is invertible and $\mathbf{A}^{-1} = \mathbf{B}$

- How to find inverse matrix of $\mathbf{A}$ (if invertible):
  1. Gauss-Jordan elimination: find a row-operating matrix $\mathbf{B}$ which transforms $\mathbf{A}$ to $\mathbf{I}$. (More feasible for human begings)
  2. Use eigen-decomposition: $\mathbf{A} = U\Lambda U^{-1}$, $\mathbf{A}^{-1} = U\Lambda^{-1}U^{-1}$
  3. Use the analytic solution:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathbf{C}^T$$

     where $\mathbf{C}$ is the adjugate matrix of $\mathbf{A}$.

Linear algebra materials

1. Introduction to Linear Algebra. Gilbert Strang.
2. Matrix Analysis and Applied Linear Algebra. Carl D. Meyer.
3. Advanced Linear Algebra. Steven Roman.
4. Linear Algebra and Its Applications. Manolis C. Tsakiris.

Part 2. Least squares

Problem settings

- Recall that our data matrix $\mathbf{X}$ and observed labels Y:

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & x_1^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \quad Y = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

- We use linear function to fit a linear mapping from $\mathbf{X}$ to Y:

$$\hat{Y} = \mathbf{X}\Theta$$

where $\Theta = \begin{pmatrix} \theta_{\mathbf{0}} \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$, such that $\hat{Y}$ is very close to Y.

Ordinary least squares

- Recall that we use $\|\hat{Y} - Y\|_2^2$ to measure the distance between real labels and estimated labels.
- Note that:

$$\|\hat{Y} - Y\|_2^2 = \sum_{i=0}^{m} (\hat{y}^{(i)} - y^{(i)})^2$$

  which is exactly the residual sum of squares (RSS).
- Ordinary least squares (OLS) estimator:

$$\hat{\Theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

  is the optimal solution that minimizes $\|\hat{Y} - Y\|_2$.

Derivation

Exercise1-1 - OLS

Assume we have a training set

Training set

Index	x	y
1	5.51	0.81
2	1.25	1.22
3	3.60	0.43
4	4.72	-0.51
5	3.91	-0.13
6	6.13	0.44
7	8.05	1.49
8	5.55	0.31
9	7.33	1.59
10	7.59	1.61

Let's try to find some linear models to fit the training data. We use the RSS objective function $J(\boldsymbol{\Theta}) = \left\|\hat{f}_{\boldsymbol{\Theta}}(\mathbf{X}) - \boldsymbol{Y}\right\|_2^2$.

★ Calculate the analytic solution of the linear model $\hat{f}_{\boldsymbol{\Theta}}(\mathrm{x}) = \theta_0 + \theta_1 \mathrm{x}$. Then, plot the line of your obtained linear model together with the data points in training set.

Exercise1-1 - OLS

Ordinary least squares (OLS) estimator:

$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

In this problem:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{10} \end{pmatrix} = \begin{pmatrix} 1 & 5.51 \\ 1 & 1.25 \\ 1 & 3.60 \\ 1 & 4.72 \\ 1 & 3.91 \\ 1 & 6.13 \\ 1 & 8.05 \\ 1 & 5.55 \\ 1 & 7.33 \\ 1 & 7.59 \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} 0.81 \\ 1.22 \\ 0.43 \\ -0.51 \\ -0.13 \\ 0.44 \\ 1.49 \\ 0.31 \\ 1.59 \\ 1.61 \end{pmatrix}$$

Exercise1-1 - OLS

Ordinary least squares (OLS) estimator:

$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

In the first step:

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 1+\cdots+1 & x_1+\cdots+x_{10} \\ x_1+\cdots+x_{10} & {x_1}^2+\cdots+{x_{10}}^2 \end{pmatrix} = \begin{pmatrix} 10 & 53.64 \\ 53.64 & 326.968 \end{pmatrix}$$

For the inverse matrix,

$$(\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 10 & 53.64 \\ 53.64 & 326.968 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

So we have,

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 0.833187235 & -0.136686658 \\ -0.136686658 & 0.025482226 \end{pmatrix}$$

Exercise1-1 - OLS

Ordinary least squares (OLS) estimator:

$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

In the second step:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \mathbf{1} & \cdots & \mathbf{1} \\ x_1 & \cdots & x_{10} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + a_{12}x_1 & \cdots & a_{11} + a_{12}x_{10} \\ a_{21} + a_{22}x_1 & \cdots & a_{21} + a_{22}x_{10} \end{pmatrix}$$

So we have,

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} 0.080044 & \cdots & -0.20426 \\ 0.003720 & \cdots & 0.056723 \end{pmatrix}$$

Exercise1-1 - OLS

Ordinary least squares (OLS) estimator:

$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

In the last step:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y = \begin{pmatrix} a_{11} + a_{12}x_1 & \cdots & a_{11} + a_{12}x_{10} \\ a_{21} + a_{22}x_1 & \cdots & a_{21} + a_{22}x_{10} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix}$$

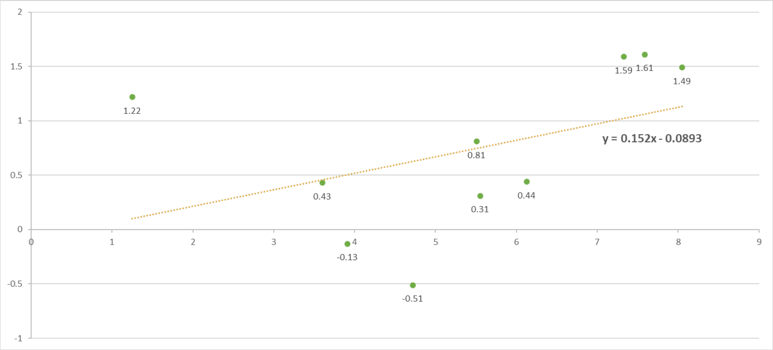
$$= \begin{pmatrix} (a_{11} + a_{12}x_1)y_1 + \cdots + (a_{11} + a_{12}x_{10})y_{10} \\ (a_{21} + a_{22}x_1)y_1 + \cdots + (a_{21} + a_{22}x_{10})y_{10} \end{pmatrix}$$

So we have,

$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y = \begin{pmatrix} -0.0893031 \\ 0.151995 \end{pmatrix}$$

Exercise1-1 - OLS

Plot the line of the linear model together with the data points.



Exercise1-2 - OLS

Assume we have a training set

Training set

Index	x	y
1	5.51	0.81
2	1.25	1.22
3	3.60	0.43
4	4.72	-0.51
5	3.91	-0.13
6	6.13	0.44
7	8.05	1.49
8	5.55	0.31
9	7.33	1.59
10	7.59	1.61

Let's try to find some linear models to fit the training data. We use the RSS objective function $J(\boldsymbol{\Theta}) = \left\| \hat{f}_{\boldsymbol{\Theta}}(\mathbf{X}) - \boldsymbol{Y} \right\|_2^2$.

★ Suppose we want to increase the model complexity, by considering y as a linear function of both x and x^2: $\hat{f}_{\boldsymbol{\Theta}}(\mathrm{x}) = \theta_0 + \theta_1 \mathrm{x} + \theta_2 \mathrm{x}^2$. In this case, calculate the analytic solution of the model and plot the curve of the model, together with the data points in training set.

Exercise1-2 - OLS

Ordinary least squares (OLS) estimator:

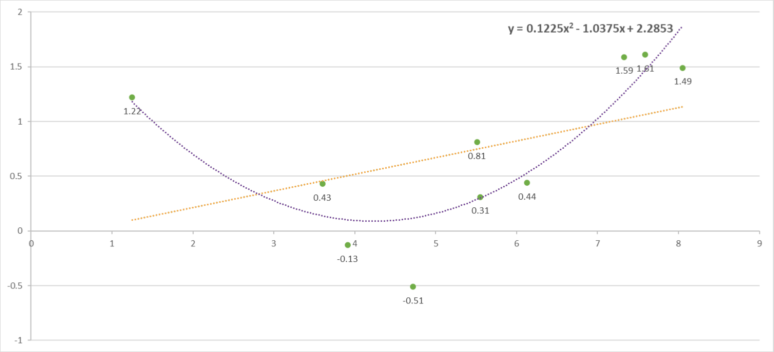
$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

In this problem:

$$\Theta = \begin{pmatrix} \theta_{\mathbf{0}} \\ \theta_1 \\ \theta_2 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{1} & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ \mathbf{1} & x_{10} & x_{10}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{1} & 5.51 & 5.51^2 \\ \mathbf{1} & 1.25 & 1.25^2 \\ \mathbf{1} & 3.60 & 3.60^2 \\ \mathbf{1} & 4.72 & 4.72^2 \\ \mathbf{1} & 3.91 & 3.91^2 \\ \mathbf{1} & 6.13 & 6.13^2 \\ \mathbf{1} & 8.05 & 8.05^2 \\ \mathbf{1} & 5.55 & 5.55^2 \\ \mathbf{1} & 7.33 & 7.33^2 \\ \mathbf{1} & 7.59 & 7.59^2 \end{pmatrix}$$

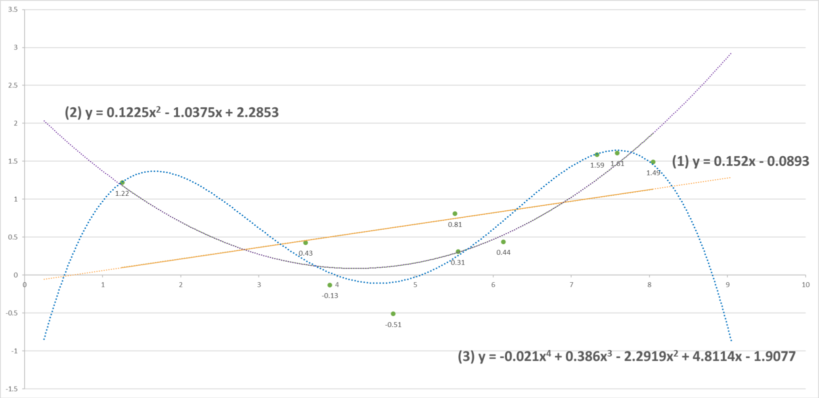
Exercise1-2 - OLS

$$\Theta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y = \begin{pmatrix} 2.28533 \\ -1.03753 \\ 0.122519 \end{pmatrix}$$



Exercise1-3 - OLS

Observe the three functions on the figure, please point out which could be faced with underfitting, which could be faced with overfitting, and which one is relatively a good one.

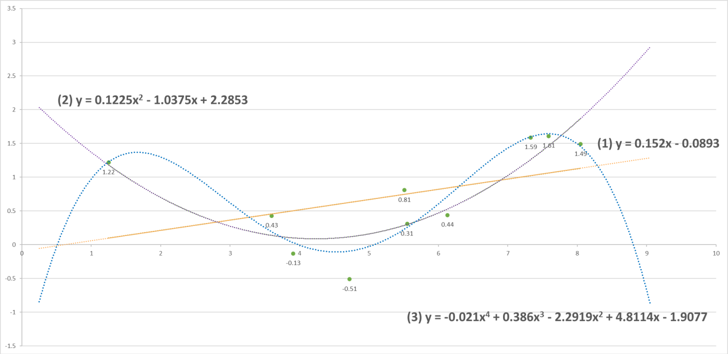


Exercise1-3 - OLS

Calculate the values of prediction error on the test data to verify your thoughts through the following data set:

$$(5.84, 0.89), (0.61, 1.79), (4.23, -0.15), (6.50, 1.63), (0.89, 1.27),$$

$$(3.75, 0.91), (5.73, 0.88), (3.10, 1.41), (6.47, 1.69), (4.59, -0.46).$$



Exercise1-3 - OLS

Calculate the values of prediction error on the test data to verify your thoughts through the following data set:

$$(5.84, 0.89), (0.61, 1.79), (4.23, -0.15), (6.50, 1.63), (0.89, 1.27),$$

$$(3.75, 0.91), (5.73, 0.88), (3.10, 1.41), (6.47, 1.69), (4.59, -0.46).$$

$$E_{mean} = \frac{1}{10}\sum_{i=1}^{10} \|\hat{f}_{\Theta}(x^{(i)}) - y^{(i)}\|_2^2$$

Model 1: $\mathrm{E}_{mean} = 0.8754$
Model 2: $\mathrm{E}_{mean} = 0.4714$
Model 3: $\mathrm{E}_{mean} = 0.5292$