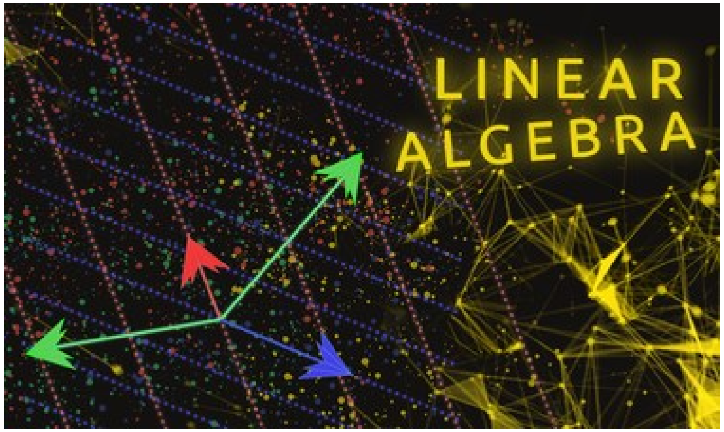
[redacted]



LINEAR ALGEBRA

Least Square Method

Nooshin Maghsoodi

NOSHIRVANI UNIVERSITY

QR factorization

If one has a matrix A with linearly independent columns, one can split it as

$$A = QR$$

where Q is an orthogonal matrix and R is upper-triangular.

How? *Apply the Gram-Schmidt algorithm to the columns of A!*

(the GS yields an orthogonal basis, but one can normalize it in the end)

Let $A = [\boldsymbol{a}_1 \vdots \boldsymbol{a}_2 \vdots \boldsymbol{a}_3]$. The Gram-Schmidt algorithm will yield orthonormal $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ such that

$$\mathrm{span}(\boldsymbol{a}_1) = \mathrm{span}(\boldsymbol{x}_1);$$

$$\mathrm{span}(\boldsymbol{a}_1, \boldsymbol{a}_2) = \mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2);$$

$$\mathrm{span}(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3) = \mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3).$$

(these equalities basically say that x_1 is constructed just from a_1, then x_2 is constructed only from a_1 and a_2, etc.)

$$\boldsymbol{a}_1 = r_{11}\boldsymbol{x}_1$$

$$\boldsymbol{a}_2 = r_{12}\boldsymbol{x}_1 + r_{22}\boldsymbol{x}_2$$

$$\boldsymbol{a}_3 = r_{13}\boldsymbol{x}_1 + r_{23}\boldsymbol{x}_2 + r_{33}\boldsymbol{x}_3$$

(this is just a consequence of the abstract fact of equality between spans)

or

$$\underbrace{[\boldsymbol{a}_1 \vdots \boldsymbol{a}_2 \vdots \boldsymbol{a}_3]}_{A} = \underbrace{[\boldsymbol{x}_1 \vdots \boldsymbol{x}_2 \vdots \boldsymbol{x}_3]}_{Q} \cdot \underbrace{\begin{bmatrix} \boldsymbol{r}_{11} & \boldsymbol{r}_{12} & \boldsymbol{r}_{13} \\ 0 & \boldsymbol{r}_{22} & \boldsymbol{r}_{23} \\ 0 & 0 & \boldsymbol{r}_{33} \end{bmatrix}}_{R}$$

QR factorization continued

Example. Compute the QR factorization of $A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix}$.

Solution. We need to compute two matrices: Q and R.

To compute Q, run the Gram-Schmidt algorithm for the columns of A:

(don't forget to normalize in the end, for we need an orthonormal basis, not merely orthogonal)

$$\boldsymbol{x}_1 := \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

$$\boldsymbol{x}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{0}{5}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\boldsymbol{x}_3 := \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} - \frac{0}{5}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{-1}{1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

Now normalize:

$$\begin{bmatrix} \frac{1}{\sqrt{5}} \\ 0 \\ \frac{2}{\sqrt{5}} \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -\frac{2}{\sqrt{5}} \\ 0 \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

QR factorization continued

Example. Compute the QR factorization of $A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix}$.

$$\Rightarrow \quad Q = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & \frac{-2}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{bmatrix}$$

The simplest way to find R is to solve for R in $A = QR$:

$$A = QR \Rightarrow Q^{-1}A = R \Rightarrow \underline{Q^T A = R}$$

($Q^{-1} = Q^T$ for orthogonal matrices)

$$R = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{-2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{5} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & \sqrt{5} \end{bmatrix}$$

Answer.

$$\underbrace{\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & \frac{-2}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \sqrt{5} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & \sqrt{5} \end{bmatrix}}_{R}$$

QR FACTORIZATION CONTINUED

The QR Factorization

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

دانشگاه صنعتی نوشیروانی بابل

LEAST SQUARE

- Sometimes, $Ax = b$ has no solution.
- When a solution is demanded and none exists, the best one can do is to find an x that makes Ax as close as possible to b.

residual is $r = Ax - b$

least squares problem: choose x to minimize $\|Ax - b\|^2$

$\|Ax - b\|^2$ is the *objective function*

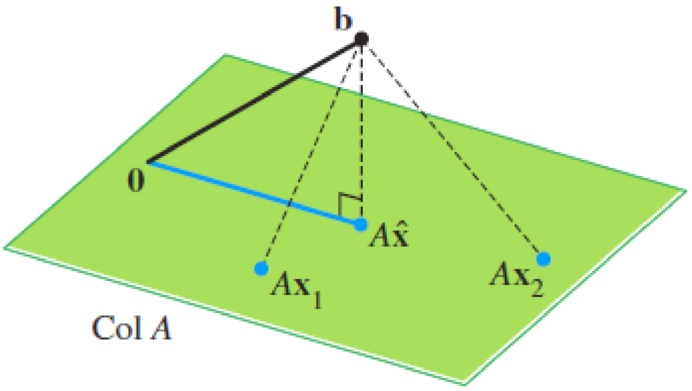
If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

دانشگاه صنعتی نوشیروانی بابل

LEAST SQUARE



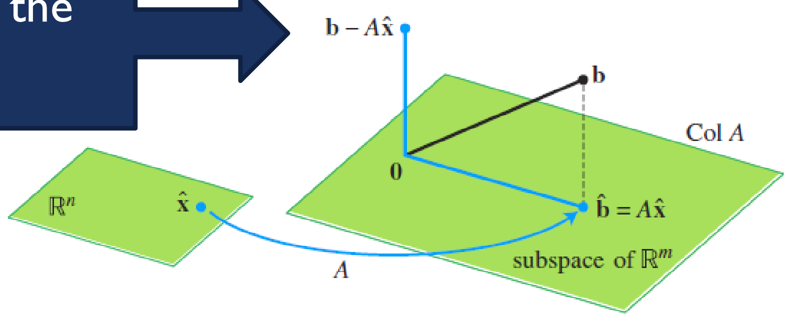
$$\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\, \mathbf{b}$$

Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

LEAST SQUARE

b-b^ is orthogonal to the ColA



$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$$

$$A^T\mathbf{b} - A^TA\hat{\mathbf{x}} = \mathbf{0}$$

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

EXAMPLE

Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

SOLUTION

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\Rightarrow \quad \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\Rightarrow \quad \begin{aligned} \hat{\mathbf{x}} &= (A^T A)^{-1} A^T \mathbf{b} \\ &= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned}$$

دانشگاه صنعتی نوشیروانی بابل

EXAMPLE2

Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

EXAMPLE2 (CONTINUE)

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Rightarrow \quad \hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

دانشگاه صنعتی نوشیروانی بابل

THEOREME

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix $A^T A$ is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} \tag{4}$$

دانشگاه صنعتی نوشیروانی بابل

AN ALTERNATIVE WAY OF COMPUTATION

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 12. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} \tag{6}$$

دانشگاه صنعتی نوشیروانی بابل

EXAMPLE

Find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

SOLUTION

$$A = QR = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

Then

$$Q^T\mathbf{b} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

The least-squares solution $\hat{\mathbf{x}}$ satisfies $R\mathbf{x} = Q^T\mathbf{b}$; that is,

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

This equation is solved easily and yields $\hat{\mathbf{x}} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}$.

دانشگاه صنعتی نوشیروانی بابل

SOME APPLICATIONS

- Least Square Data Fitting
- Least Square Classification

LEAST SQUARE DATA FITTING

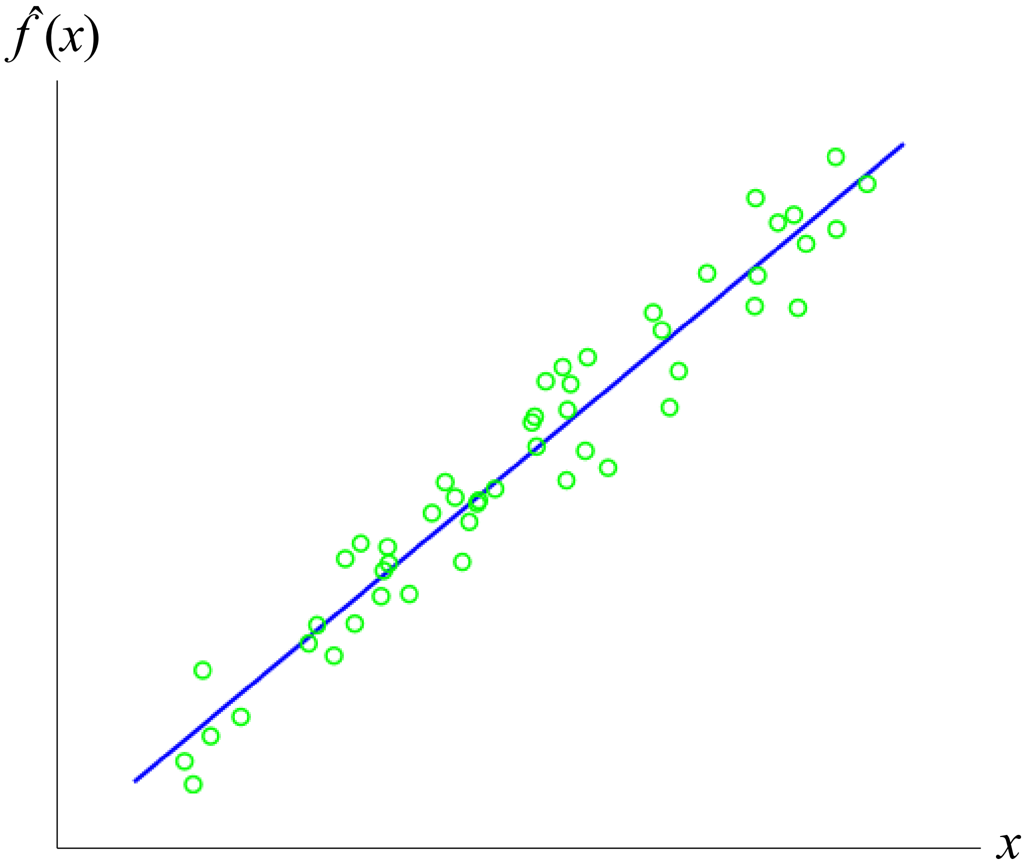
- we believe a scalar y and an n-vector x are related by *model*

$$y \approx f(x)$$

- x is called the *independent variable*
- y is called the *outcome* or *response variable*
- $f : \mathbf{R}^n \to \mathbf{R}$ gives the relation between x and y
- often x is a feature vector, and y is something we want to predict
- we don’t know f, which gives the ‘true’ relationship between x and y

دانشگاه صنعتی نوشیروانی بابل

Example



LEAST SQUARE DATA FITTING

- we are given some *data*

$$x^{(1)}, \ldots, x^{(N)}, \qquad y^{(1)}, \ldots, y^{(N)}$$

  also called *observations*, *examples*, *samples*, or *measurements*

- $x^{(i)},\ y^{(i)}$ is ith *data pair*
- $x_j^{(i)}$ is the jth component of ith data point $x^{(i)}$

دانشگاه صنعتی نوشیروانی بابل

LEAST SQUARE DATA FITTING

- choose *model* $\hat{f} : \mathbf{R}^n \to \mathbf{R}$, a *guess* or *approximation* of f
- *linear in the parameters* model form:

$$\hat{f}(x) = \theta_1 f_1(x) + \cdots + \theta_p f_p(x)$$

- $f_i : \mathbf{R}^n \to \mathbf{R}$ are *basis functions* that we choose
- θ_i are *model parameters* that we choose
- $\hat{y}^{(i)} = \hat{f}(x^{(i)})$ is (the model's) *prediction* of $y^{(i)}$
- we'd like $\hat{y}^{(i)} \approx y^{(i)}$, *i.e.*, model is consistent with observed data

LEAST SQUARE DATA FITTING

- *prediction error* or *residual* is $r_i = y^{(i)} - \hat{y}^{(i)}$
- *least squares data fitting*: choose model parameters θ_i to minimize RMS prediction error on data set

$$\left(\frac{(r^{(1)})^2 + \cdots + (r^{(N)})^2}{N}\right)^{1/2}$$

- this can be formulated (and solved) as a least squares problem

LEAST SQUARE DATA FITTING

- express $y^{(i)}$, $\hat{y}^{(i)}$, and $r^{(i)}$ as N-vectors
  - $y^{\mathrm{d}} = (y^{(1)}, \ldots, y^{(N)})$ is vector of outcomes
  - $\hat{y}^{\mathrm{d}} = (\hat{y}^{(1)}, \ldots, \hat{y}^{(N)})$ is vector of predictions
  - $r^{\mathrm{d}} = (r^{(1)}, \ldots, r^{(N)})$ is vector of residuals

- define $N \times p$ matrix A with elements $A_{ij} = f_j(x^{(i)})$, so $\hat{y}^{\mathrm{d}} = A\theta$

- least squares data fitting: choose θ to minimize

$$\|r^{\mathrm{d}}\|^2 = \|y^{\mathrm{d}} - \hat{y}^{\mathrm{d}}\|^2 = \|y^{\mathrm{d}} - A\theta\|^2 = \|A\theta - y^{\mathrm{d}}\|^2$$

- $\hat{\theta} = (A^T A)^{-1} A^T y$ (if columns of A are linearly independent)

- $\|A\hat{\theta} - y\|^2 / N$ is *minimum mean-square (fitting) error*

دانشگاه صنعتی نوشیروانی بابل

LEAST SQUARE CLASSIFICATION

- data fitting with outcome that takes on (non-numerical) values like
  - true or false
  - spam or not spam
  - dog, horse, or mouse
- outcome values are called *labels* or *categories*
- data fitting is called *classification*
- we start with case when there are two possible outcomes
- called *Boolean* or *2-way* classification
- we encode outcomes as $+1$ (true) and -1 (false)
- classifier has form $\hat{y} = \hat{f}(x)$, $f : \mathbf{R}^n \rightarrow \{-1, +1\}$

دانشگاه صنعتی نوشیروانی بابل