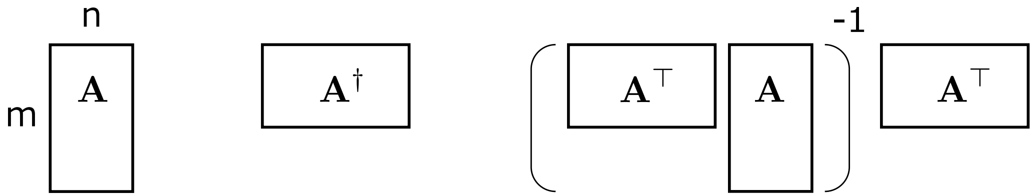
5. Least square method and line fitting

- Pseudoinverse
- Overdetermined system of linear equations
- Line fitting

Pseudoinverse (aka Moore-Penrose pseudoinverse or generalized inverse)

- Assuming that a $m \times n$ matrix A is a real matrix and $A^T A$ is invertible, the pseudoinverse $A^{\dagger}$ for matrix A is defined to be

$$\mathbf{A}^{\dagger} \equiv (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}$$



- The following always holds:

$$\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{I}$$

  - This is because:

$$\mathbf{A}^{\dagger}\mathbf{A} = ((\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top})\mathbf{A} = (\mathbf{A}^{\top}\mathbf{A})^{-1}(\mathbf{A}^{\top}\mathbf{A}) = \mathbf{I}$$

- Note that if $m \neq n$, the following always holds:

$$\mathbf{A}\mathbf{A}^{\dagger} \neq \mathbf{I}$$

Calculating a pseudoinverse

- Function pinv gives the pseudoinverse of a given matrix

```
>> A=randn(5,3)
A =
  -1.000354   0.027611   0.065035
  -3.013282  -0.687265  -0.462170
  -1.345817  -0.410357   1.915242
  -0.480726   0.027323   1.544261
  -0.512782   0.230256  -0.269629
>> pinv(A)
ans =
  -0.3005504  -0.1638335   0.0394693  -0.1490451  -0.3649408
   1.1103074  -0.5201691  -0.5397881   0.7475318   1.6065569
   0.0075412  -0.1606289   0.2720726   0.2571976  -0.0259860
```

- The left multiplication to A yields an identity matrix

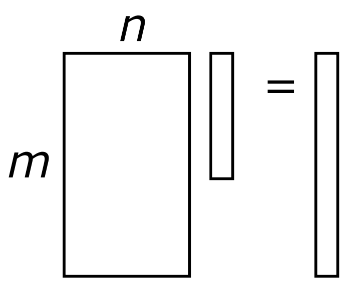
```
>> pinv(A)*A
ans =
   1.0000e+00   2.7756e-16  -1.5266e-16
  -5.5511e-16   1.0000e+00   7.2164e-16
   2.9490e-17   7.9797e-17   1.0000e+00
```

Remark: the right multiplication does not yield an identity

Overdetermined system of linear equations

- Consider a system of linear equations with a more number of equations than unknowns
  - A: m x n matrix (m>n)

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$



m>n → Called overdetermined

m<n → Called underdetermined

  - In general, an overdetermined system does not have a solution
- We calculate a “solution” as follows:

$$\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$$

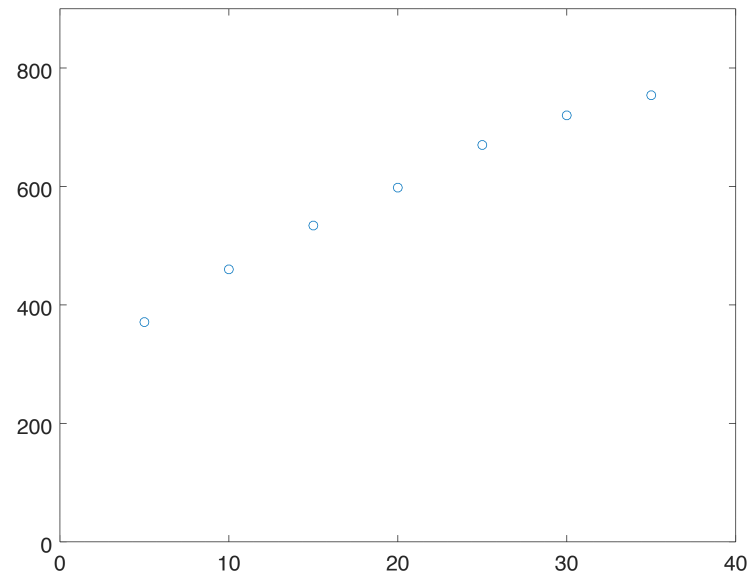
  - It can be shown that this solution x minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$
  - This solution is thus called the least square solution

Line fitting: an example

- Salary and years of service of employees in Japan

Years of service	<5	<10	<15	<20	<25	<30	<35
Salary (mil. JPY)	370.8	459.4	533.8	597.7	669.7	719.7	753.8

```
>> years=5:5:35
years =
     5    10    15    20    25    30    35
>> income=[371,460,534,598,670,720,754];
>> plot(years,income,”o”)
>> axis([0,40,0,900])
>> set(gca,”fontsize”,14)
```



Line fitting: least square method (1/2)

- Fit a line $y=ax+b$ to a set of points $\{(x_1,y_1), \cdots, (x_N,y_N)\}$ so that the difference in y axis will be small for each (x_i,y_i)

$$\varepsilon_i \equiv \|y_i - \hat{y}_i\| = \|y_i - (ax_i + b)\|$$

- To do so, find (a,b) that minimizes the sum of the differences for all the points

$$\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} \|y_i - (ax_i + b)\|^2$$

The right hand side can be rewritten as:

$$\sum_i \|y_i - (ax_i + b)\|^2 = \left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} ax_1 + b \\ ax_2 + b \\ \vdots \\ ax_N + b \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right\|^2$$

$$ax_i + b = \begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

Line fitting: least square method (2/2)

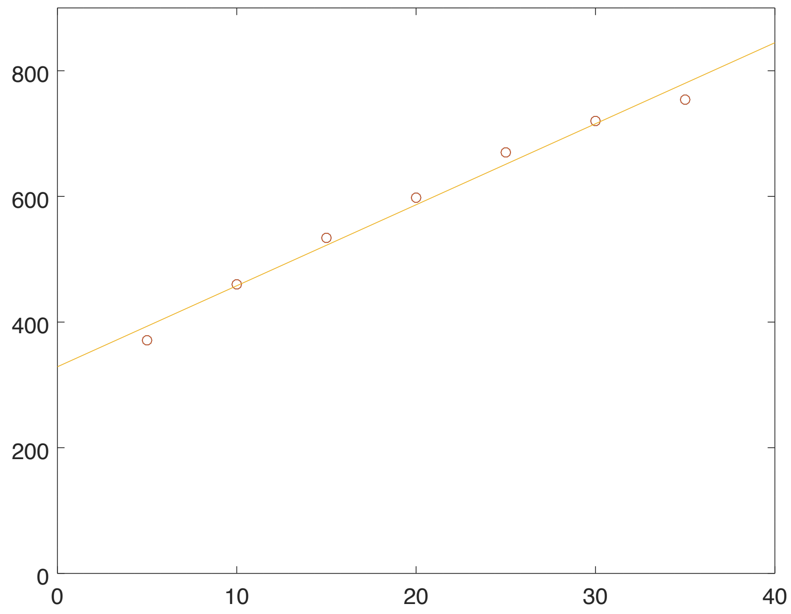
- Thus, the problem reduces to solution of a linear equation Xp=y

$$\|\mathbf{X}\mathbf{p}-\mathbf{y}\|^2 \rightarrow \min \qquad \mathbf{X} \equiv \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_N & 1 \end{bmatrix}, \quad \mathbf{p} \equiv \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

- Its solution (i.e., least square solution) is given using pseudoinverse X† as

$$\hat{\mathbf{p}} \equiv \mathbf{X}^{\dagger}\mathbf{y}$$

```
>> X=ones(7,2);
>> X(:,1)=years';
>> y=income';
>> p=pinv(X)*y;
>> hold on
>> xx=0:1:40;
>> plot(xx,p(1)*xx+p(2))
```



Exercises 5.1

- The table to the right shows the number of Nobel laureates per capita (i.e., divided by population) and chocolate consumption per capita for different countries
- It has been discovered that there is a strong link between these two cultural traits (Nobel laureates and chocolate consumption)
    - Franz H. Messerli, Chocolate Consumption, Cognitive Function, and Nobel Laureates, the New England Journal of Medicine, 367, 1562-1564, 2012
- Fit a line to the data and plot the results
    - You can download a file (‘Nobel_vs_choco.txt‘) from the course page

	Nobel laureates per capita	Chocolate consumption per capita (kg/y/head)
Sweden	31.855	6.6
Switzerland	31.544	10.8
Denmark	25.255	8.6
Austria	24.332	7.9
Norway	23.368	9.8
UK	18.875	10.3
Ireland	12.706	8.8
Germany	12.668	11.4
USA	10.706	5.1
Hungary	9.038	3.5
France	8.99	7.4
Belgium	8.622	6.8
Finland	7.6	7
Australia	5.451	6
Italy	3.265	3.3
Poland	3.124	4.5
Lithuania	2.836	6.1
Greece	1.857	4.5
Portugal	1.855	4.5
Spain	1.701	3.3
Japan	1.492	2.2
Bulgaria	1.421	2.2
Brazil	0.05	2.5

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3743834/