Lecture 7. Correlation coefficient

Correlation coefficient

- **PCA: Covariance matrix and correlation matrix**
- Scatterplot
- Three frameworks for correlation coefficient
    - Naïve approach
    - Regression: Correlation and determinacy; properties and meaning
    - Probability: Gaussian distribution
- Weird correlation case studies
- Different function and/or different criteria: Nature-inspired optimization
- Homework 6

Correlation Coefficient: Conventional approach to PCA

* =

Covariance matrix:

Given a N×V data matrix *X*, compute its centered version Y and the V×V feature covariance matrix B:

a. **Center matrix X by finding, for each feature, its mean and subtracting it from all the feature values, Y=X-m(X)**

b. **Compute square matrix A=Y′*Y and divide it by N or N-1 (do the latter if you think that the result is going to be used as an estimate of the covariance matrix of a multivariate density function, I rather divide by N):**

   B= Y′*Y/N.

(*v*,*w*) entry in B: $b_{vw} = \frac{1}{N}\sum_{i=1}^{N}(x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w})$

*   =

Correlation Coefficient: Conventional approach to PCA, 2

Covariance matrix:

Given a N×V data matrix *X*, its V×V feature covariance matrix $B=[b_{vw}]$:

$$b_{vw} = \frac{1}{N}\sum_{i=1}^{N}(x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w}),$$

$\overline{x_v}, \overline{x_w}$ **- means**

Correlation matrix:

Matrix *X* is normalized by standard deviations, the covariances b_{vw} are correlation coefficients

$$b_{vw} = \frac{1}{\sigma_v \sigma_w}\sum_{i=1}^{N}(x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w})/N,$$

σ_v, σ_w **- standard deviations**

Correlation coefficient

- PCA: Covariance matrix and correlation matrix
- **Scatterplot**
- Three frameworks for correlation coefficient
  - Naïve approach
  - Regression: Correlation and determinacy; properties and meaning
  - Probability: Gaussian distribution
- Weird correlation case studies
- Different function and/or different criteria: Nature-inspired optimization
- Homework 6

Meaning of correlation, 1

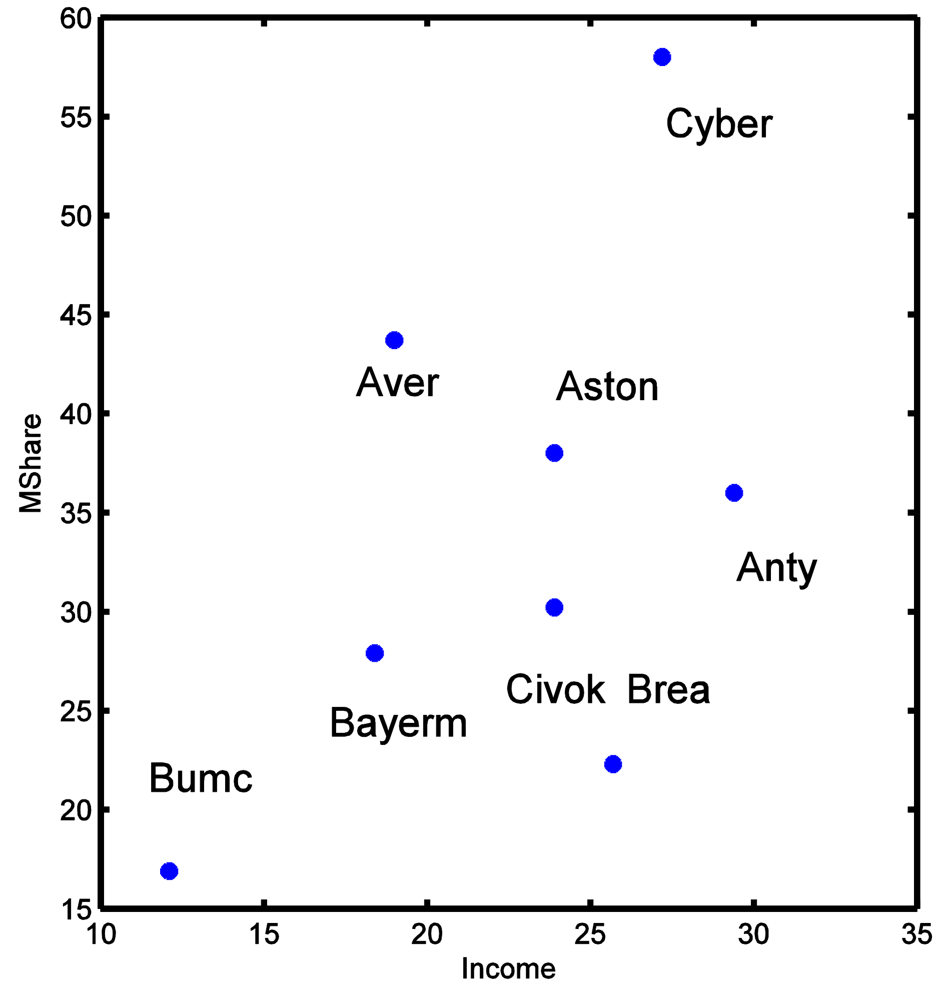
Illustrative: take two features

Company name	Income, $mln	MSha,%	Nsu	EC	Sector
Aversi	**19.0**	**43.7**	2	No	Utility
Antyos	**29.4**	**36.0**	3	No	Utility
Astonite	**23.9**	**38.0**	3	No	Industrial
Bayermart	**18.4**	**27.9**	2	Yes	Utility
Breaktops	**25.7**	**22.3**	3	Yes	Industrial
Bumchist	**12.1**	**16.9**	2	Yes	Industrial
Civok	**23.9**	**30.2**	4	Yes	Retail
Cyberdam	**27.2**	**58.0**	5	Yes	Retail

Meaning of correlation,2: Scatterplot

Take two features

nam	x	y
Aver	19	43.7
Anty	29.4	36
Aston	23.9	38
Bayerm	18.4	27.9
Brea	25.7	22.3
Bumc	12.1	16.9
Civok	23.9	30.2
Cyber	27.2	58

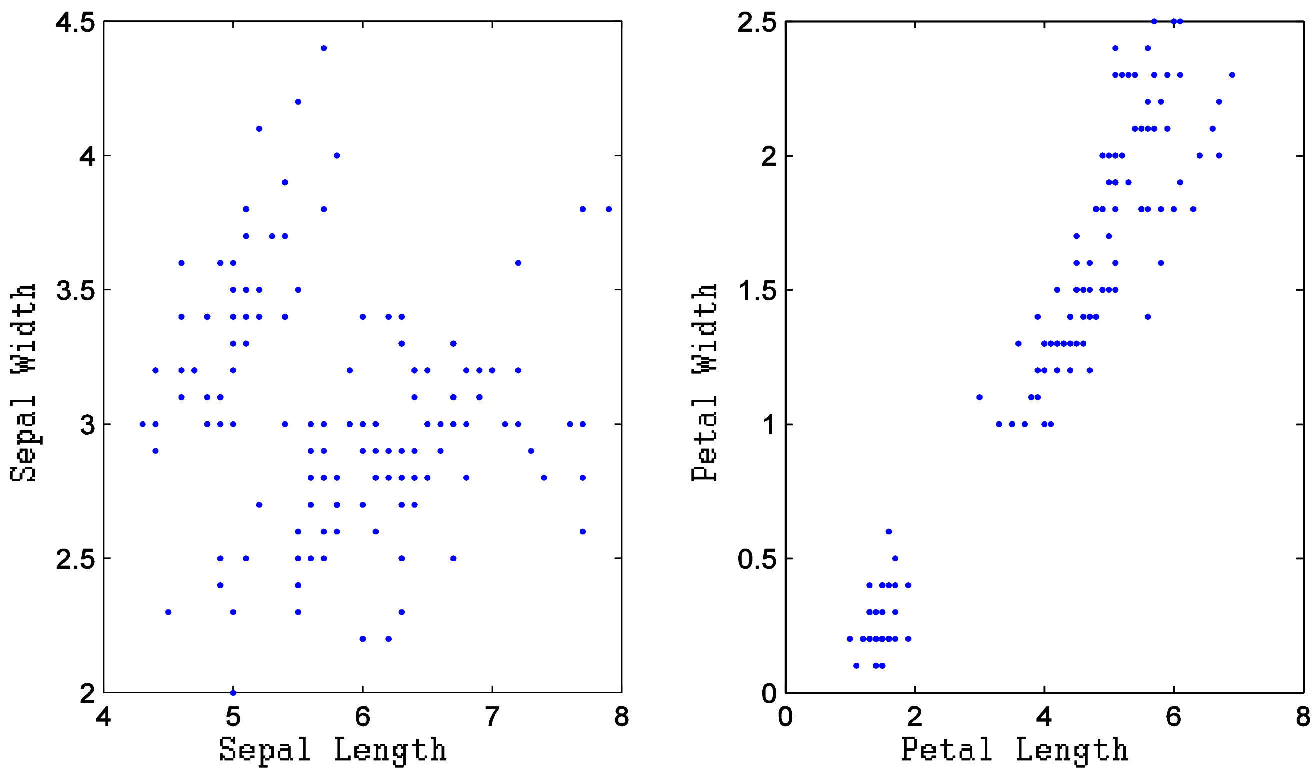


MatLab Command:

>> plot(x,y,'b*'); text(x,y,nam); axis([10 35 15 60])

Iris Scatterplots

Sepal plot Petal plot (noticeable correlation)



The views:

Do differ!

Correlate?

MatLab:

>> subplot(1,2,1);plot(iris(:,1),iris(:,2),'b.');

>> subplot(1,2,2);plot(iris(:,3),iris(:,4),'b.');

Correlation coefficient

- PCA: Covariance matrix and correlation matrix
- Scatterplot
- **Three frameworks for correlation coefficient**
  - **Naïve approach**
  - Regression: Correlation and determinacy; properties and meaning
  - Probability: Gaussian distribution
- Weird correlation case studies
- Different function and/or different criteria: Nature-inspired optimization
- Homework 6

Naïve interpretation of correlation

- $b_{vw} = \frac{1}{\sigma_v \sigma_w} \sum_{i=1}^{N} (x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w})/N,$

= $<\boldsymbol{yv}, \boldsymbol{yw}>$ where

$$\boldsymbol{y_v} = (\boldsymbol{x_v} - \overline{\boldsymbol{x_v}}) / \|\boldsymbol{x_v} - \overline{\boldsymbol{x_v}}\|$$

$$\boldsymbol{y_w} = (\boldsymbol{x_w} - \overline{\boldsymbol{x_w}}) / \|\boldsymbol{x_w} - \overline{\boldsymbol{x_w}}\|$$

Cosine of angle between y_v and y_w

Between -1 and 1. 0 ~ orthogonality, 1 – same, –1 – converse ($180^0 = \pi$)

Correlation coefficient

- PCA: Covariance matrix and correlation matrix
- Scatterplot
- **Three frameworks for correlation coefficient**
  - Naïve approach
  - **Regression: Correlation and determinacy; properties and meaning**
  - Probability: Gaussian distribution
- Weird correlation case studies
- Different function and/or different criteria: Nature-inspired optimization
- Homework 6

2D Linear regression 1

A bit of History

Francis Galton (1822-1911), another

grandson of Erasmus Darwin,

obsessed with the idea that

talent is inherited,

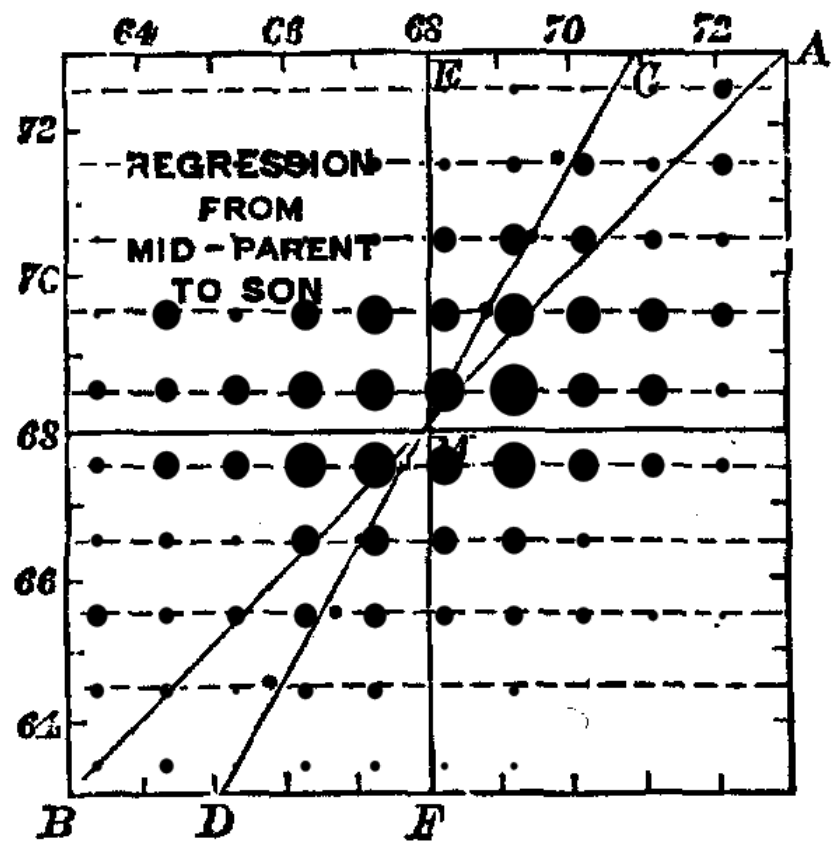
finds that the height of son

regresses to the mean,

from father's height (1885) –

This explains the term.

FIG. 10.

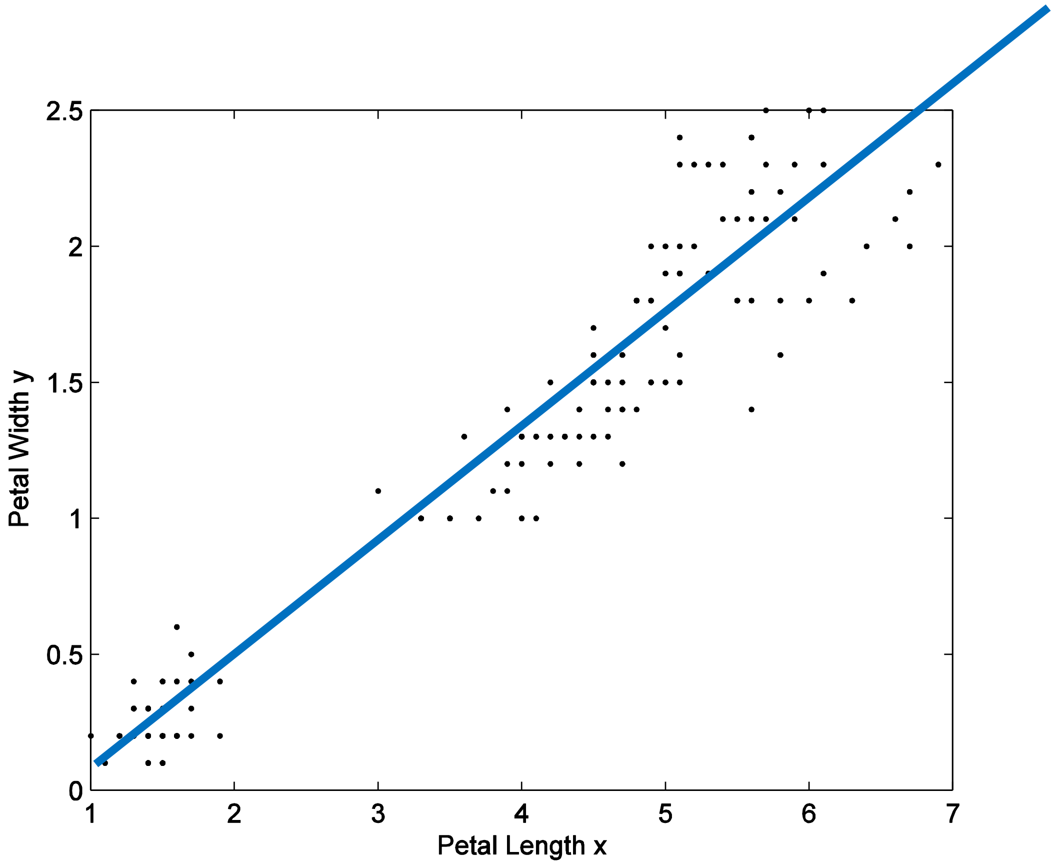


2D Linear regression, 2

Iris Petal Width: how can we express it

through Petal Length linearly

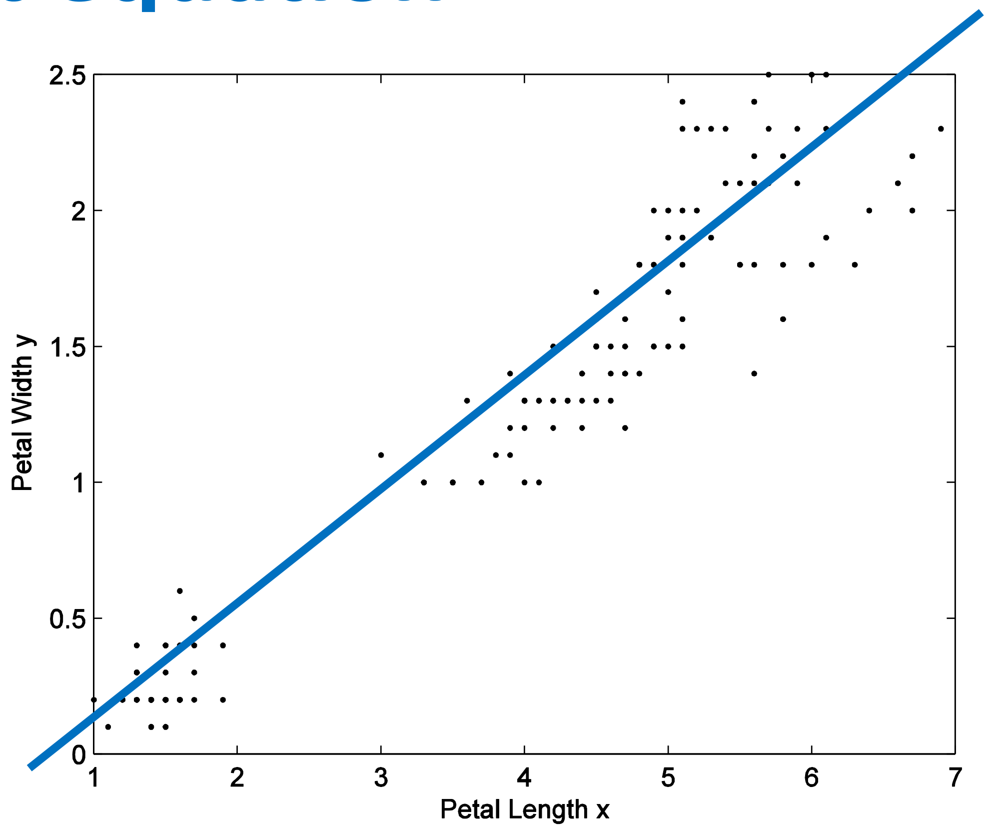
PeWi=a*PeLe+b



2D Linear regression, 3

Iris How can we fit equation

PeWi=a*PeLe+b



Meaning of a:

a = Change in PeWi at PeLe changed by 1

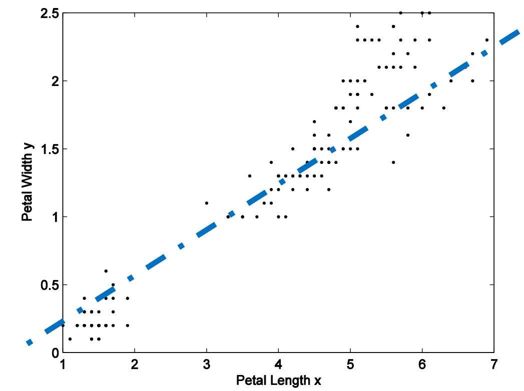
(slope)

b = expected PeWi at PeLe=0 **(This requires a bit of fantasy,,,)**

(intercept)

2D Linear regression, 4

How can we express y=ax+b with minimum error? Maths:



At entity *i=1, 2, . . ., N* equation

$$y_i = ax_i + b + e_i$$

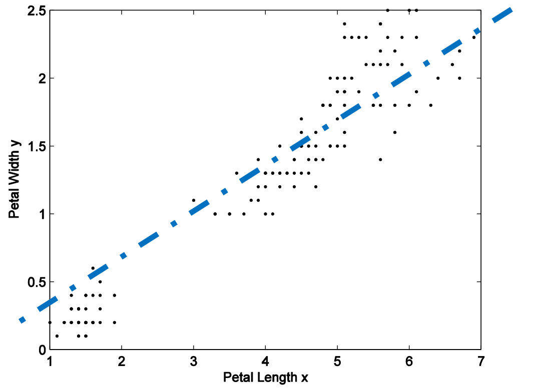
where e_i is error (residual)

Idea: Find *a* and *b* minimizing errors e_i

2D Linear regression ,5

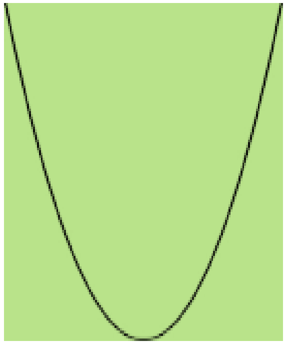
Problem: Find *a* and *b* minimizing errors squared

$$L(a,b) = \sum_{i=1}^{N} (y_i - a x_i - b)^2$$



(least-squares criterion)

***L(a,b)* is parabolic over *a*, *b*:**



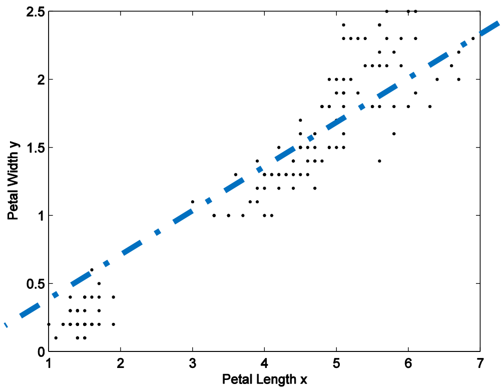
Therefore, first-order optimality conditions from calculus should work:

$$\frac{\partial L}{\partial a} = 2\sum_{i=1}^{N} (y_i - a x_i - b)(-x_i) = 0 \qquad (*)$$

$$\frac{\partial L}{\partial b} = 2\sum_{i=1}^{N} (y_i - a x_i - b)(-1) = 0 \qquad (**)$$

2D Linear regression, 6

Soving first-order optimality equations:



$$(*)\quad 2\sum_{i=1}^{N}(y_i - ax_i - b)(-x_i) = 0$$

$$(**)\quad 2\sum_{i=1}^{N}(y_i - ax_i - b)(-1) = 0$$

Divide (**) by -2 and transfer *b* to the right:

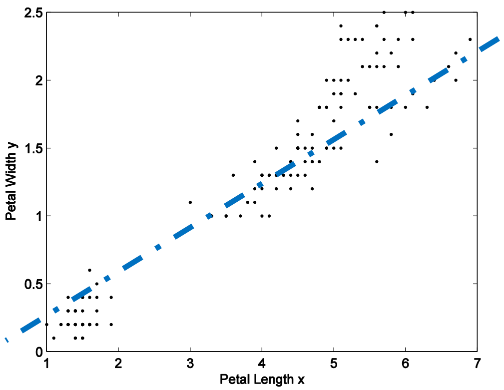
$$\sum_{i=1}^{N} y_i - a\sum_{i=1}^{N} x_i = Nb \quad ,$$

therefore

$$b = \bar{y} - a\bar{x},$$ **where $\bar{y}, \bar{x}$– means of *y*, *x*, respectively**

2D Linear regression, 7

Soving first-order optimality equations:



Now we have

$$(*) \quad 2\sum_{i=1}^{N}(y_i - ax_i - b)(-x_i) = 0$$

$$(**) \quad b = \bar{y} - a\bar{x},$$

where $\bar{y}, \bar{x}$– means of *y*, *x*, respectively.

It remains to find *a* from (*). Put this *b* in (*), divide by -2:

$$\sum_{i=1}^{N}(y_i - ax_i - \bar{y} + a\bar{x})x_i = 0 \quad .$$

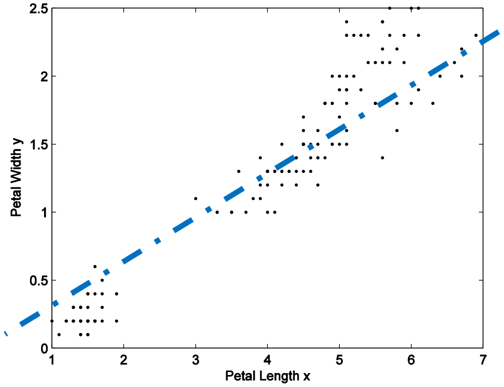
Let us collect *a*-items on the left, the others on the right:

$$a\sum_{i=1}^{N}(x_i - \bar{x})x_i = \sum_{i=1}^{N}(y_i - \bar{y})x_i$$ **. This implies**

$$a = \frac{\sum_{i=1}^{N}(y_i - \bar{y})\, x_i}{\sum_{i=1}^{N}(x_i - \bar{x})\, x_i}$$

2D Linear regression, 8

Polishing first-order optimality equations:



(**) $b = \bar{y} - a\bar{x}$

(*) $a = \frac{\sum_{i=1}^{N}(y_i - \bar{y})x_i}{\sum_{i=1}^{N}(x_i - \bar{x})x_i}$

Notice: $\sum_{i=1}^{N}(x_i - \bar{x}) = \sum_{i=1}^{N}(y_i - \bar{y}) = 0$

Therefore

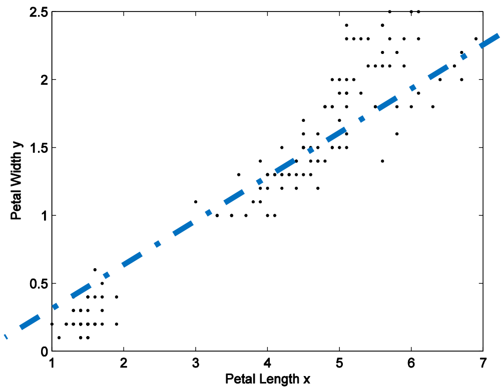
$$a = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})/N}{\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})/N}$$

2D Linear regression, 9

Polishing first-order optimality equations:

$$(**) \qquad b = \bar{y} - a\bar{x}$$

$$(*) \qquad a = \frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})/N}{\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})/N}$$



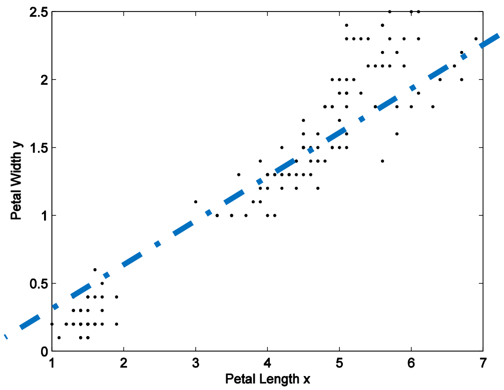
Notice: the denominator is the variance of *x*, $\sigma^2(x)$!!!

Introduce a symmetric expression, correlation coefficient

$$\rho = \frac{\sum_{i=1}^{N}(y_i - \bar{y})\,(x_i-\bar{x})/N}{\sigma(x)\sigma(y)}$$

2D Linear regression, 10

Polishing first-order optimality equations:



$$\boldsymbol{b = \bar{y} - a\bar{x}} \qquad (**)$$

$$\boldsymbol{a = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})/N}{\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})/N}} \qquad (*)$$

The denominator is the variance of *x*, $\sigma^2(x)$!!!

Using a symmetric expression, correlation coefficient

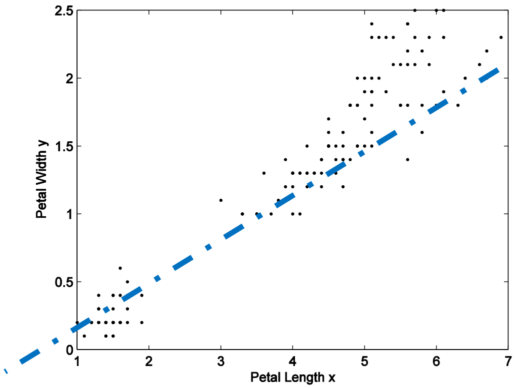
$$\rho = \frac{\sum_{i=1}^{N}(y_i - \bar{y})\,(x_i - \bar{x})/N}{\sigma(x)\sigma(y)}$$

leads to (*) rewritten as

$$\boldsymbol{a = \rho\frac{\sigma(y)}{\sigma(x)}} \qquad (*)$$

2D Linear regression, 11

Polishing first-order optimality equations:



$$b = \bar{y} - a\bar{x} \qquad (**)$$

$$a = \rho \frac{\sigma(y)}{\sigma(x)} \qquad (*)$$

where

$$\rho = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})/N}{\sigma(x)\sigma(y)}$$

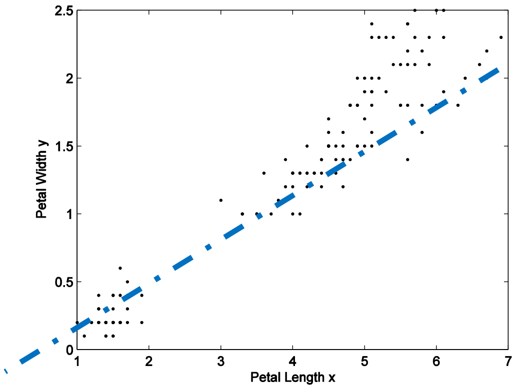
Have we found a solution, the values *a* and *b* minimizing the residuals squared *L(a,b)*?

Yes, we have.

Remains to be done: find the minimum value of *L(a,b)*.

2D Linear regression, 12

Finding minimum *L(a,b)*: Put optimal



$$b = \bar{y} - a\bar{x} \qquad (**)$$

$$a = \rho\frac{\sigma(y)}{\sigma(x)} \qquad (*)$$

into formula for *L*:

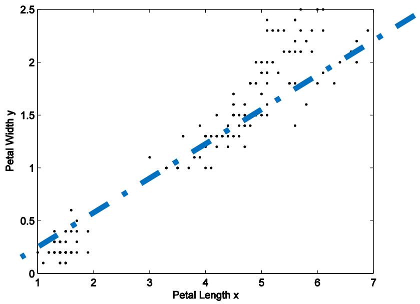
$$L(a,b) = \sum_{i=1}^{N}(y_i - ax_i - b)^2 = \sum_{i=1}^{N}\left(y_i - \rho\frac{\sigma(y)}{\sigma(x)}x_i - \bar{y} + \rho\frac{\sigma(y)}{\sigma(x)}\bar{x}\right)^2 \quad \text{(i)}$$

$$L(a,b) = \sum_{i=1}^{N}\left[(y_i - \bar{y}) - \rho\frac{\sigma(y)}{\sigma(x)}(x_i - \bar{x})\right]^2 =$$

$$= \sum_{i=1}^{N}(y_i - \bar{y})^2 - 2\rho\frac{\sigma(y)}{\sigma(x)}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) + \rho^2\frac{\sigma^2(y)}{\sigma^2(x)}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad \text{(II)}$$

$$L(a,b) = N\sigma^2(y) - 2N\rho^2\sigma^2(y) + N\rho^2\sigma^2(y) = N\sigma^2(y)(1 - \rho^2) \quad \text{(iii)}$$

2D Linear regression, 13: all solved



Final linear regression optimality equations:

$$b = \bar{y} - a\bar{x} \qquad (**)$$

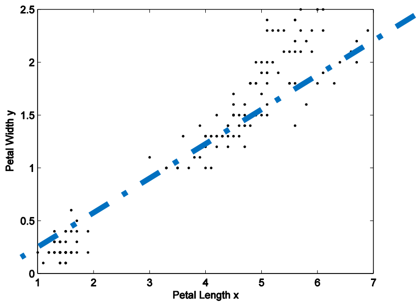
$$a = \rho \frac{\sigma(y)}{\sigma(x)} \qquad (*)$$

The minimum

$$L(a,b)/N = \sum_{i=1}^{N}(y_i - ax_i - b)^2/N = \sigma^2(y)(1-\rho^2) \qquad (***)$$

So what?

2D Correlation and determinacy coefficients: properties and meaning 1



Final linear regression optimality equations:

$$b = \bar{y} - a\bar{x} \qquad (**)$$

$$a = \rho \frac{\sigma(y)}{\sigma(x)} \qquad (*)$$

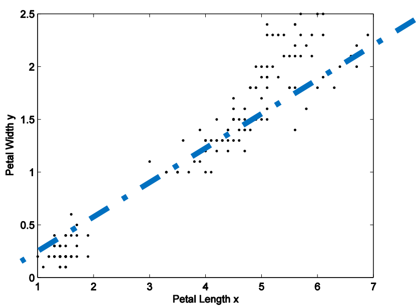
The minimum value of

$$L(a,b)/N = \sum_{i=1}^{N}(y_i - ax_i - b)^2/N = \sigma^2(y)(1-\rho^2) \qquad (***)$$

Equation (***) : ρ^2, the determinacy coefficient, is the proportion of the variance $\sigma^2(y)$ taken into account by the linear regression of *y* over *x*.

Value *L(a,b)/N* in (***) is referred to as the residual variance.

Correlation coefficient: four properties, 2



Final linear regression optimality equations:

$$\boldsymbol{b = \bar{y} - a\bar{x}} \qquad (**)$$

$$\boldsymbol{a = \rho \frac{\sigma(y)}{\sigma(x)}} \qquad (*)$$

$$\boldsymbol{L(a,b)/N = \sum_{i=1}^{N}(y_i - ax_i - b)^2/N = \sigma^2(y)(1-\rho^2)} \qquad (***)$$

(i) The determinacy coefficient ρ^2 is within interval [0,1]; the correlation coefficient ρ , within [-1, +1]. Indeed, (1-ρ^2)≥0 because L ≥0.

(ii) Coefficient ρ is 1 or -1 if and only if regression equation y=ax+b is true for every i=1,2,...,N with no errors. Indeed, L=0 only if all the items are zero.

(iii) Coefficient ρ is 0 if and only if the slope *a*=0, because of (*).

(iv) The sign of ρ is the sign of the slope *a*; therefore, *x* and *y* are related positively if $\rho>0$, and negatively, if $\rho<0$.

2D Correlation coefficient: properties and meaning, 3

(i) The determinacy coefficient ρ^2 is within interval [0,1]; the correlation coefficient ρ , within [-1, +1]. Indeed, $(1-\rho^2)\geq 0$ because L $\geq$0.

(ii) Coefficient ρ is 1 or -1 if and only if regression equation y=ax+b is true for every i=1,2,…,N with no errors. Indeed, L=0 only if all the items are zero.

(iii) Coefficient ρ is 0 if and only if the slope *a*=0, because of (*).

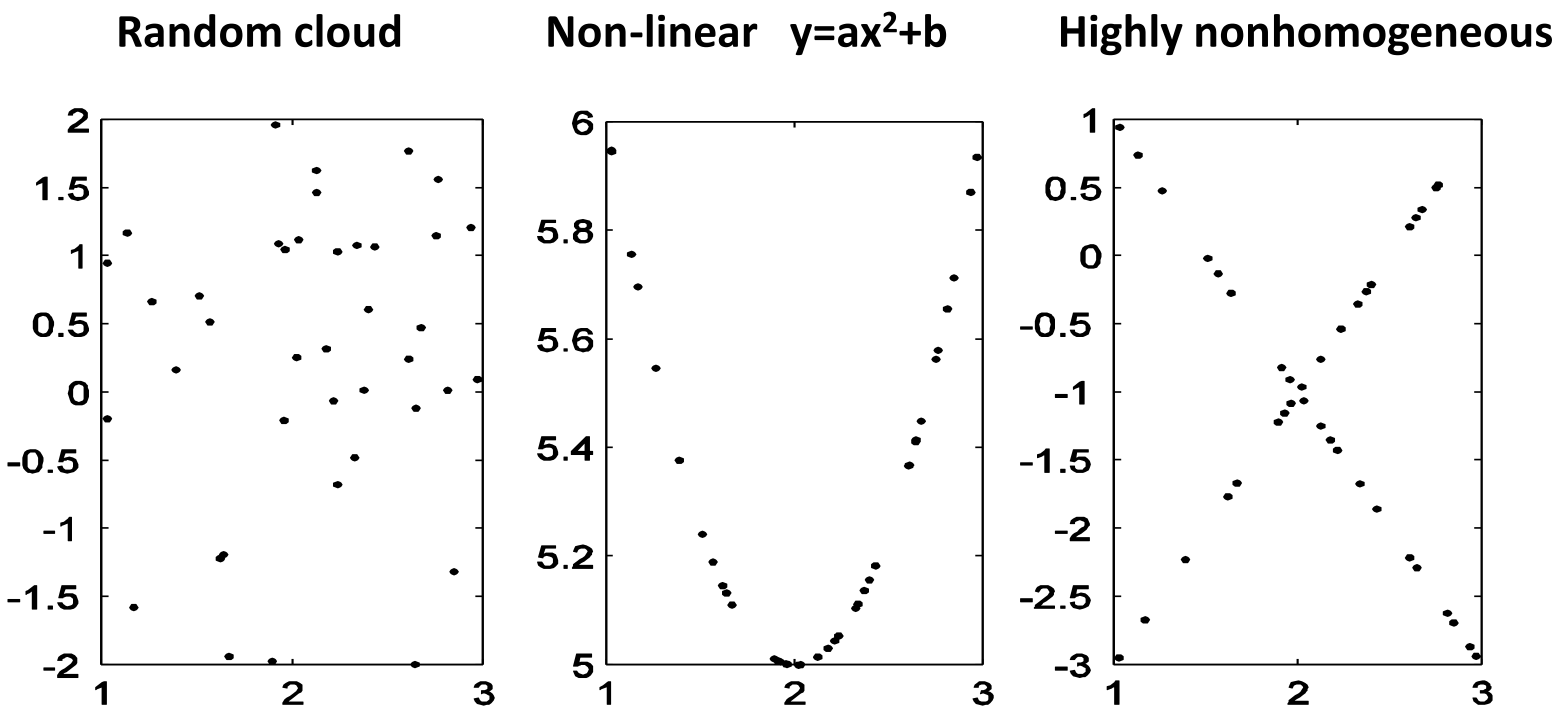
(iv) The sign of ρ is the sign of the slope *a*; therefore, *x* and *y* are related positively if $\rho>0$, and negatively, if $\rho<0$.

These show that correlation coefficient is a measure of degree of a linear relation between x and y.

Meaning of $\rho = 0$: No relation? (Nope!)

$$a = \rho \frac{\sigma(y)}{\sigma(x)} \qquad (*)$$

(iii) Coefficient ρ=0 if and only if the slope a=0 in the regression equation.

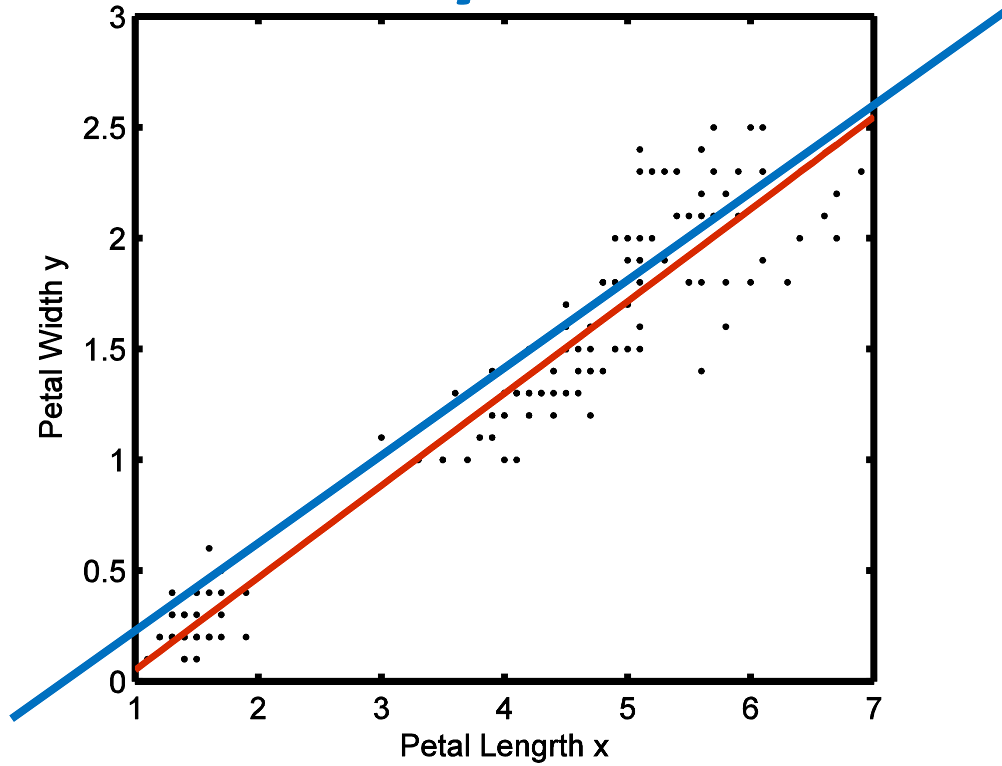


Meaning of ρ : Good Case

Iris Petal Width through Petal Length y=a*x+b

$b = \bar{y} - a\bar{x}$ (**)

$a = \rho \frac{\sigma(y)}{\sigma(x)}$ (*)



>> x=iris(:,3);

>> y=iris(:,4);

>> ρ =corr(x,y)% ρ =0.9629

Even inspite that points (x_i,y_i)

are not exactly on a line, determinacy ρ ^{2}=92.7% (7.3% of y-scatter left unexplained)

slope = 0.4158

intercept = -0.3631 (red line)

High determinacy warrants no high precision:

Iris Petal Width=0.4158*x- 0.3631 (א)

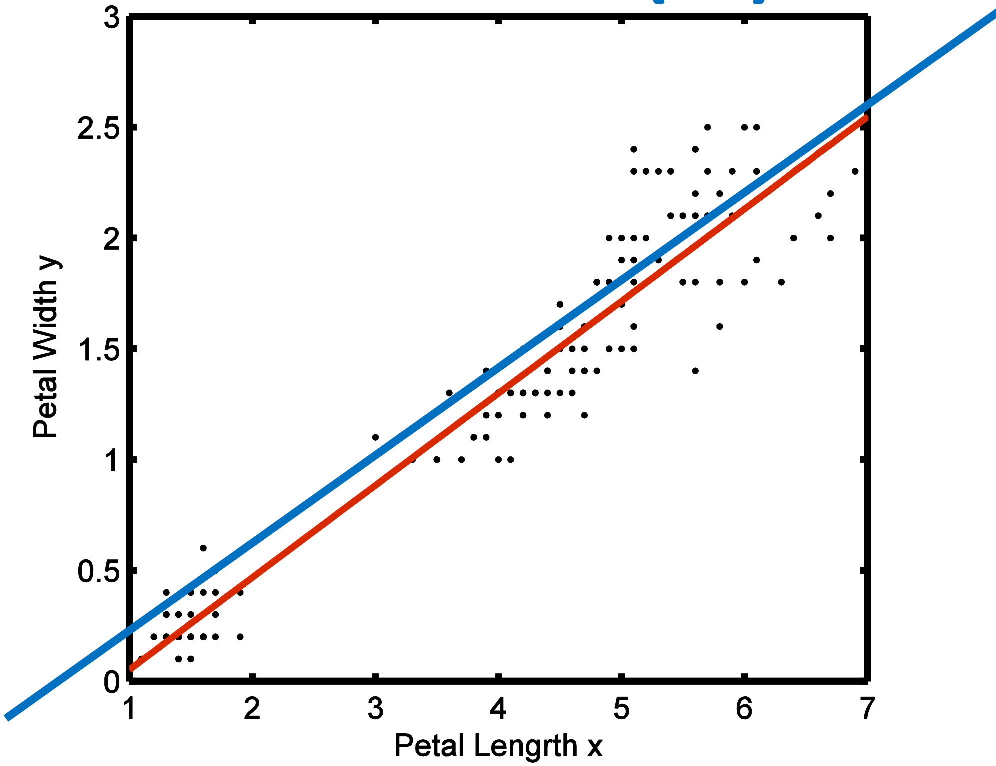
Although points (x_i,y_i) do not fit

the **regression line**, ρ^2=92.7% -

Almost unity!

Test for the errors at prediction

n.	x	y	PWא	error,%
23	1.4	0.1	0.22	119.0
51	4.5	1.5	1.51	0.5
86	4.3	1.3	1.42	9.6
138	5.0	1.9	1.72	-9.7
142	5.7	2.5	2.01	-19.7



Average precision 20.6% (s=25.8):

not that high

Correlation coefficient

- PCA: Covariance matrix and correlation matrix
- Scatterplot
- Three frameworks for correlation coefficient
  - Naïve approach
  - Regression: Correlation and determinacy; properties and meaning
    - **Weird correlation case studies**
  - Probability: Gaussian distribution
- Different function and/or different criteria: Nature-inspired optimization
- Homework 6

Correlation: Weird case,1

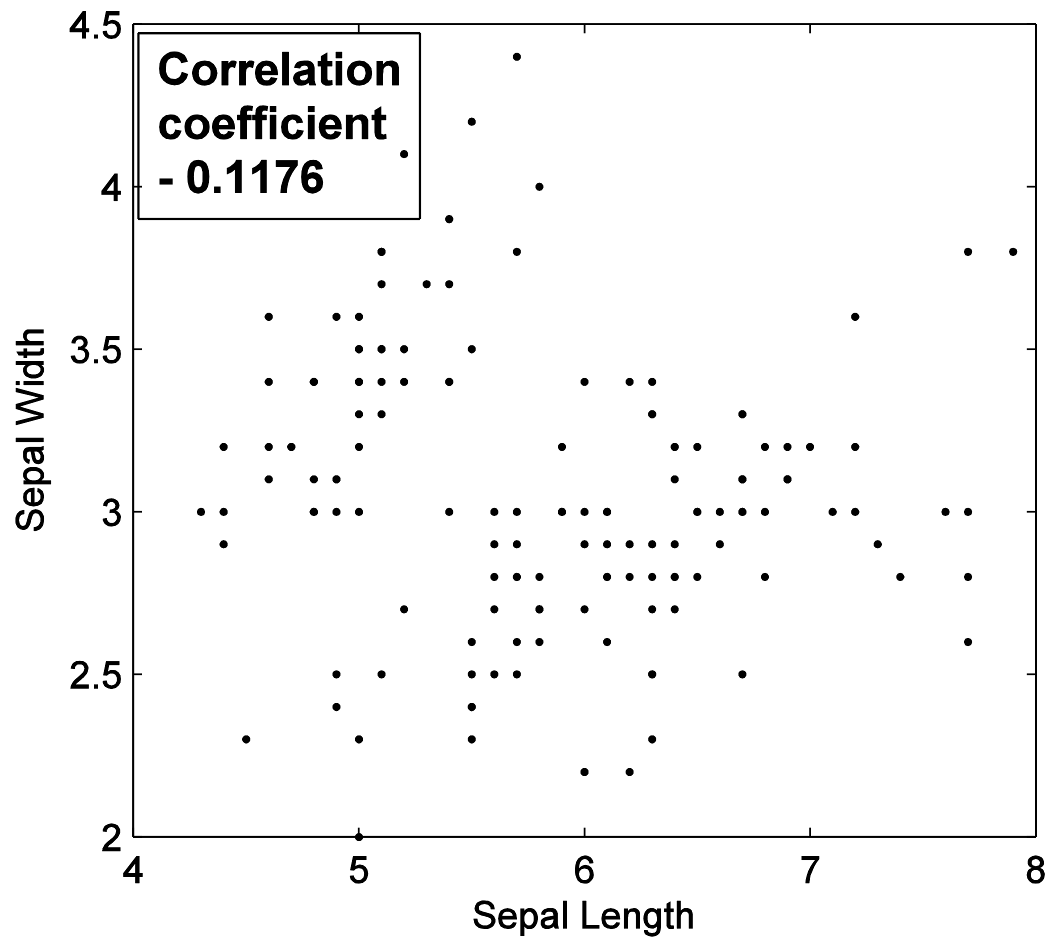
Iris Sepal Length and Sepal Width (other pair)

This is highly unnatural.

The SL and SW should go

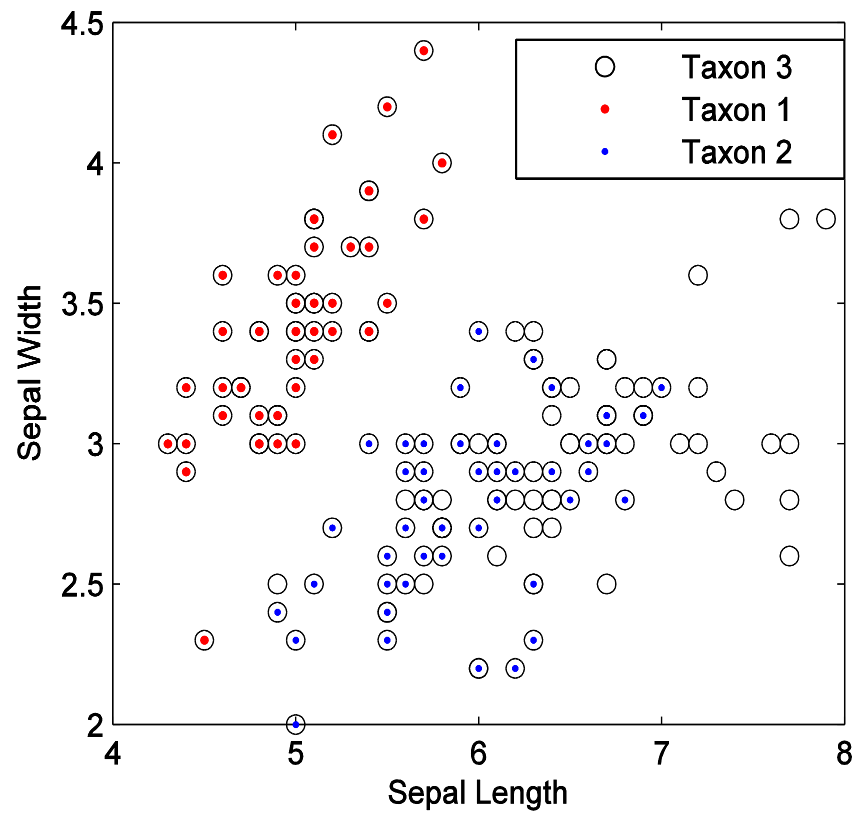
along, with a positive,

if not high, correlation.



Correlation: Weird case, 2

Iris Corr(Sepal Length, Sepal Width)=-0.1176



This highly unnatural value

comes from the non-homogeneity:

a mix of three taxa

correlation within each,

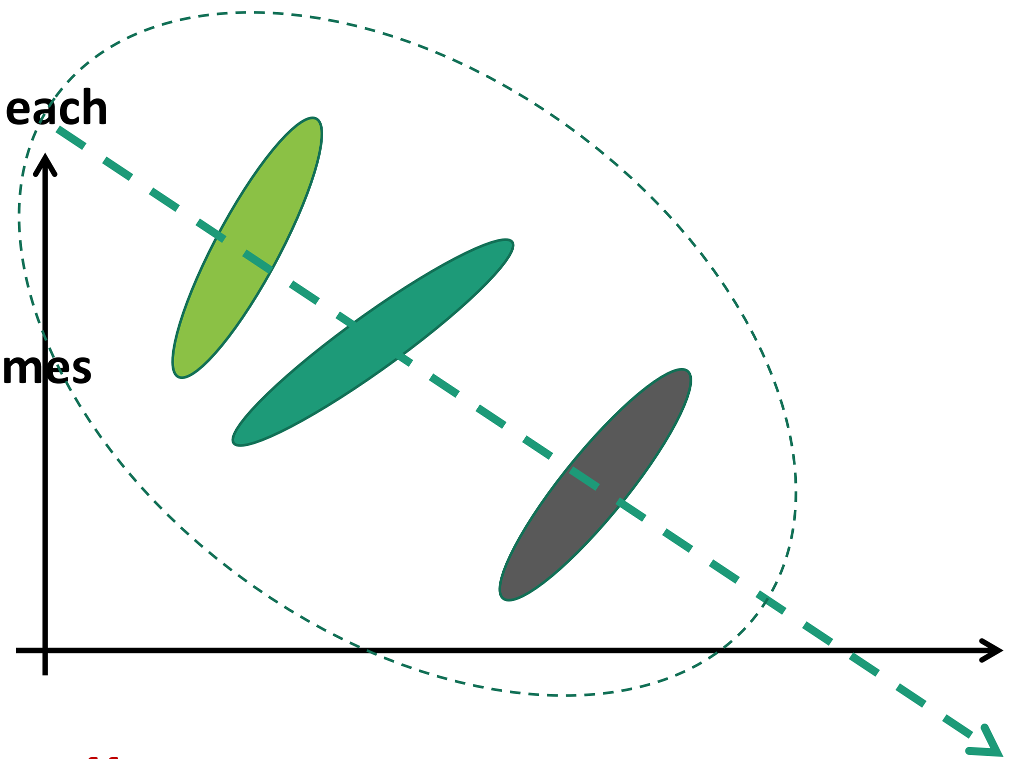
0.74, 0.53, 0.46,

is positive, if not quite high

Correlation: Weird case, 3

A scheme of FALSE negative correlation by merging different groupings

A high positive correlation within each group, red, blue, grey; yet a negative correlation overall !!!



Instances of data manipulation, sometimes unintentional, should have made great politicians (B. Disraeli, UK Prime-Minister) to say even if they have not:

THREE LEVELS of LIE: "A LIE, DAMNED LIE, and STATISTICS!",

Correlation : Weird case , 4 A quiz

A scheme of FALSE POSITIVE correlation by merging different groupings

A high NEGATIVE correlation within each group; yet a positive correlation overall !!!

Can you guess a proper picture for this?
Try proposing a scatter-plot for this case.

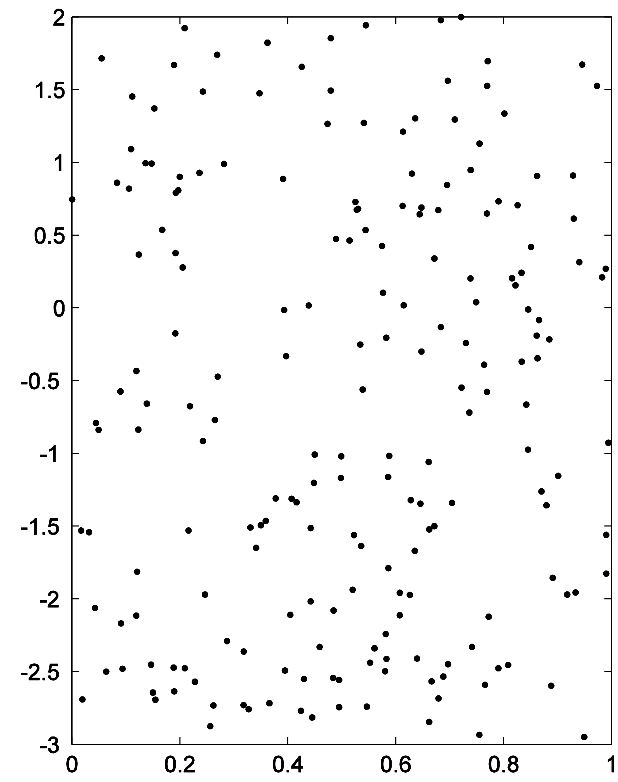
THREE LEVELS of LIE: "A LIE, DAMNED LIE, and STATISTICS!", (I. Dizraeli ? – no)

Correlation and regression: Weird Case 2

Inflated correlation, 1:

Random features generated:

```
>> g=rand(200,1);
>> h=5*rand(200,1)-3;
>> plot(g,h, 'k.')
>> corr(g,h) =0.06 % close to 0
```



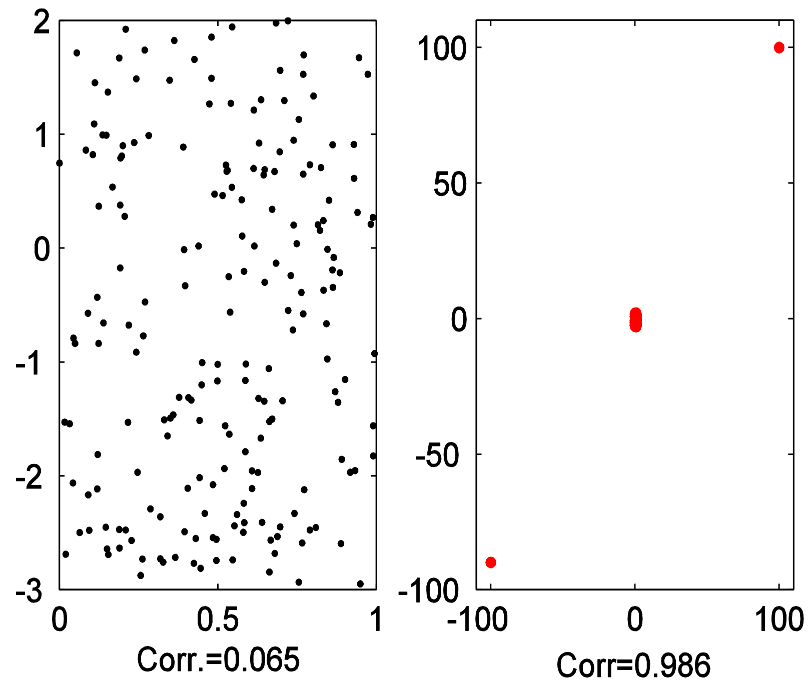
Correlation and regression: Weird Case 2, 2

Inflated correlation, 2: by adding outliers

Random features generated:

On the left:
>> g=rand(200,1);
>> h=5*rand(200,1)-3;
>> rho=0.0650

On the right
TWO outliers are added:
>> g(201)=100; g(202)=-100;
>> h(201)=-100; h(202)=-90;
>> rho=0.9862 % almost a unity!



Some observed unexplained correlations

- Social drinking and earnings - drinkers earn more money (B.L. Peters, E. Stringham (2006)*, Journal of Labor Research*, 27(3), 411-421).
- Chocolate consumption and the numbers of Nobel prize winners, both relative to the population size (F.H. Messerli (2012)*, The New England Journal of Medicine*, *367*(16), 1562).
- Numbers of: (a) newborn babies and (b) brooding pairs of storks observed across 17 countries (R. Matthews, *Teaching Statistics*, 2000, 22, 2, 36-38).

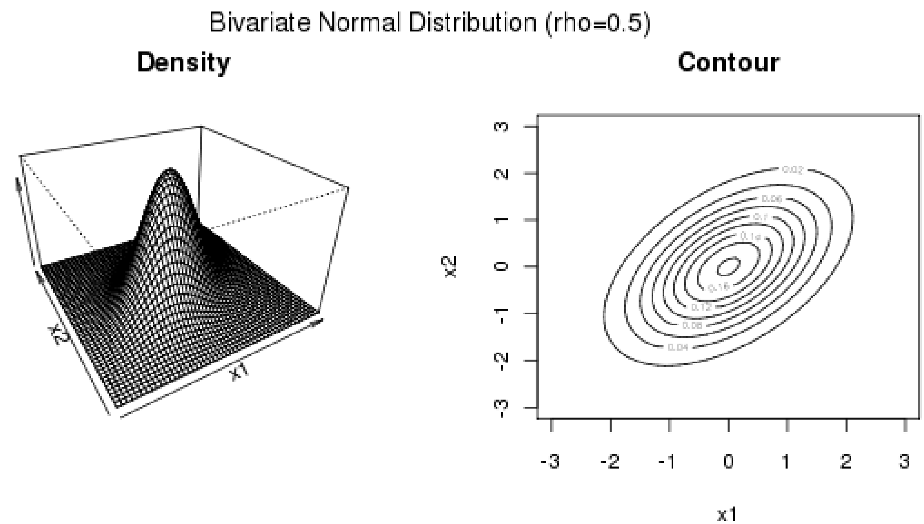
Correlation coefficient

- PCA: Covariance matrix and correlation matrix
- Scatterplot
- Three frameworks for correlation coefficient
    - Naïve approach
    - Regression: Correlation and determinacy; properties and meaning
        - Weird correlation case studies
    - **Probability: Gaussian distribution**
- Different function and/or different criteria: Nature-inspired optimization
- Homework 6

Week 7. K. Pearson's highly creative insight (probabilistic perspective at Correlation coefficient:)

At the standard Multivariate Gaussian

$f(u, \Sigma) = C\exp\{-u^T \Sigma^{-1} u/2\}$ *where* $u=(u_1, u_2, \ldots, u_v)$;



Bivariate case

after z-scoring

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

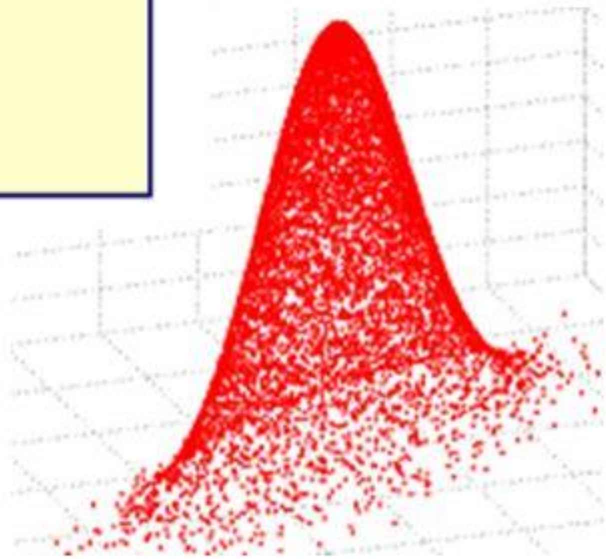
K. Pearson: Correlation coefficient is a sample-based estimate of the parameter ρ in the Gaussian density function under the conventional assumption of independent random sampling.

Multivariate Gaussian Distribution

- In *d-dimensional* space, the Gaussian pdf is:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]}$$

$$p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$



Meaning of correlation coefficient

- A high value of ρ does not warrant any good precision of the regression
- Value of ρ can be inflated by adding a few outliers
- Zero value of ρ does not warrant low association between a target and predictor – this may indicate either
  - High non-homogeneity
  - High non-linearity

- Probabilistic framework: parameter of bivariate Gaussian

Correlation coefficient

- PCA: Covariance matrix and correlation matrix
- Scatterplot
- Three frameworks for correlation coefficient
  - Naïve approach
  - Regression: Correlation and determinacy; properties and meaning
  - Probability: Gaussian distribution
- **Different function and/or different criteria: Nature-inspired optimization**
- Homework 6

A better criterion: **mathematically hopeless**

- Relative error to minimize

$$\Delta(i)=|y_i-ax_i-b|/|y_i|$$

$$\Sigma_i\Delta(i) \Rightarrow \min_{a,\, b}$$

- A very complex function; no good methods within the classical approach (take an admissible solution, then hone and polish it)

Nature-inspired approach to regression: cases of non-linearity or different criteria

- Relative error to minimize

$$\Delta(i)=|y_i-ax_i-b|/|y_i|$$

$$\sum_i \Delta(i) \Rightarrow \min_{a,\, b}$$

- A very complex function; no good methods of classical styles: take an admissible solution, then hone and polish it

Nature-inspired approach for criterion $\Sigma_i \Delta(i) \Rightarrow \min_{a,\, b}$ with $\Delta(i)=|y_i-ax_i-b|/|y_i|,1$

- Rather than honing a single admissible solution, run a nature-inspired evolutionary process for a population of admissible solutions
- First find an area A of admissibility for pairs (a,b)
- Take a population of random p admissible pairs (a(k),b(k)), k=1,2,…,p.
- Define rules for: (1) the population to evolve, (2) elite maintenance, (3) halt.
- Run an evolution process according to the rules.

Nature-inspired approach for $\Sigma_i \Delta(i) \Rightarrow \min_{a,\, b}$ with $\Delta(i)=|y_i-ax_i-b|/|y_i|$, 2

- Find an area A of admissibility. At each i,j=1,…,N, take a(ij), b(ij) from yi=a(ij)xi+b(ij), yj=a(ij)xj+b(ij): a(ij)=(yi-yj)/(xi-xj), b(ij)=yi-a(ij)xi. Sort all a(ij) and b(ij), remove upper and lower percentile t%(about 15%) in intervals [a1,a2] and [b1,b2], so that A=[a1,a2] $\times$[b1,b2].
- Take a population, that is, p $\times$ 2 array f of random p pairs (a(k),b(k)) from A, k=1,2,…,p.

Nature-inspired approach for $\Sigma_i\Delta(i) \Rightarrow \min_{a,\,b}$ with $\Delta(i)=|y_i-ax_i-b|/|y_i|$, 3

- Define rules for:

(1) the population to evolve:

(1a) take the average a and b over the population and replicate them p times in $p\times 2$ array mf, (1b) define fn=f+randn(p,2).*mf/20, (1c) trim fn to remain in A as array fr;

(2) Elite policy

(2a) find in f the best pair (a*,b*) [elite], store it, and replicate it in $p \times 2$ array el;

(2b) define next generation array f'=0.7fr+0.3el and go to step 1 with f=f'. Change the elite if current (a*,b*) are better than those stored.

(3) Halt after 10000 iterations and output (a*,b*).

Nature-inspired approach for the least squares with $y=ax^b$, 1

- Find an area A of admissibility, as above, by the linearized equation log(y)=log(a)+blog(x) by using transformed variables z=log(y) and v=log(x), so that the equations are about z(ij) and v(ij).
- Run the nature-inspired process above.

Example: take x and y generated as follows: x=10*rand(1,50),

for i=1:50;yy=2*x(i)^1.07+2*randn; y(i)=max(yy,1.01); end;

Nature-inspired approach for the least squares with $y=ax^b$, 2

Example: Generate

x=10*rand(1,50),

for i=1:50;yy=2*x(i)^1.07+2*randn; y(i)= yy | 1.01; end;

Nature Inspired process results:

a = 2.0293, b = 1.0760, the average squared error = 0.0003

Linearized regression z=c+bv results :

a = 3.0843, b = 0.8011, the average squared error 4.41.

Most important: a wrong regularity:

b=0.8 is less than 1, whereas the generated b=1.07.

SUMMARY

1. **Scatter plot:** **just a Cartesian representation in 2D**
2. **Correlation coefficient: 3 perspectives**
3. **Linear regression:** **a convenient format to summarize relation between two features**
4. **Determinacy and correlation in the approximation perspective:** **due to the linearity and least-squares criterion; ρ^2 scoring the extent of *y*-variance taken into account; ρ expressing, vaguely, the extent of linear relation between *x* and *y***
5. **Correlation and regression:** **Useful, but be aware of non-homogeneity ["just lies, damned lies, and statistics"].**
6. **Different function and/or different criteria: Nature-inspired optimization**

Correlation coefficient

- **PCA: Covariance matrix and correlation matrix**
- **Scatterplot**
- **Three frameworks for correlation coefficient**
  - **Naïve approach**
  - **Regression: Correlation and determinacy; properties and meaning**
  - **Probability: Gaussian distribution**
- **Different function and/or different criteria: Nature-inspired optimization**
- **Homework 6**

Homework 6

- 1. Find two features in your dataset with more or less "linear-like" scatterplot.
- 2. Display the scatter-plot.
- 3. Build a linear regression of one of the features over the other. Make a comment on the meaning of the slope.
- 4. Find the correlation and determinacy coefficients, and **comment** on the meaning of the latter.
- Make a prediction of the target values for given two or three predictor' values; make a comment
- Compare the mean relative absolute error of the regression on all points of your set and the determinacy coefficient and make comments