Lecture 6 – Correlation

Text

COMP20008

School of Computing and Information Systems

Announcements

Assignment 1 due 23:59 September 5th

- Coding challenge and teamwork evaluation

Assignment 1 due 23:59 September 9th

- Code reviews due

Assignment 1 oral assessments in Week 8 (15 Sept – 19 Sept)

- **See canvas announcements to book your oral assessment**

Assignment 2

- Plan to be released some time next week–
- Same dataset scenario, more open ended (group) report and oral assessment (in-person)

Past exams

- Planned to be posted also some time next week

Agenda

Discuss correlations between pairs of features in a dataset

- Why useful and important
- Pitfalls

Methods for computing correlation

- Pearson correlation
- Mutual information (another method to compute correlation)

Correlations in life

As a child grows, so does his/her clothing size.

As you drink more coffee, the number of hours you stay awake increases.

The longer amount of time you spend in the bath, the more wrinkly your skin becomes.

The older a man gets, the less hair that he has.

https://towardsdatascience.com/3-beautiful-real-life-correlations-fed9e855da52,
https://examples.yourdictionary.com/positive-correlation-examples.html,
https://examples.yourdictionary.com/negative-correlation-examples.html

What is Correlation?

Correlation is used to detect pairs of variables that might have some relationship.

AND

Text

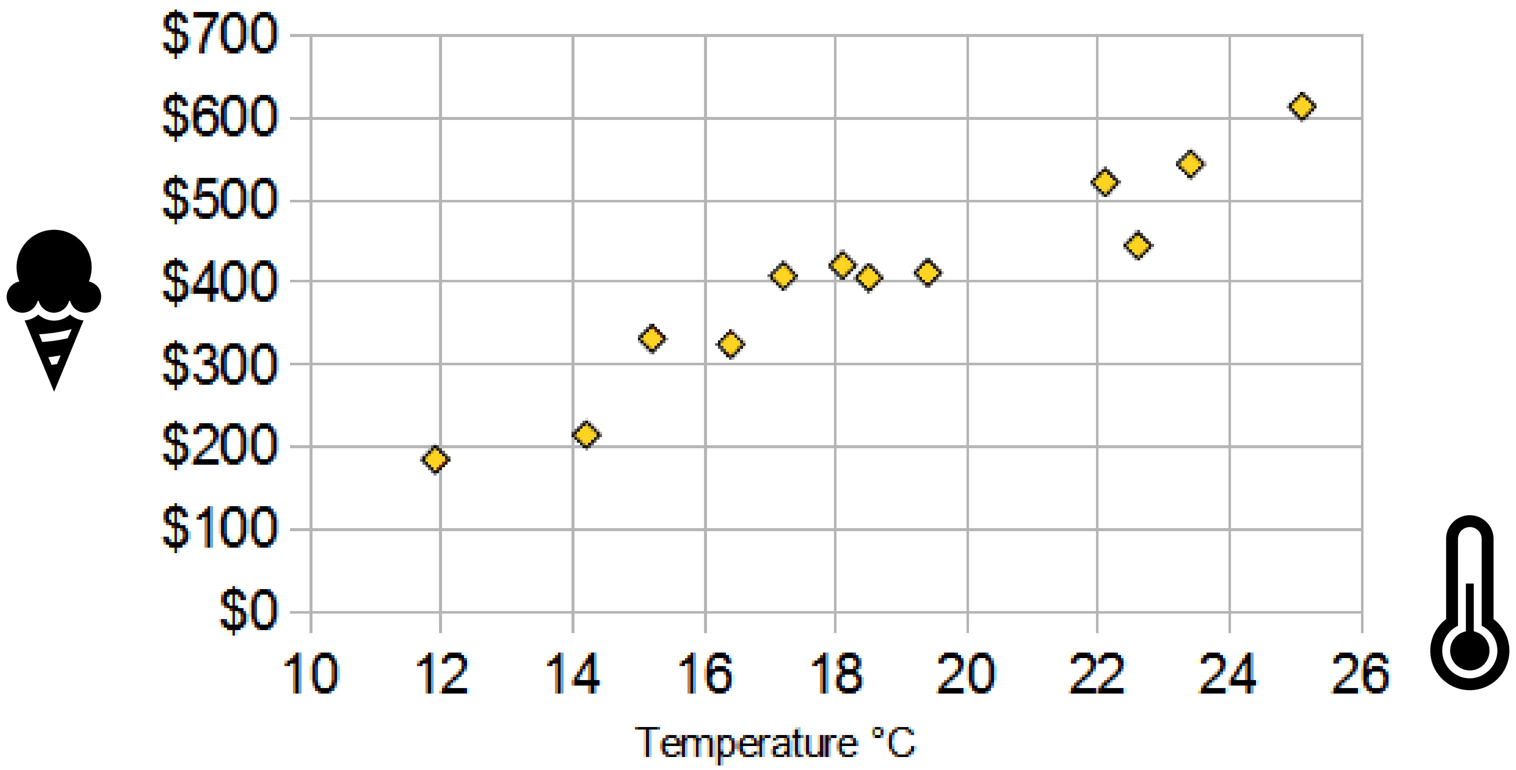
How strong is the relationship?

Ice Cream Sales vs Temperature	
Temperature °C	Ice Cream Sales
14.2°	$215
16.4°	$325
11.9°	$185
15.2°	$332
18.5°	$406
22.1°	$522
19.4°	$412
25.1°	$614
23.4°	$544
18.1°	$421
22.6°	$445
17.2°	$408

https://www.mathsisfun.com/data/correlation.html

What is Correlation?

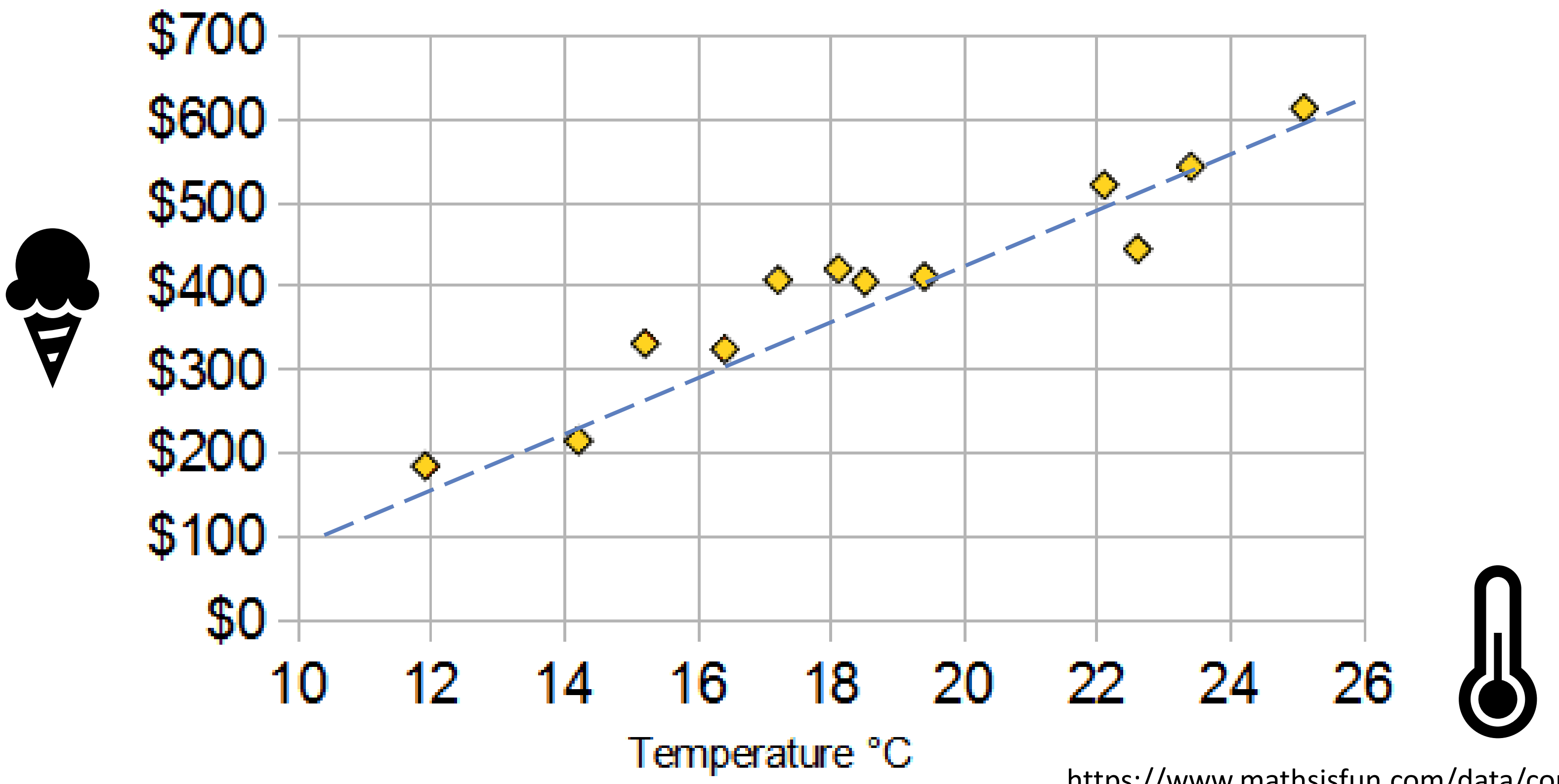
Visually can be identified via inspecting scatter plots



https://www.mathsisfun.com/data/correlation.html

What is Correlation?

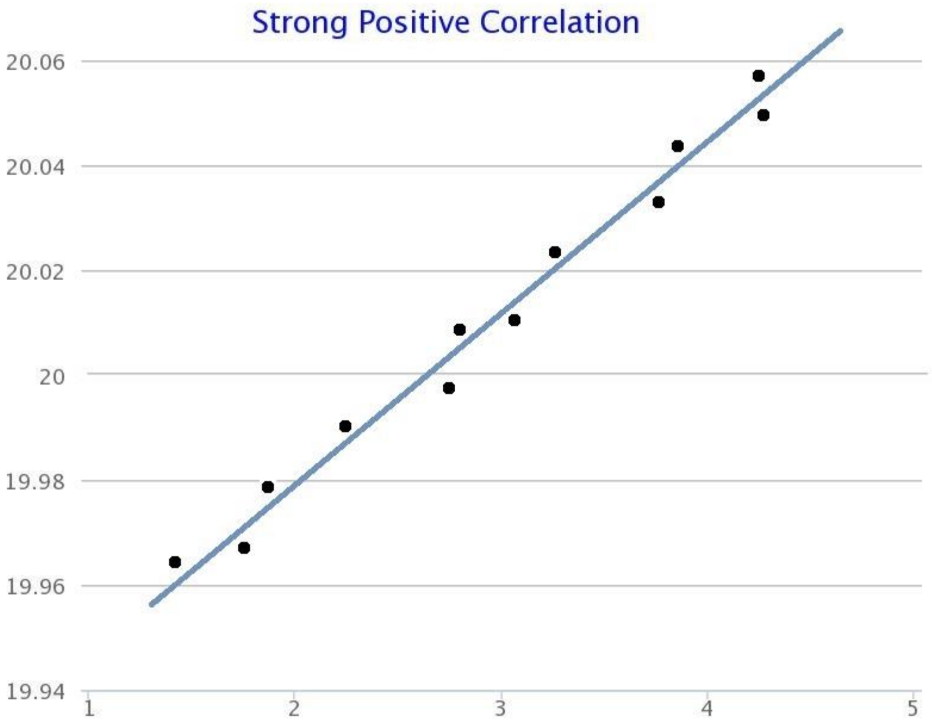
Linear relations

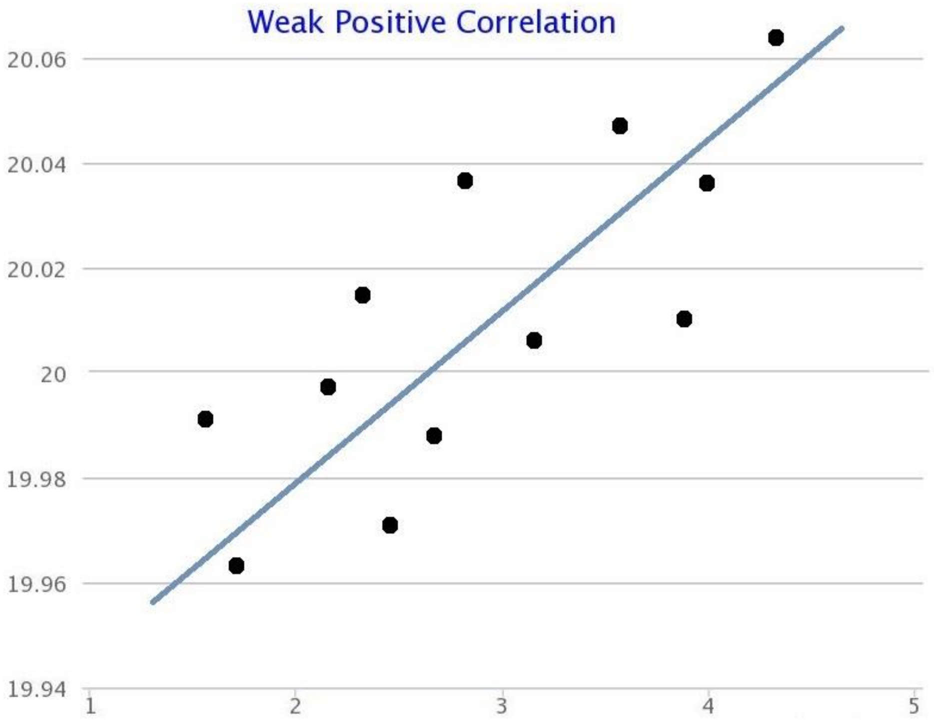


https://www.mathsisfun.com/data/correlation.html

What is Correlation?

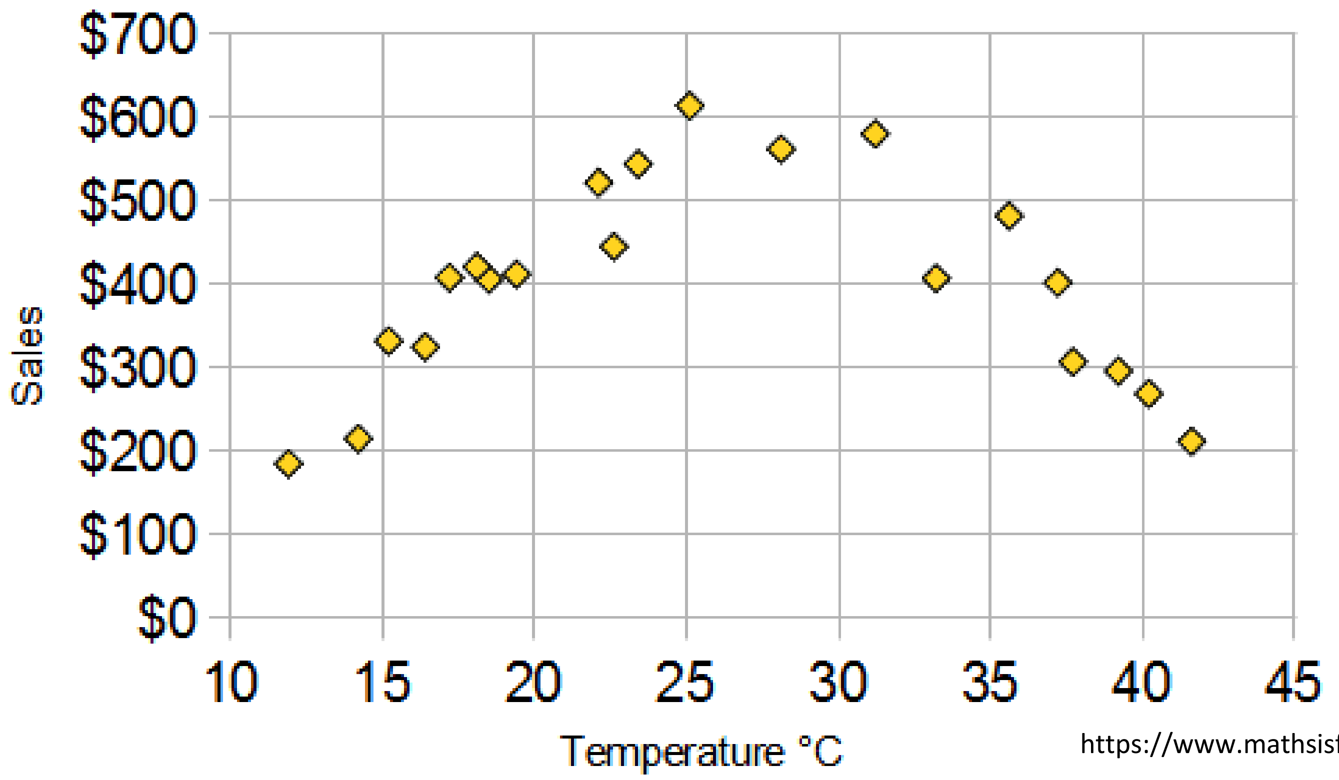
Correlation strength





https://images.app.goo.gl/zZXtjBLR2BcjRpK79

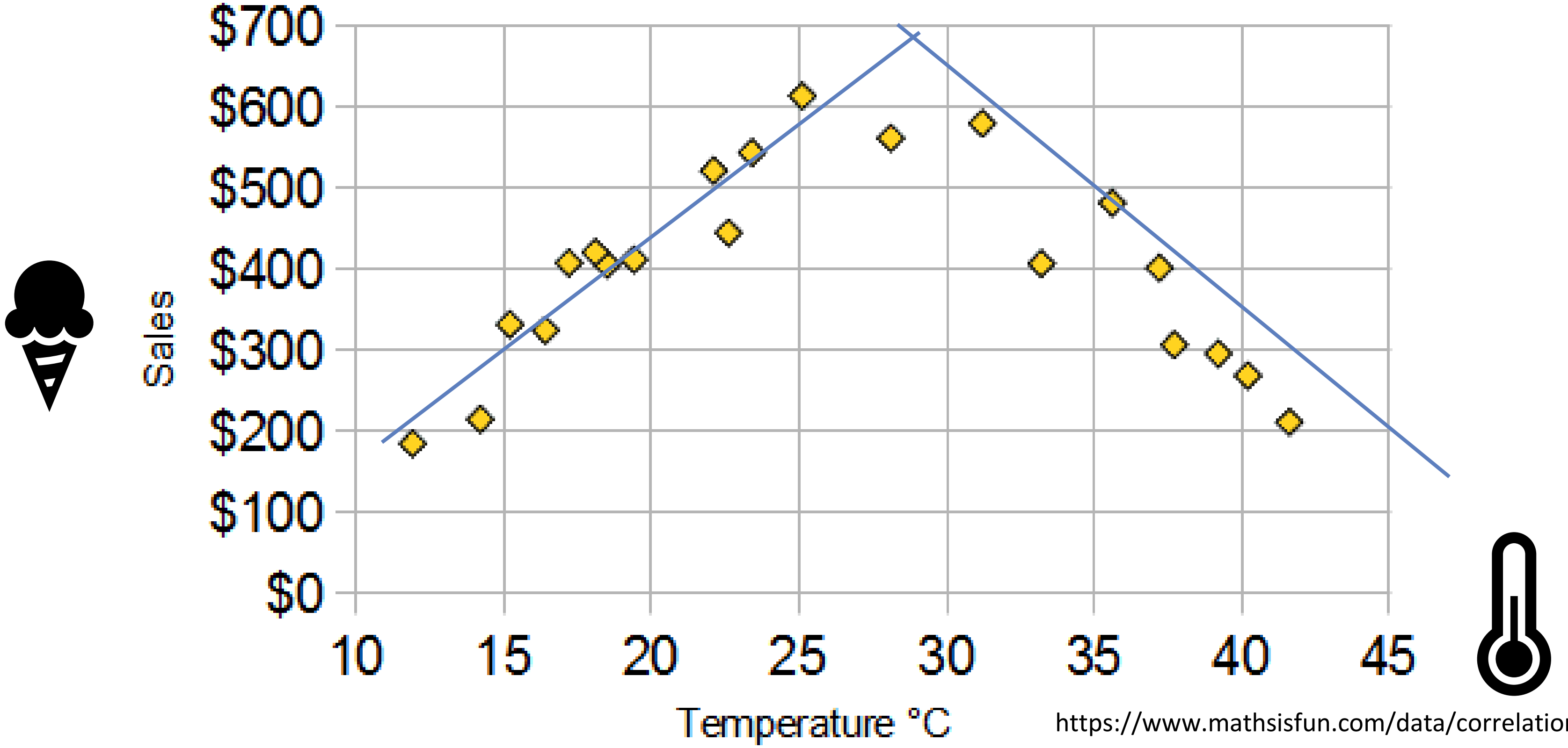
Example of non-linear correlation



https://www.mathsisfun.com/data/correlation.html

It gets so hot that people aren't going near the shop, and **sales start dropping**

Example of non-linear correlation



https://www.mathsisfun.com/data/correlation.html

It gets so hot that people aren't going near the shop, and **sales start dropping**

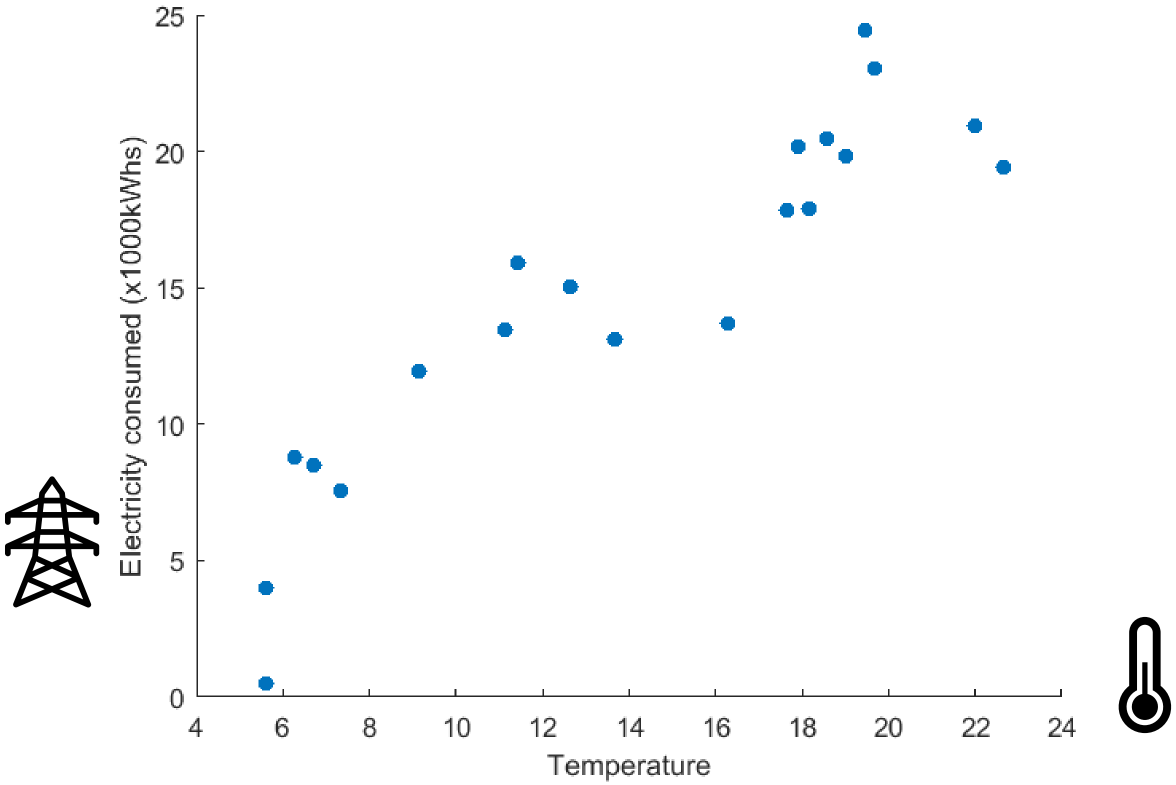
Motivation Behind Correlation

Greater understanding of data

Can hint at potential causal relationships

Business decision based on correlation: increase electricity production when temperature increases

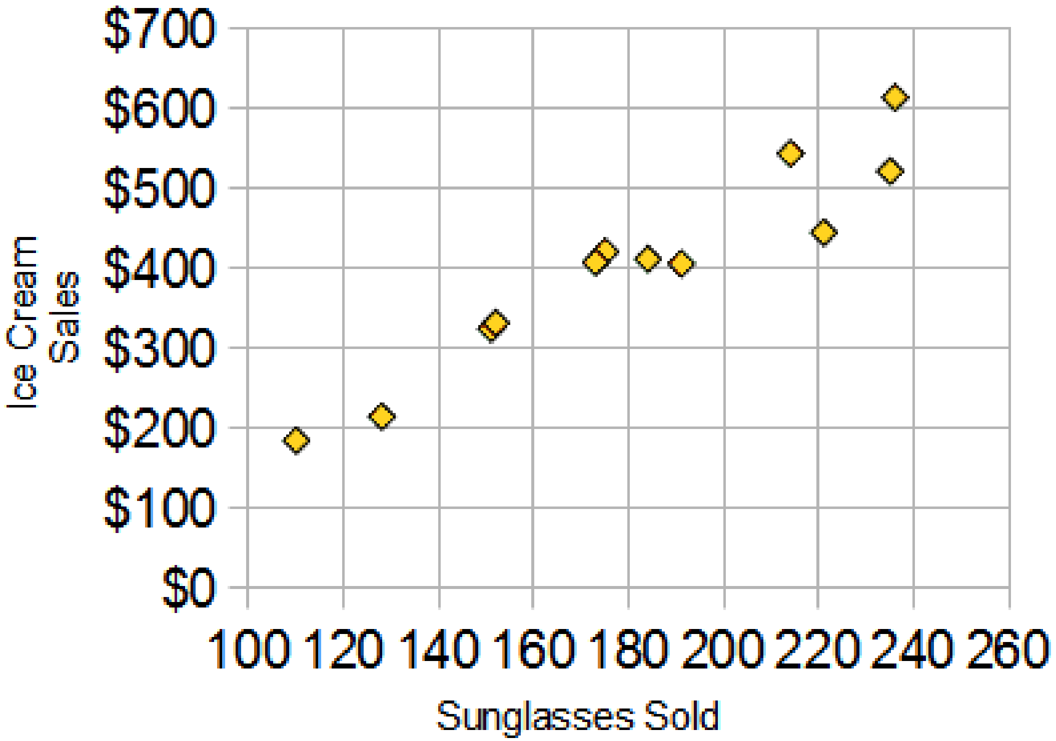
colleration != causality(cause and effect)

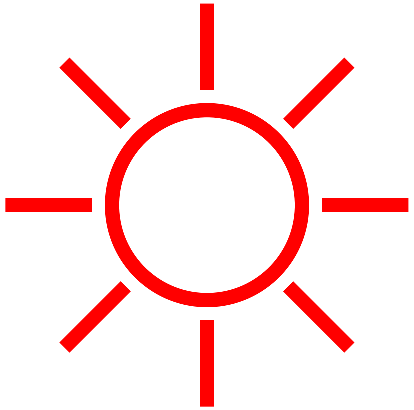


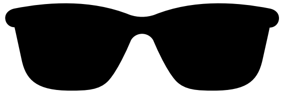
Example of Correlated Variables

strong linear relation, however not causation

Correlation does not necessarily imply causality!







Example rank correlation

"If a university has a higher-ranked football team, then is it likely to have a higher-ranked basketball team?"

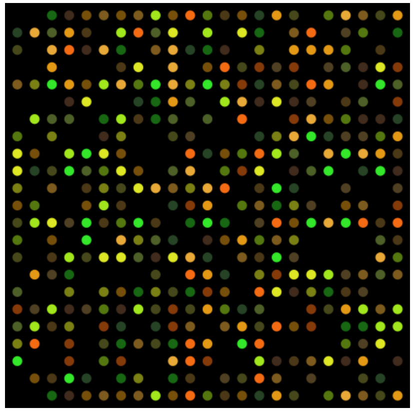
Football ranking	University team
1	Melbourne
2	Monash
3	Sydney
4	New South Wales
5	Adelaide
6	Perth

Basketball ranking	University team
1	Sydney
2	Melbourne
3	Monash
4	New South Wales
5	Perth
6	Adelaide

Biology example - Microarray data

Each chip contains thousands of tiny probes corresponding to the genes (20k - 30k genes in humans). Each probe measures the activity (expression) level of a gene

	Gene 1 expression	Gene 2 expression	…	Gene 20K expression
	0.3	1.2	…	3.1



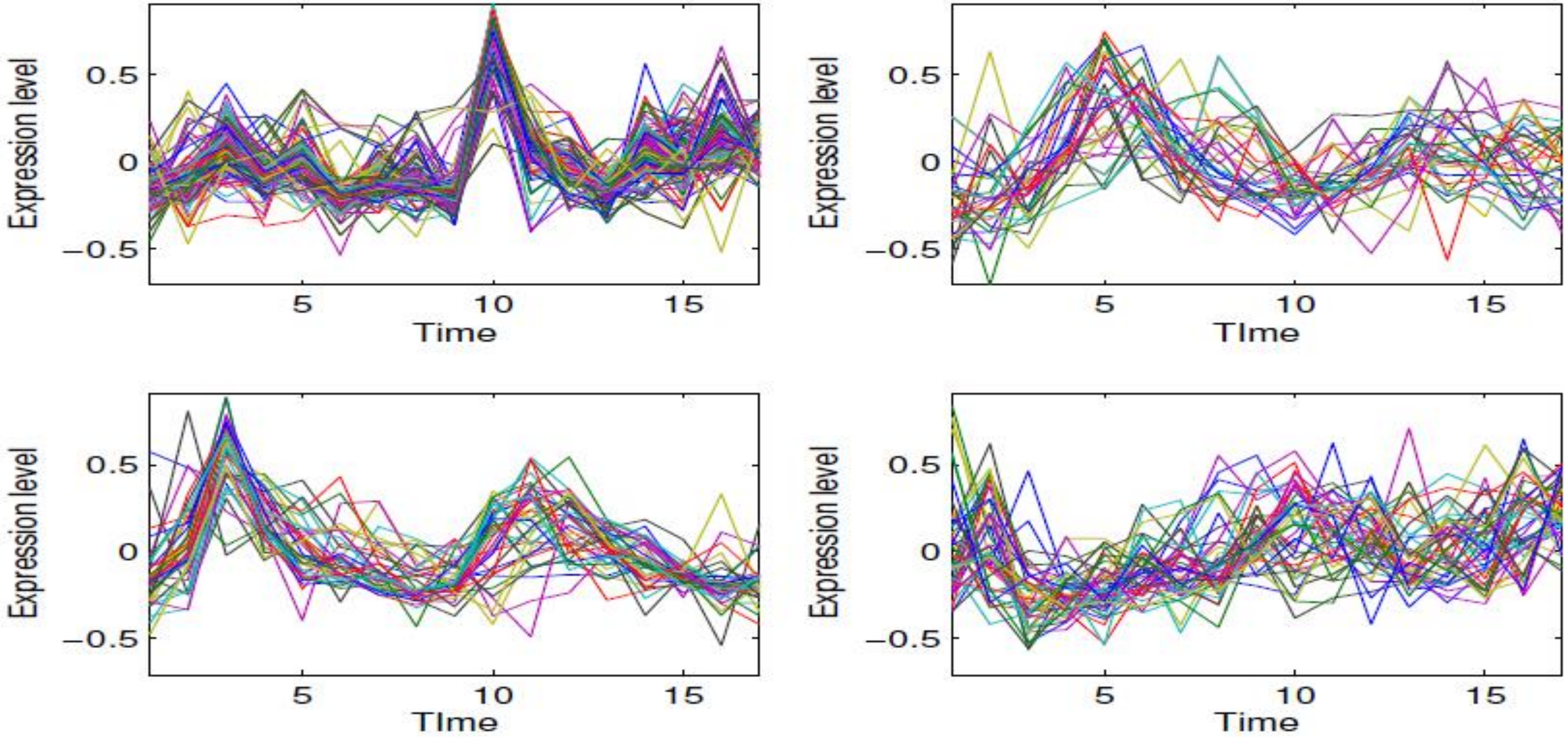
Microarray dataset

	Gene 1	Gene 2	Gene 3	…	Gene n
Time 1	2.3	1.1	0.3	…	2.1
Time 2	3.2	0.2	1.2	…	1.1
Time 3	1.9	3.8	2.7	…	0.2
…	…	…	…	…	…
Time m	2.8	3.1	2.5	…	3.4

- Each row represents measurements at some time
- Each column represents levels of a gene

Correlation analysis on Microarray data

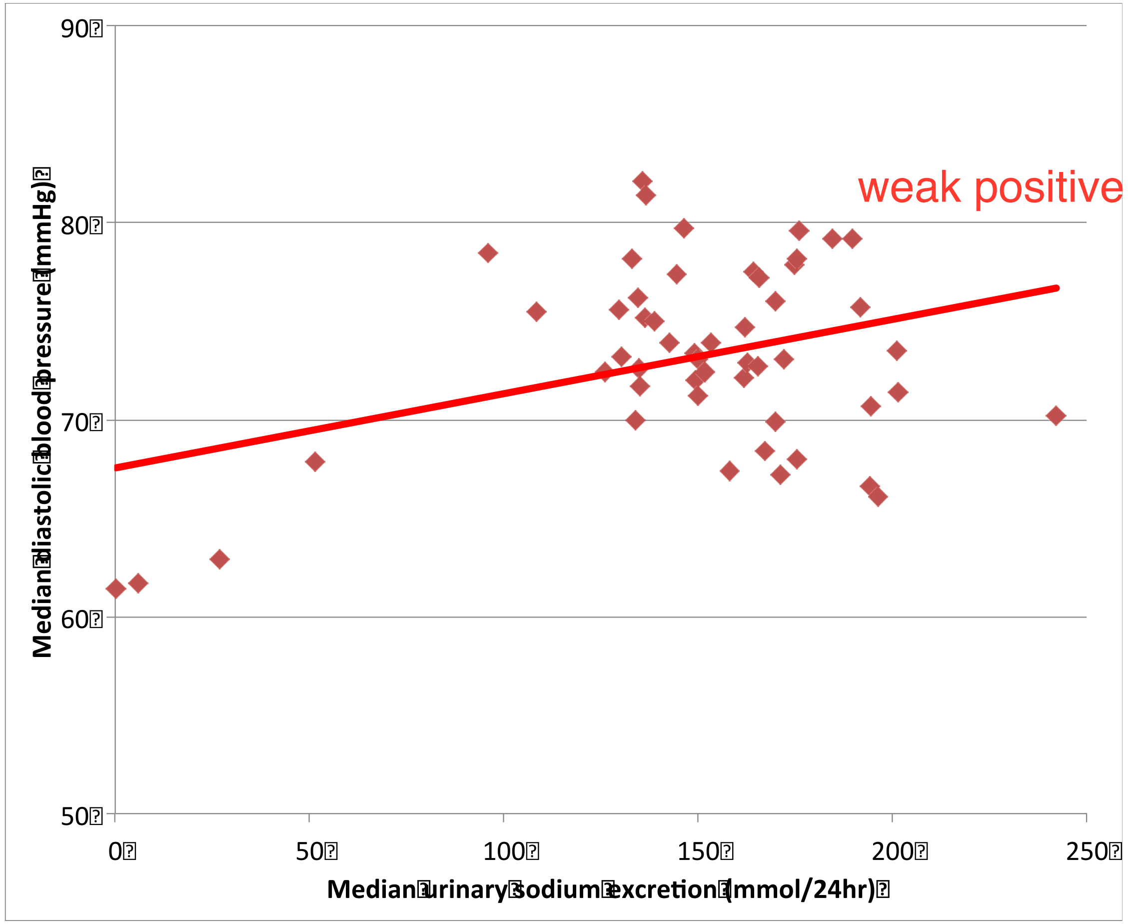
Can reveal genes that exhibit similar patterns ⇨ similar or related functions ⇨ Discover functions of unknown genes



Association Between Salt and High Blood Pressure

Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion.

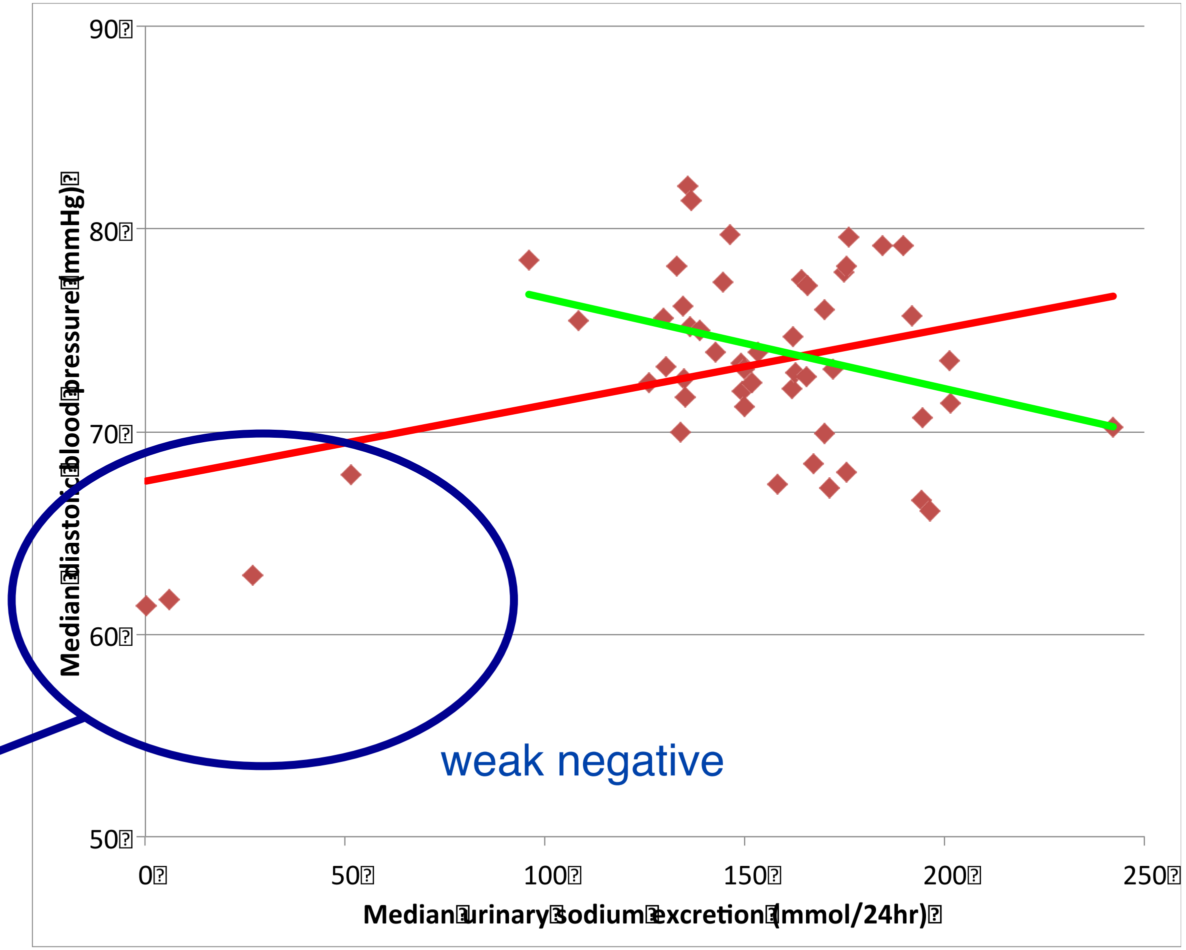
British Medical Journal; 297: 319-328, 1988.



Or Does It!?

Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion.

British Medical Journal; 297: 319-328, 1988.



If we exclude these four 'outliers', which are non-industrialised countries with non-salt diets, we get a quite different result!

Spurious Correlation

Correlation ≠ Causality

https://assets.businessinsider.com/real-maps-ridiculous-correlations-2014-11?jwsource=cl

https://images.app.goo.gl/FVr8BhxWmQMxCB5f7

Why is correlation important?

Discover relationships

One step towards discovering causality **A causes B**

Example: Smoking causes lung cancer

- Feature ranking: for building better predictive models

  A good feature to use, is a feature that has high correlation with the outcome we are trying to predict

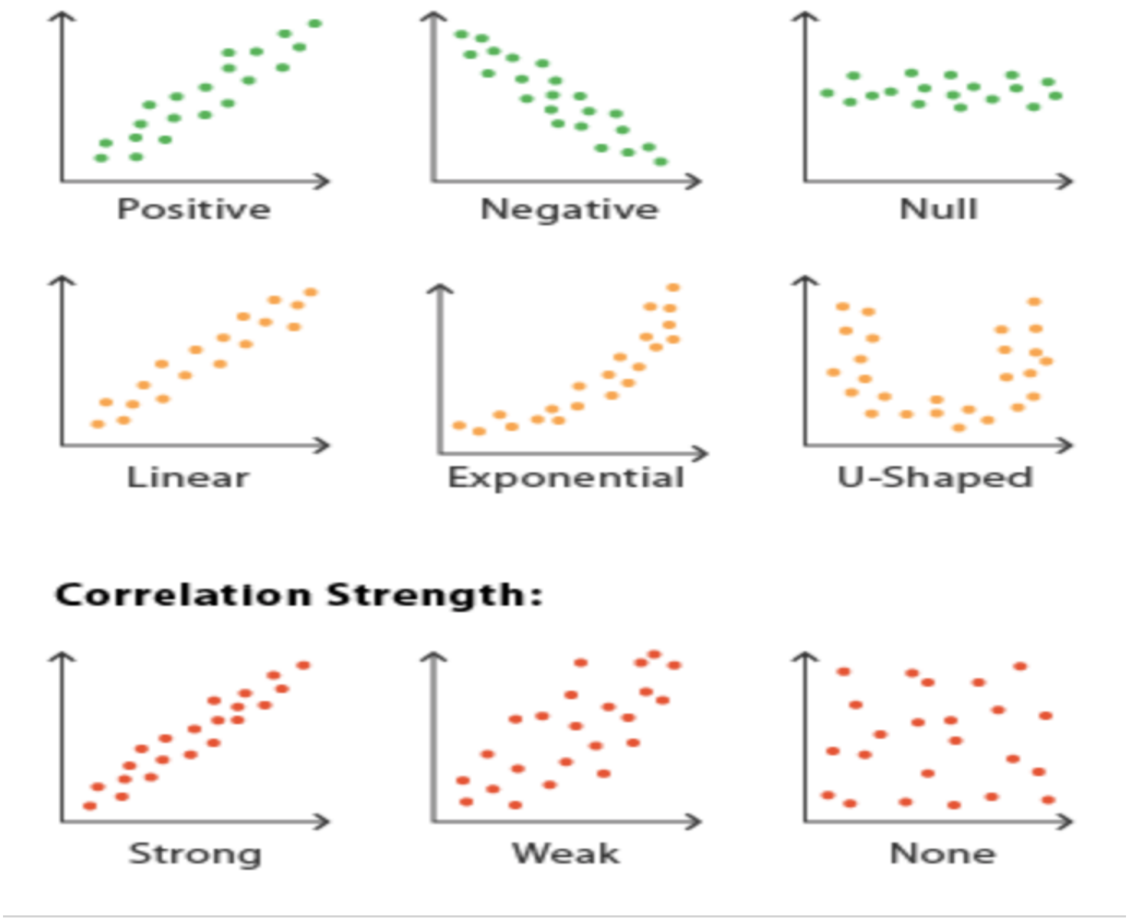
Break

Pearson Correlation

COMP20008

School of Computing and Information Systems

Correlation



Problems of Euclidean distance

Objects can be represented with different measure scales

	Day 1	Day 2	Day 3	…	Day m
Temperature	20	22	16	…	33
#Ice-creams	50223	55223	45098	…	78008
#Electricity	102034	105332	88900	…	154008

d(temp,ice-cr)= 540324

d(temp,elect)= 12309388

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

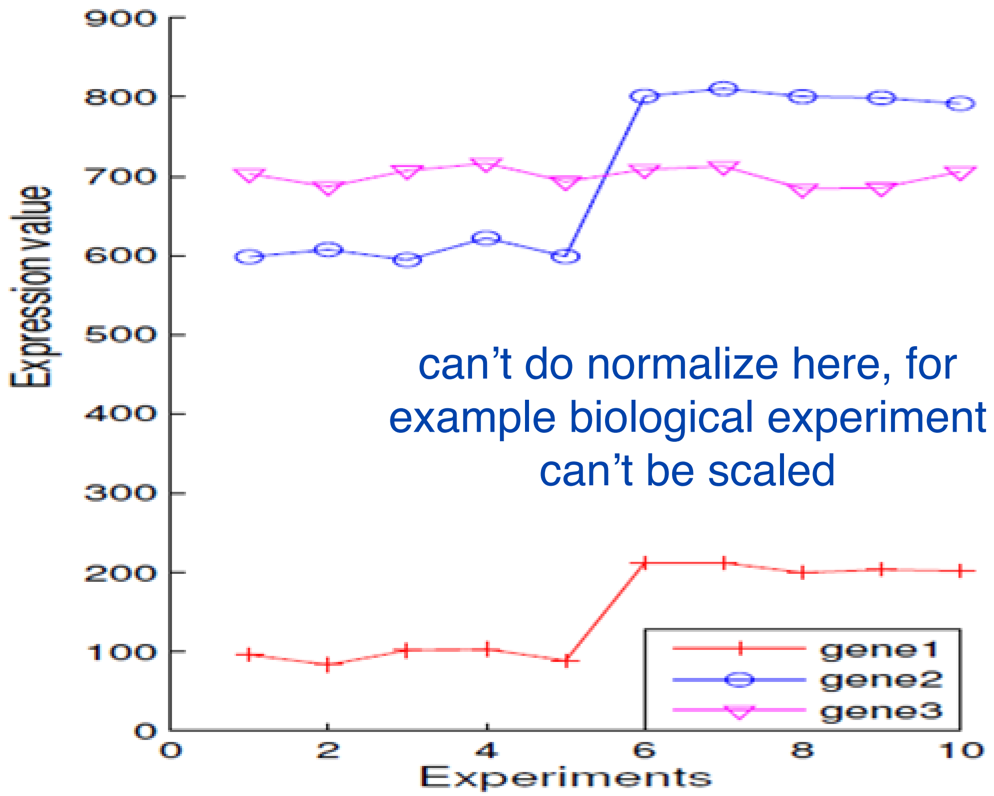
$$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Euclidean distance: does not give a clear intuition about how well variables are correlated

Problems of Euclidean distance

Cannot discover variables with similar behaviours/dynamics but at different scale

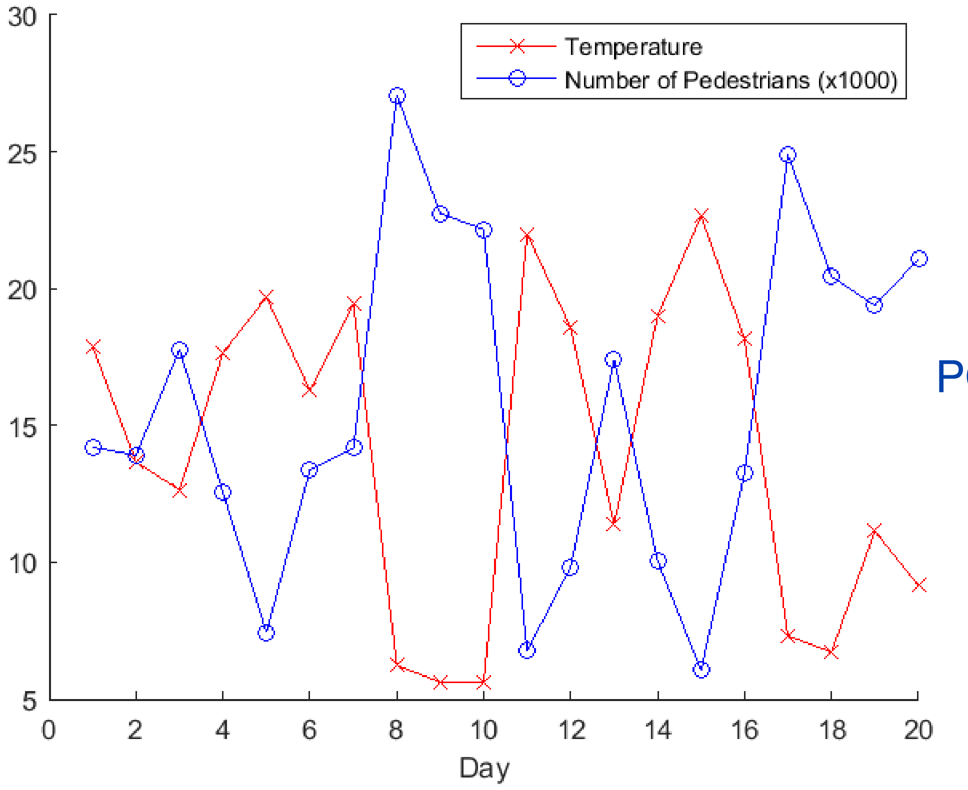
eucludean doesn't give the intuition on how data move or which direction



can't do normalize here, for example biological experiment can't be scaled

Problems of Euclidean distance

Cannot discover variables with similar behaviours/dynamics but in the opposite direction (negative correlation)



POLARITY

Assessing linear correlation – Pearson correlation

We will define a correlation measure r_{xy}, assessing samples from two features x and y

- Assess how close their scatter plot is to a straight line (a linear relationship)

Pearson Product Moment Correlation

Range of r_{xy} lies within [-1,1]:

- 1 for perfect positive linear correlation
- -1 for perfect negative linear correlation
- 0 means no correlation
- Absolute value $|r|$ indicates strength of linear correlation

Pearson's correlation coefficient (r)

$$r = r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \bar{x})^2) \times (\sum_{i=1}^{n}(y_i - \bar{y})^2)}}$$

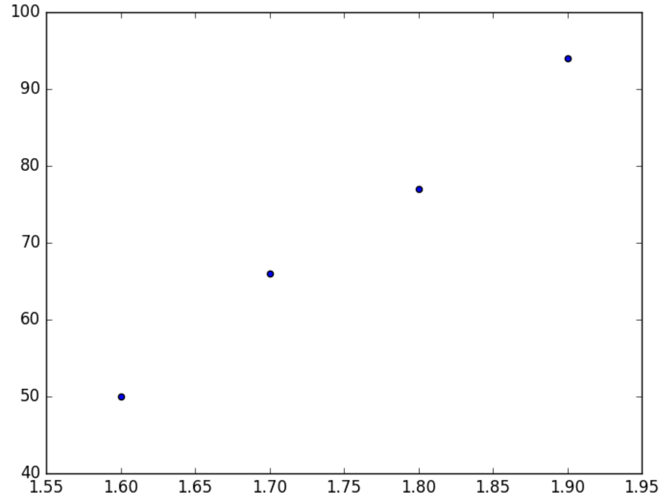
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$r = r_{xy} = \frac{\sum_{i=1}^{n} x_i^* \times y_i^*}{\sqrt{(\sum_{i=1}^{n} {x_i^*}^2) \times (\sum_{i=1}^{n} {y_i^*}^2)}} \qquad x_i^* = x_i - \bar{x} \qquad y_i^* = y_i - \bar{y}$$

Pearson coefficient example

Height (x)	Weight (y)
1.6	50
1.7	66
1.8	77
1.9	94

- How do the values of x and y move (vary) together?
- Big values of x with big values of y?
- Small values of x with small values of y?



Pearson coefficient example

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \bar{x})^2).(\sum_{i=1}^{n}(y_i - \bar{y})^2)}}$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1.6	50	-0.15	-21.75	3.2625	0.0225	473.0625
1.7	66	-0.05	-5.75	0.2875	0.0025	33.0625
1.8	77	0.05	5.25	0.2625	0.0025	27.5625
1.9	94	0.15	22.25	3.3375	0.0225	495.0625
$\bar{x}$ = 1.75	$\bar{y}$ = 71.75			$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ = 7.15	$\sum_{i=1}^{n}(x_i - \bar{x})^2$ = 0.05	$\sum_{i=1}^{n}(y_i - \bar{y})^2$ = 1028.75

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \bar{x})^2).(\sum_{i=1}^{n}(y_i - \bar{y})^2)}} = \frac{(7.15)}{\sqrt{0.05 \text{ x } 1028.75}} = 0.996933$$

linear strong positive colleration

Interpreting Pearson correlation values

In general it depends on your domain of application. Jacob Cohen has suggested

- 0.5 is large
- 0.3-0.5 is moderate
- 0.1-0.3 is small
- less than 0.1 is trivial

Properties of Pearson's correlation

colleration isn't affected by scaling

Range within [-1,1]

Is sensitive to outliers

Can only detect linear relationships

$$y = a.x + b + noise$$

Cannot detect non-linear relationships

$$y = x^3 + noise$$

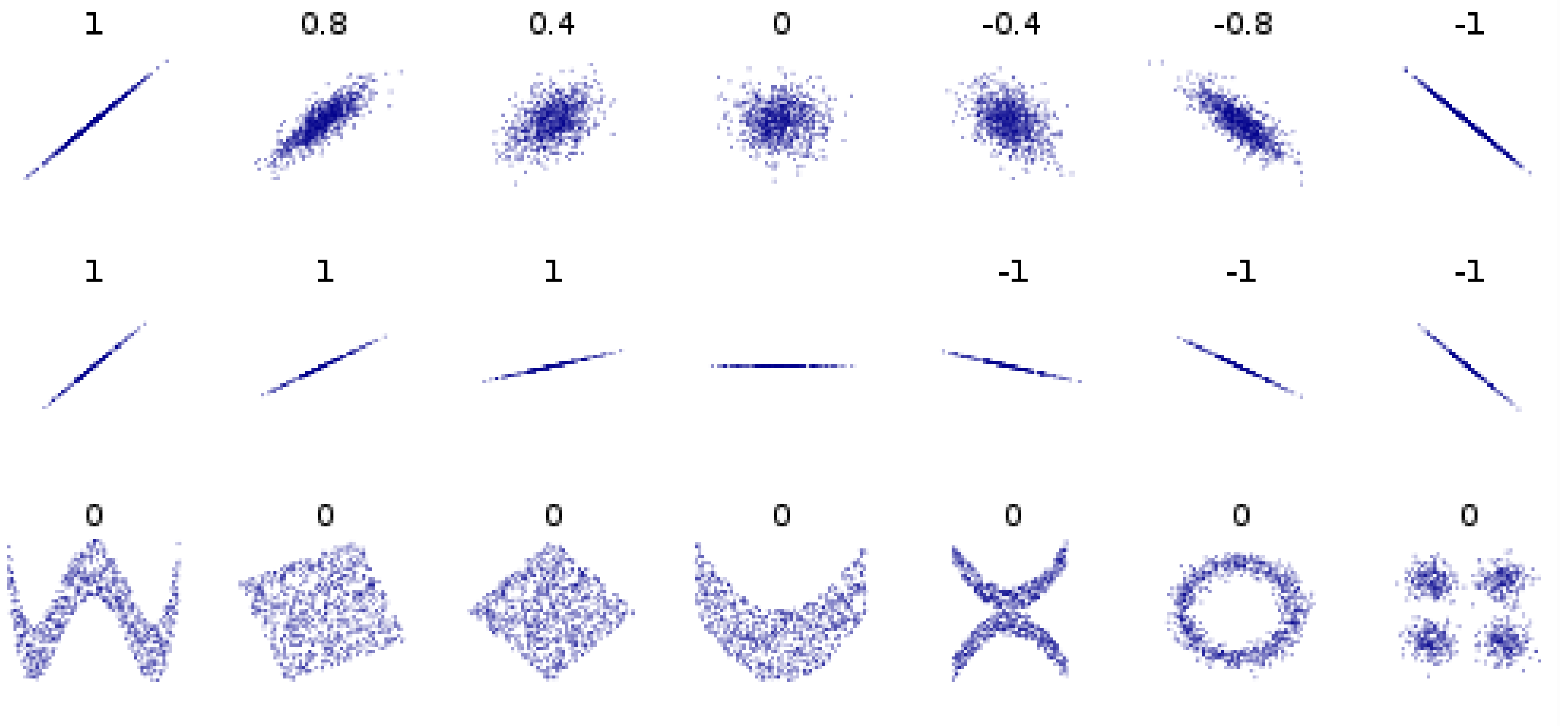
Scale invariant: r(x,y)= r(x, Ky)

- Multiplying a feature's values by a constant K makes no difference

Location invariant: r(x,y)= r(x, K+y)

- Adding a constant K to one feature's values makes no difference

Pearson correlation examples



https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

2016 exam question 2a)

2. a) Richard is a data wrangler. He does a survey and constructs a dataset recording average time/day spent studying and average grade for a population of 1000 students:

Student Name	*Average time per day studying*	*Average Grade*
...	...	...

i) (3 marks) Richard computes the Pearson correlation coefficient between *Average time per day studying* and *Average grade* and obtains a value of 0.85. He concludes that more time spent studying causes a student's grade to increase. Explain the limitations with this reasoning and suggest two alternative explanations for the 0.85 result.

the conclusion is too strong, it's not causation relationship
might be because of outliers, third variables or cofounding causes like student IQ

Summary

Correlation <> Causality

- http://tylervigen.com/spurious-correlations

Google trends correlation

- https://trends.google.com/trends/

Break

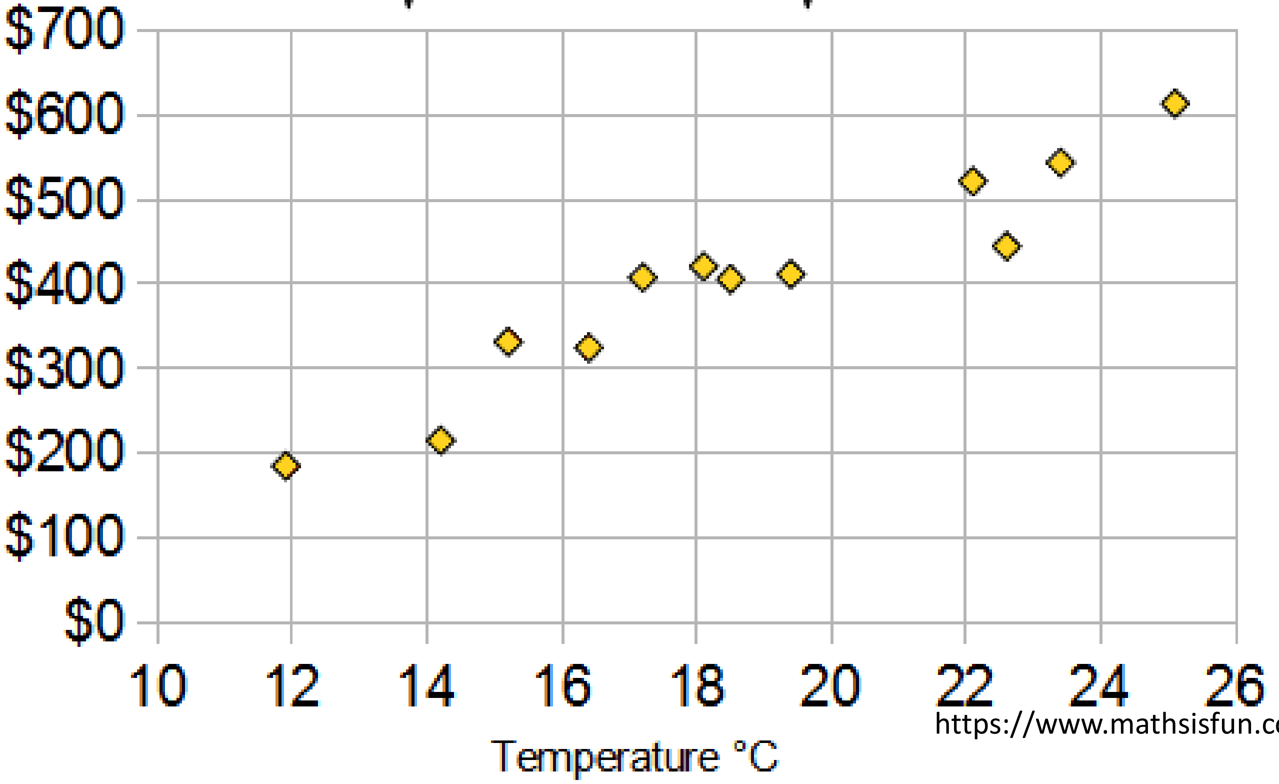
Mutual Information

COMP20008

School of Computing and Information Systems

Recap: Pearson correlation – assess linear correlation between two features

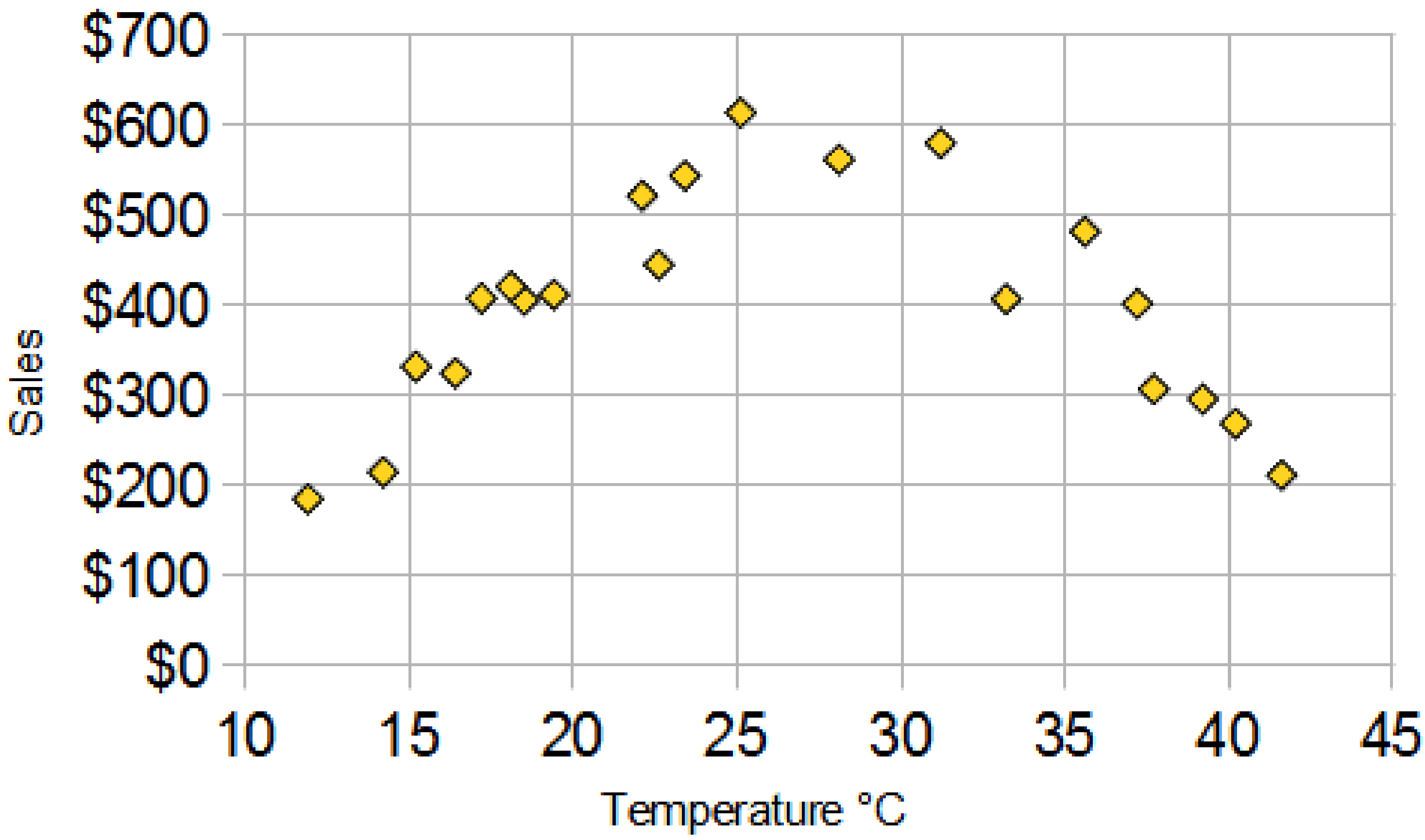
$$r = r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$



https://www.mathsisfun.com/data/correlation.html

What about non-linear correlation?

Pearson correlation is not suitable for this scenario (value less than 0.1)



https://www.mathsisfun.com/data/correlation.html

Mutual Information

$$MI(X,Y) = H(Y) - H(Y|X)$$
$$= H(X) - H(X|Y)$$

Entropy

Conditional Entropy

Then.. After MI we will discuss NMI

Entropy

Mutual Information

Entropy – intuition

Entropy is a measure used to quantify the amount of uncertainty in an outcome

Randomly select an element from

A. {1,1,1,1,1,1,1,2} versus

B. {1,1,2,2,3,3,4,4}

In which case are you more uncertain about the outcome of the selection? Why?

- More uncertain, less predictable => high entropy (b.)
- Less uncertain, more predictable => low entropy (a.)

Another example

Consider the sample of all people in this subject. Each person is labelled young (<30 years) or old (>=30 years)

Randomly select a person and inspect whether they are young or old.

- How surprised am I likely to be by the outcome? low

Suppose I repeat the experiment using a random sample of people catching the train to the city in peak hour?

- How surprised am I likely to be by the outcome? high

A recap on logarithms (to the base 2)

$y = \log_2 x$ (y is the solution to the question "To what power do I need to raise 2, in order to get x?'')

2*2*2*2=16 which means $\log_2 16 = 4$ (16 is 2 to the power 4)

$\log_2 32 = 5$

$\log_2 30 = 4.9$

$\log_2 1.2 = 0.26$

$\log_2 0.5 = -1$

In what follows, we'll write log instead of $\log_2$ to represent binary log

Entropy

Given a feature $\boldsymbol{X}$, assuming $\boldsymbol{X}$ has k number of categories (bins), then $H(\boldsymbol{X})$ is its entropy.

$$H(\boldsymbol{X}) = -\sum_{i=1}^{k} p_i \log_2 p_i$$

p_i: proportion of points in the i-th category (bin)

For example:

X: Flipping a special coin, head: 75%, tail: 25%

$$H(X) = -[(p_H \log_2 p_H) + (p_T \log_2 p_T)] = -[0.75 \log_2 0.75 + 0.25 \log_2 0.25] = \mathbf{0.81}$$

Entropy practice example

A	B	B	A	C	C	C	C	A

We have 3 categories (A,B,C) for a feature $\boldsymbol{X}$

9 objects, each object is in one category.

$$H(\boldsymbol{X}) = -\sum_{i=1}^{k} p_i \log_2 p_i$$

What is the entropy of this sample data of 9 objects?

P_A=3/9 P_B=2/9 P_C=4/9

Answer: $H(\boldsymbol{X}) = 1.53$

How would you compute the entropy for the “Likes to sleep” feature?

Person	Likes to sleep
1	Yes
2	No
3	Maybe
4	Never
5	Yes
6	Yes
7	Never
8	Yes

$\boldsymbol{X}$: Likes to sleep

$$H(\boldsymbol{X}) = -\sum_{i=1}^{k} p_i \log_2 p_i$$

URL: https://go.unimelb.edu.au/8b7p

Properties of entropy

$H(X) \geq 0$

Entropy is maximized for uniform distribution (highly uncertain what value a randomly selected object will have)

Entropy – when using log base 2 – measures uncertainty of the outcome in bits. This can be viewed as the information associated with learning the outcome

Variable discretisation

Pre-processing: continuous features are first discretised into bins (categories). E.g. small [0,1.4], medium (1.4,1.8), large [1.8,3.0]

Object	Height	Discretised Height
1	2.03	large
2	1.85	large
3	1.23	small
4	1.31	small
5	1.72	medium
6	1.38	small
7	0.94	small

Variable discretisation: Techniques

Equal-width bin

Divide the range of the continuous feature into equal length intervals (bins). If speed ranges from 0-100, then the 10 bins are [0,10), [10,20), [20,30), …[90,100]

Equal-frequency bin

Divide range of continuous feature into equal frequency intervals (bins). Sort the values and divide so that each bin has roughly same number of objects

Domain knowledge: assign thresholds manually

Car speed:

- 0-40: slow
- 40-60: medium
- >60: high

Discretisation example

Given the values 2, 2, 3, 10, 13, 15, 16, 17, 19, 19, 20, 20, 21

A 3-bin **equal width** discretization

- Bin-width: $\frac{21-2}{3} = 6.333$
- $[2, 8.333), [8.333, 14.666), [14.666, 21]$

Discretisation example

Given the values 2, 2, 3, 10, 13, 15, 16, 17, 19, 19, 20, 20, 21

A 3-bin **equal frequency** discretization

- $[2, 13), [13, 19), [19, 21]$ – by hand
- $(1.999, 13], (13, 19], (19, 21]$ – pandas

pandas.qcut

`pandas.qcut(x, q, labels=None, retbins=False, precision=3, duplicates='raise')` [source]

Quantile-based discretization function.

Discretize variable into equal-sized buckets based on rank or based on sample quantiles. For example 1000 values for 10 quantiles would produce a Categorical object indicating quantile membership for each data point.

```python
import pandas as pd
X = pd.Series([2, 2, 3, 10, 13, 15, 16, 17, 19, 19, 20, 20, 21])
pd.qcut(X, q=3)
```

```
0     (1.999, 13.0]
1     (1.999, 13.0]
2     (1.999, 13.0]
3     (1.999, 13.0]
4     (1.999, 13.0]
5      (13.0, 19.0]
6      (13.0, 19.0]
7      (13.0, 19.0]
8      (13.0, 19.0]
9      (13.0, 19.0]
10     (19.0, 21.0]
11     (19.0, 21.0]
12     (19.0, 21.0]
dtype: category
Categories (3, interval[float64, right]): [(1.999, 13.0] <
(13.0, 19.0] < (19.0, 21.0]]
```

Break

Mutual Information

$$MI(X,Y) = H(Y) - H(Y|X)$$
$$= H(X) - H(X|Y)$$

Entropy

Conditional Entropy

Then.. After MI we will discuss NMI

Conditional entropy - intuition

Suppose I randomly sample a person. I check if they wear glasses – how surprised am I by their age given their wearing glasses status?

Person	WearGlasses(X)	Age (Y)
1	No	young
2	No	young
3	No	young
4	No	young
5	Yes	old
6	Yes	old
7	Yes	old

Conditional entropy H(Y|X)

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

Measures how much information is needed to describe outcome Y, given that outcome X is known.

Suppose X is Height and Y is Weight.

Object	Height (X)	Weight (Y)
1	big	light
2	big	heavy
3	small	light
4	small	light
5	small	light
6	small	light
7	small	heavy

Conditional entropy H(Y|X)

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

$$H(Y|X) = {}^{2}/_{7} * H(Y|X = big) + {}^{5}/_{7} * H(Y|X = small)$$

$$H(Y|X = x) = -\sum_{y \in Y} p(y|x) \log_2 p(y|x)$$

$$H(Y|X = big) = -\left[{}^{1}/_{2} \log_2 {}^{1}/_{2} + {}^{1}/_{2} \log_2 {}^{1}/_{2}\right] = 1$$

$$H(Y|X = small) = -\left[{}^{4}/_{5} \log_2 {}^{4}/_{5} + {}^{1}/_{5} \log_2 {}^{1}/_{5}\right] = 0.721$$

$$H(Y|X) = {}^{2}/_{7} * 1 + {}^{5}/_{7} * 0.721 = 0.801$$

Object	Height (X)	Weight (Y)
1	big	light
2	big	heavy
3	small	light
4	small	light
5	small	light
6	small	light
7	small	heavy

Mutual information definition

$$MI(X,Y) = H(Y) - H(Y|X)$$
$$= H(X) - H(X|Y)$$

$$H(Y|X) = \sum_{x \in X} p(x)H(Y|X = x)$$

where X and Y are features (columns) in a dataset

- MI (mutual information) is a measure of correlation
  - The amount of information about Y we gain by knowing X, or
  - The amount of information about X we gain by knowing Y

Mutual information example 1

$$MI(X,Y) = H(Y) - H(Y|X)$$

$$H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y)$$

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

$$H(Y) = -\left[{}^5/_7 \log_2 {}^5/_7 + {}^2/_7 \log_2 {}^2/_7\right] = 0.8631$$

$$H(Y|X) = {}^6/_7 * H(Y|X = small) + {}^1/_7 * H(Y|X = big)$$

$$= {}^6/_7 * \left[-{}^5/_6 \log_2 {}^5/_6 - {}^1/_6 \log_2 {}^1/_6\right] + {}^1/_7 * [-1 * \log_2 1] = 0.5571$$

$$MI(X,Y) = 0.8631 - 0.5571 = 0.3059$$

Object	Height (X)	Weight (Y)
1	small	light
2	big	heavy
3	small	light
4	small	light
5	small	light
6	small	light
7	small	heavy

Mutual information example 2

$$MI(X,Y) = H(Y) - H(Y|X)$$

$$H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y)$$

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

$$H(Y) = 1.379$$

$$H(Y|X) = 0.965$$

$$MI(X,Y) = H(Y) - H(Y|X) = 0.414$$

Object	Height (X)	Weight (Y)
1	big	light
2	big	heavy
3	small	light
4	small	jumbo
5	medium	light
6	medium	light
7	small	heavy

Properties of Mutual Information

The amount of information shared between two variables X and Y

MI(X,Y)

- large: X and Y are highly correlated (more dependent)
- small: X and Y have low correlation (more independent)

$0 \leq MI(X,Y) \leq \infty$

Sometimes also referred to as ‘Information Gain’

Break

Mutual information: normalisation

- MI(X,Y) is always at least zero, may be larger than 1
- In fact, one can show it is true that
  - $0 \leq \text{MI}(X,Y) \leq \min(H(X),H(Y))$
- Thus if want a measure in the interval [0,1], we can define normalised mutual information (NMI).
  - $\text{NMI}(X,Y) = \text{MI}(X,Y) / \min(H(X),H(Y))$
  - $\text{NMI}(X,Y) = \text{MI}(X,Y) / \max(H(X),H(Y))$
  - $\text{NMI}(X,Y) = \text{MI}(X,Y) / \text{mean}(H(X),H(Y))$ balance approach

NMI(X,Y)

- large: X and Y are highly correlated (more dependent)
- small: X and Y have low correlation (more independent)

Normalised Mutual Information

Example 1:

$$MI = 0.3059$$
$$NMI = 0.517$$

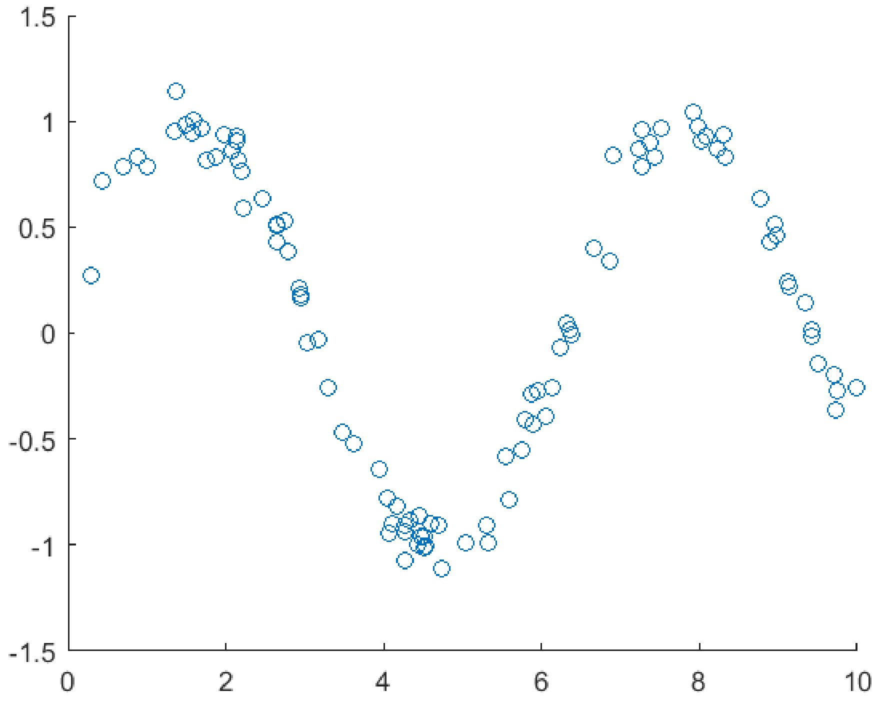
Object	Height (X)	Weight (Y)
1	small	light
2	big	heavy
3	small	light
4	small	light
5	small	light
6	small	light
7	small	heavy

Example 2:

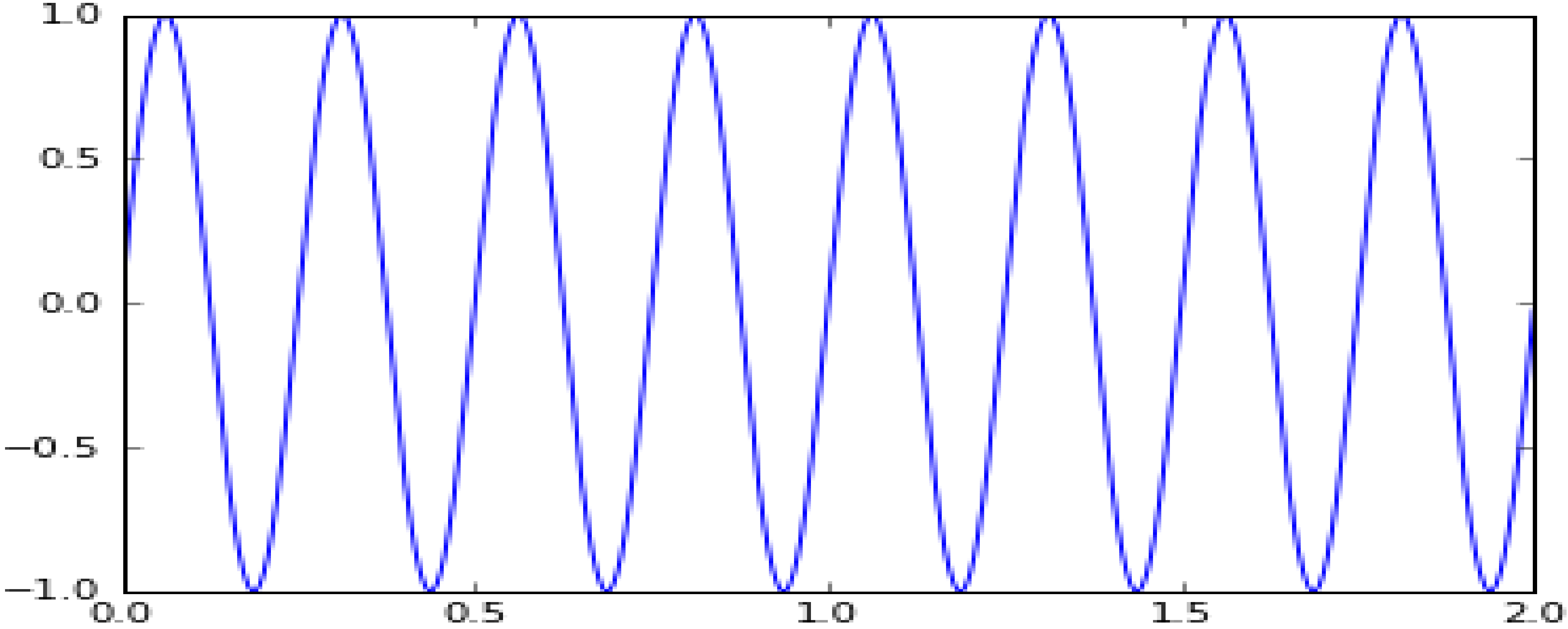
$$MI = 0.414$$
$$NMI = 0.300$$

Object	Height (X)	Weight (Y)
1	big	light
2	big	heavy
3	small	light
4	small	jumbo
5	medium	light
6	medium	light
7	small	heavy

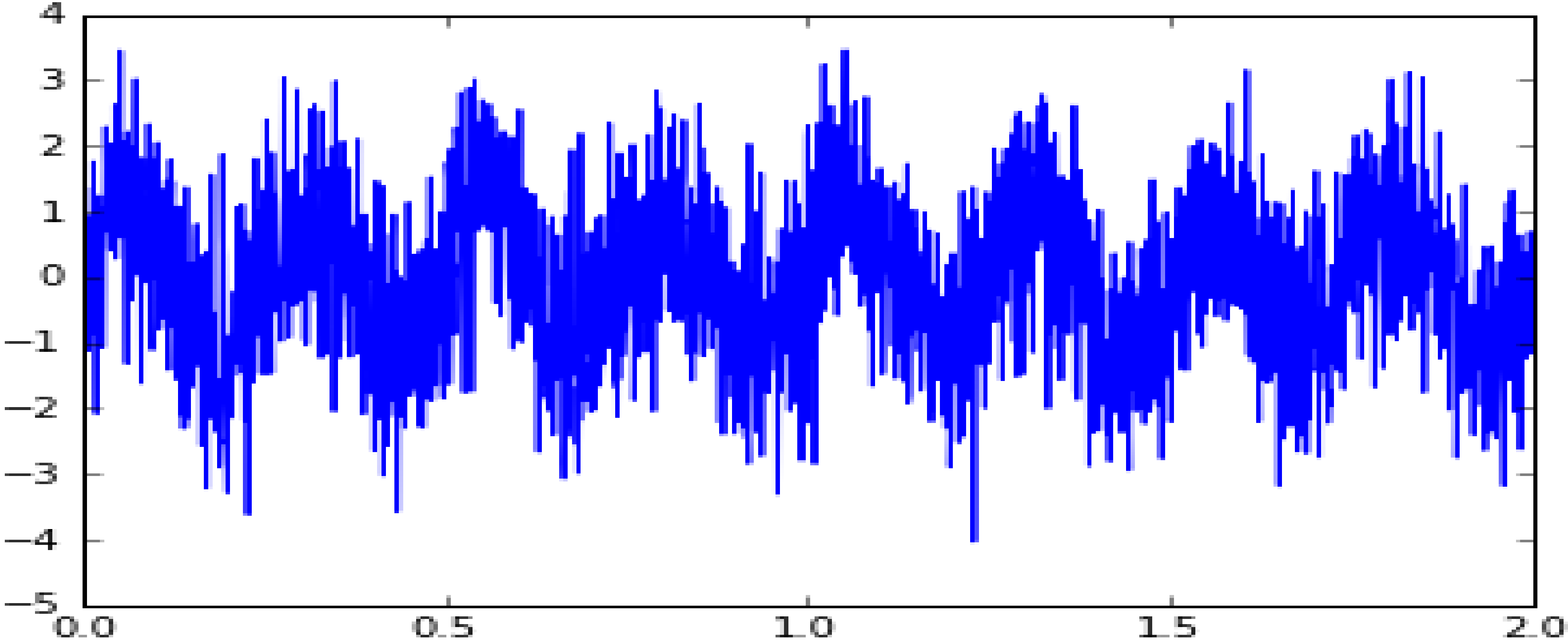
Pearson correlation=-: -0.0864 Normalised mutual information (NMI)= 0.43 (3-bin equal spread bins)



Pearson correlation=-0.1
Normalised mutual information (NMI)=0.84



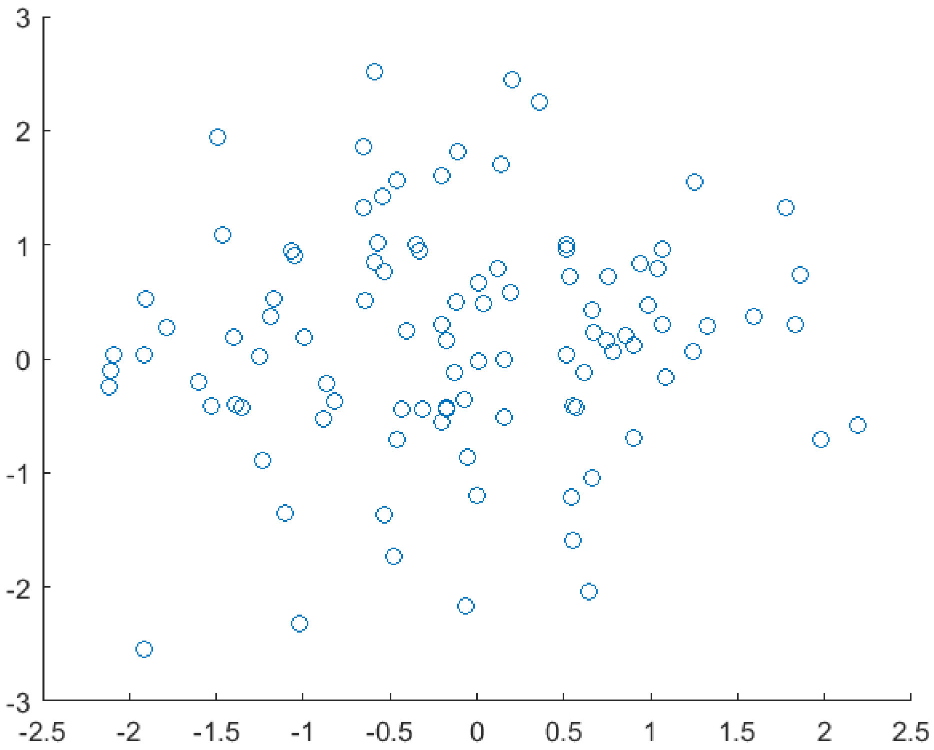
Pearson correlation=-0.05
Normalised mutual information (NMI)=0.35



Examples

Pearson?

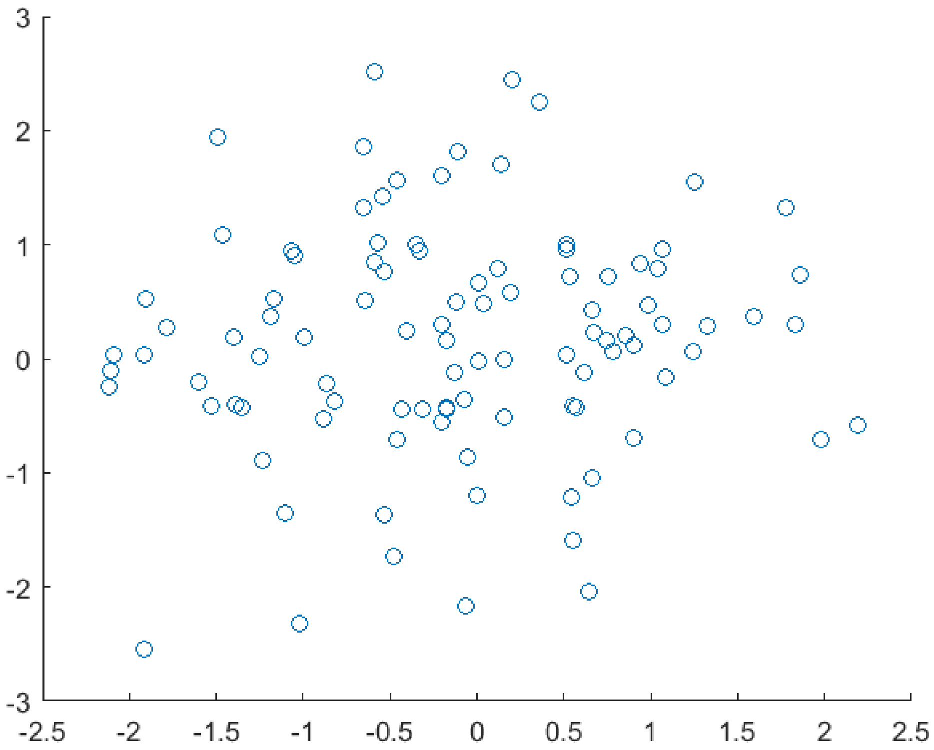
NMI?



Examples

Pearson: 0.08

NMI: 0.009



Computing MI with class features

Identifying features that are highly correlated with a class feature

HoursSleep	HoursExercise	HairColour	HoursStudy	Happy (class feature)
12	20	Brown	low	Yes
11	18	Black	low	Yes
10	10	Red	medium	Yes
10	9	Black	medium	Yes
10	10	Red	high	No
7	11	Red	high	No
6	15	Brown	high	No
2	13	Brown	high	No

Compute MI(HoursSleep, Happy), MI(HoursExercise, Happy),
and MI(HoursStudy, Happy), MI(HairColour, Happy). Retain most predictive feature(s)

Advantages and disadvantages of MI

Advantages

- Can detect both linear and non-linear dependencies (unlike Pearson)
- Applicable and very effective for use with discrete features (unlike Pearson correlation)

Disadvantages

- If feature is continuous, it first must be discretised to compute mutual information. This involves making choices about what bins to use.
- This may not be obvious. Different bin choices will lead to different estimations of mutual information.

Choose the statement which is False

Mutual information can be used to assess how much one feature is associated with another

Mutual information can detect the existence of non-linear relationships in the data

If you have non-linear data, you can use Pearson correlation

Mutual information doesn't indicate the "direction" of a relationship, just its strength

URL: https://go.unimelb.edu.au/8b7p

Question 2aii) from 2016 exam

a) Richard is a data wrangler. He does a survey and constructs a dataset recording average time/day spent studying and average grade for a population of 1000 students:

Student Name	Average time per day spent studying	Average Grade
…	….	….

ii) Richard separately discretises the two features *Average time per day spent studying* and *Average grade*, each into 2 bins. He then computes the normalised mutual information between these two features and obtains a value of 0.1, which seems surprisingly low to him. Suggest two reasons that might explain the mismatch between the normalised mutual information value of 0.1 and the Pearson Correlation coefficient of 0.85. Explain any assumptions made.

End

Acknowledgements

Materials are partially adopted from …

- Previous COMP2008 slides
- Pauline Lin, Chris Ewin, Uwe Aickelin and others

See also

- Correlation and Causality – Don’t Confuse Them (13-2): https://www.youtube.com/watch?v=dDlgAiTHNqI