Fundaments of Machine learning for and with engineering applications

Enrico Riccardi[1]

Department of Mathematics and Physics, University of Stavanger (UiS).[1]

Sep 10, 2025

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

Correlated variables

Let's make a few example of correlated variables:

1. Basketball playing skill and hight
2. Age and hair loss
3. Oil and gas production vs oil price
4. Wind and wind turbine efficiency
5. Wind and energy production
6. Rain and energy production
7. Solar irradiance and energy production

How to approach

There are a set of questions that one shall pose when relating two variables.

1. Statistical dependence

Two variables have their distribution and, even if very similar, are unrelated

1. Causal dependence

Two variables depend on each other.

Discussion point

How does this relate to soft and hard modeling?

Visualization

- Scatterplots (MatPLotLib)

It is one of the simplest ways to graphically display their relationship (it can be 3D).

- Heathmaps (seabon)
- Correlation matrix plots (pandas)

Correlation

The covariance or joint variance between two random variables is an extension of the concept of variance.

$$Cov[XY] = \sigma_{xy} = E[(X - \bar{Z})(Y - \bar{Y})] = \frac{1}{N-1}/sumNi = 1(x_i - \bar{X})(y_i - \bar{Y}) = \frac{N}{N-1}(E[XY] - E$$

- Generalization of variance
- Consider the covariance of a variable with itself

Cov [XY]

Variance: always positive Covariance: positive or negative

Correlation Analysis

- The correlation between two random variables is a measure of the strength of their linear relationship
- Parametric Correlation

Measures a linear (Pearson) dependence between two variables (x and y) is known as parametric correlation test because it depends

xxx

Pearson's ρ Value

The correlation coefficient ρ between to

- It is closely linked to the concept of covariance:
- Assumes normal distribution
- ρ ranges between -1 (per

xxxx

Correlation examples

Strong negative, weak and Strong Positive -1 0 +1

Example in python

python

note: the implementation gives the covariance matrix (Cov(x, y)

Correlation Coefficient - Interpretation

What do I do with this number

The variation of x *explains* a corr coef variation of y

What does it mean to explain?

Correlation is not causation!!

Regression Coefficient - interpretation

Plotting two variables on a scatter plot, a streight line means a $\rho = +/-1$

Regression Coefficient - examples

Python

Regression Coefficient - limitation

Anscorbe's Quartet: Four different pairs of variables

4 distributions with the same means (7.5), standard deviation (4.12), correlation (0.81) and regression line (y=3 + 0.5x)

plots

Spearman Rank Correlation

THe Spearman correlation evaluates a monotonic relationship between two variables - Continuous or Ordinal and it is based on the *ranked* values for each variable rather than the raw data-

Plots Not monotonic fails (more advanced methods for this)

Spearman Rank Correlation

- Rank correlation compares the ranks (ordering)

Pearson vs Spearman

Figures

Spearman Rank Correlation

- Calculated the same way as the Person correlation coefficient but using ranks instead of values

missing

Spearman Rank Correlation example

python

Kendall tau

quick, not so much used

Correlation does NOT indicate Causation

Fancy slides here

Graphing Bivariate Data

Scatterplots between two variables is one of the most

R2 coefficient of determination

Scatterplots combined with Histograms

plots

Uncorrelated and Independent Random Variables

The two random variables X and Y are said to be

Uncorrelate if : $Cov(X, Y) = 0$ **Independent** if : $f_{XY}(x, y) = f_X(x)f_Y(y)$

Correlation versus Dependence

- Uncorrelated Random Variables

Random variables are uncorrelated if there is no linear relationship between them Mathematically, two random variables X and Y are uncorrelated if their covariance is zero. Cov(X,Y)=0

*

Causation Implies Dependency

Statistical dependency: Variables are causally related myst be statistically dependent

- Causal Dependency

Dependency Does not Imply Causation

- Correlation without Causation

Two variables might be correlated (and hence dependent) due to a coincidence, a lurking variable, or confounding factor.

- Common cause

Two variables might be dependent because they are both influenced by a third variable