Linear Correlation

Christian Permann

Motivation: Linear Correlation

- Used to find how two variables correlate linearly
- Good first indication of simple correlation
- Many other methods based on this

Explanation

- Correlation Coefficient: $$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}}$$
- Can take on values between -1 and 1
  - 1 indicates perfect direct linear correlation
  - 0 indicates no linear correlation
  - -1 indicates perfect indirect linear correlation

Applications

- Simple statistical analysis
  - Finding variables that can be represented from one another
  - Finding redundant variables
  - Gaining an idea about data loss if one variable is to be derived from another
  - Evaluating the quality of linear interpolation

Conclusion

- Outsourcing of repeatedly performed computations allows for large speedup
- Linear correlation is perfectly suited for the DataFlow paradigm