Unit 3: Correlation and Regression

Correlation and regression are methods used in ML to determine the strength of an association between two variables (Crawford, 2006) . **Correlation** assesses the linear relationship between 2 variables, providing a measure of both the strength and direction of the relationship (Komorowski et al., 2016). Visual inspection of scatter plots is often applied to aid in identifying the type of relationship. Can be performed in two different methods: parametric like the Pearson product-moment correlation and non-parametric like Spearman rank order correlation and Kendall methods. Importantly, in correlation both variables are treated equally that is neither is considered the predictor or the target variable (Shi & Conrad 2009).

In contrast, Shi and Conrad (2009) describe **regression** analysis as an evaluation of the relationship between one dependent variable and one or more independent variables, implying causality. Regression surpasses correlation by inferring and modeling causal relationships and predicting the dependent variable's value from given independent variables. Unlike correlation, regression relies on several underlying assumptions and encompasses both linear and nonlinear models. Linear regression can be simple or multiple, while nonlinear regression addresses more complex relationships like in logistic regression which extends to binary or discrete dependent variables.

See Jupyter notebooks for Unit 3 activities below:

**References**

Crawford, S.L. (2006). Correlation and regression. *Circulation* 114(19): 2083–2088. DOI: http://dx.doi.org/10.1161/CIRCULATIONAHA.105.586495

Komorowski, M., Marshall, D.C., Salciccioli, J.D., Crutain, Y. (2016). Exploratory Data Analysis. In: Secondary Analysis of Electronic Health Records. Springer, Cham. 185–203. <https://doi.org/10.1007/978-3-319-43742-2_15>

Shi, R., Conrad, S.A.(2009). Correlation and regression analysis. *Statistics for clinicians: Annalysis of Allergy, Asthma & Immunology* 103(4): 34–40. DOI: <http://dx.doi.org/10.1016/s1081-1206(10)60812-5>