Interesting Data Science and Machine Learning explanations

Table of Contents

[Regression Analysis 1](#_Toc114741579)

[Classification 1](#_Toc114741580)

[Performance Metrics 2](#_Toc114741581)

[Statistics 2](#_Toc114741582)

[Dimensionality Reduction 3](#_Toc114741583)

[Python 3](#_Toc114741584)

# Regression Analysis

[r - How to get correlation between two categorical variable and a categorical variable and continuous variable? - Data Science Stack Exchange](https://datascience.stackexchange.com/questions/893/how-to-get-correlation-between-two-categorical-variable-and-a-categorical-variab?noredirect=1&lq=1)

[An overview of correlation measures between categorical and continuous variables | by Outside Two Standard Deviations | Medium](https://medium.com/@outside2SDs/an-overview-of-correlation-measures-between-categorical-and-continuous-variables-4c7f85610365)

[Why is ANOVA equivalent to linear regression? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/175246/why-is-anova-equivalent-to-linear-regression)

[machine learning - Why is Euclidean distance not a good metric in high dimensions? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/99171/why-is-euclidean-distance-not-a-good-metric-in-high-dimensions?rq=1)

<https://inmachineswetrust.com/posts/drop-first-columns/>

# Classification

[r - Difference between logit and probit models - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/20523/difference-between-logit-and-probit-models?noredirect=1&lq=1)

[classification - How does a Support Vector Machine (SVM) work? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/23391/how-does-a-support-vector-machine-svm-work)

# Performance Metrics

[machine learning - Why is accuracy not the best measure for assessing classification models? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/312780/why-is-accuracy-not-the-best-measure-for-assessing-classification-models)

[regression - How to calculate Area Under the Curve (AUC), or the c-statistic, by hand - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/145566/how-to-calculate-area-under-the-curve-auc-or-the-c-statistic-by-hand?noredirect=1&lq=1)

[machine learning - Cross Validation Vs Train Validation Test - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/410118/cross-validation-vs-train-validation-test)

[machine learning - Why only three partitions? (training, validation, test) - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/9357/why-only-three-partitions-training-validation-test?noredirect=1&lq=1)

# Statistics

[Bayesian and frequentist reasoning in plain English - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/22/bayesian-and-frequentist-reasoning-in-plain-english)

[terminology - Is there a way to remember the definitions of Type I and Type II Errors? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/1610/is-there-a-way-to-remember-the-definitions-of-type-i-and-type-ii-errors?noredirect=1&lq=1)

[distributions - Why should we use t errors instead of normal errors? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/120776/why-should-we-use-t-errors-instead-of-normal-errors?noredirect=1&lq=1)

[correlation - Pearson vs Spearman vs Kendall - Data Science Stack Exchange](https://datascience.stackexchange.com/questions/64260/pearson-vs-spearman-vs-kendall/64261)

[probability - Why does a 95% Confidence Interval (CI) not imply a 95% chance of containing the mean? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/26450/why-does-a-95-confidence-interval-ci-not-imply-a-95-chance-of-containing-the)

[intuition - Intuitive explanation for dividing by $n-1$ when calculating standard deviation? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/3931/intuitive-explanation-for-dividing-by-n-1-when-calculating-standard-deviation)

[probability - What does it mean that a statistic $T(X)$ is sufficient for a parameter? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/84226/what-does-it-mean-that-a-statistic-tx-is-sufficient-for-a-parameter?noredirect=1&lq=1)

[terminology - What misused statistical terms are worth correcting? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/202879/what-misused-statistical-terms-are-worth-correcting?noredirect=1&lq=1)

[terminology - What is the difference between "likelihood" and "probability"? - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/2641/what-is-the-difference-between-likelihood-and-probability?noredirect=1&lq=1)

# Dimensionality Reduction

[pca - Making sense of principal component analysis, eigenvectors & eigenvalues - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues)

[statistics - Intuitive Way To Understand Principal Component Analysis - Mathematics Stack Exchange](https://math.stackexchange.com/questions/1146/intuitive-way-to-understand-principal-component-analysis)

[regression - PCA and proportion of variance explained - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/22569/pca-and-proportion-of-variance-explained?noredirect=1&lq=1)

# Python

[python - What does the "yield" keyword do? - Stack Overflow](https://stackoverflow.com/questions/231767/what-does-the-yield-keyword-do?rq=1)

[Highest scored 'python' questions - Stack Overflow](https://stackoverflow.com/questions/tagged/python?tab=Votes)

[python - Floor division // vs int() rounded off - Stack Overflow](https://stackoverflow.com/questions/52231739/floor-division-vs-int-rounded-off)