STATISTICS CONCEPTS USED IN THE DOMAIN OF DATA SCIENCE AND MACHINE LEARNING

XAVIER TANG

2020

Contents

Contents   iii

Contents   iv

CHAPTER 1

Distribution

This chapter deals with concepts mainly related to various probability distribution.

1.1   Bernoulli Trials

1.2   Discrete variable: Binomial Distribution

1.3   Discrete variable: Poisson Distribution

1.4   Continuous variable: Normal/Gaussian Distribution

1.5   Continuous variable: Exponential Distribution

1.6   Probability Distributed Function

1.7   Cumulative Distributed Function

CHAPTER 2

Metrics

2.1 ROC

2.2 Standard Deviation

2.3 Variance

2.4 Confusion Matrix

2.4.1 Recall

2.4.2 Precision

2.4.3 F1

2.5 P value

2.6 T value

2.7 Z value

2.8 Correlation/Pearson

2.9 Covariance

CHAPTER 3

Testing

3.1 T test

3.2 Chi-square test

3.3 Z test

3.4 A/B testing

3.4.1 Null hypothesis

3.4.2 Alternative hypothesis

3.4.3 Type I/II error

3.4.4 Statistical Power

3.5 Test of Significance

3.6 Hypothesis testing (one-way and two-way)

3.7 ANOVA

3.8 ANCOVA

3.9 One-sample/Two-sample bootstrap hypothesis test

3.10 Time series: p, d, q parameters, unit root and box test

CHAPTER 4

Thoerem

CHAPTER 5

General

5.1 Confidence Interval

5.2 Conditional Probability

5.3 Normalisation

5.4 Standardisatio

5.5 Least-squared error

5.6 R-squared error

5.7 Mean-squared error

5.8 Inferential Statistics

5.9 Bias-variance trade off