|  |  |  |
| --- | --- | --- |
| **Probability** | | |
| **Experiment**: A repeatable procedure that generates an outcome | | |
| **Sample space**: The set of all possible outcomes | | |
| **Event**: An outcome or set of possible outcomes | | |
| **Conditions for** **independence** of and : | | |
| **Bayes’ Theorem**: | | |
| **Mean** | | |
| **Expectation** | | |
| **Variance** | | |
| **Standard deviation**(s.d). | | |
|  | | |
|  | | |
|  | | |
| **Variance** | | |
|  | | |
|  | | |
| **s.d.** | | |
| **Covariance and correlation** | | |
|  | | |
|  | | |
|  | | |
|  | | |
| For 2 **dependent** variables and , | | |
| For 2 **independent** variables and , | | |
|  | | |
| **Discrete r.v.** | | |
| **Probability mass function (PMF)**: for all values of | | |
| **Cumulative distribution function (CDF)**: for all values of | | |
| **Continuous r.v.** | | |
| **Probability density function (PDF)**: | | |
| **Cumulative distribution function (CDF)**: | | |
| **Distributions** | | |
| **Bernoulli**: | | |
| **Binomial**: | | |
| **Geometric**: | | |
| **Uniform**: | | |
| **Exponential**: | | |
| **Normal**: | | |
| **Central Limit Theorem**: For any r.v. where n is large, sum () and average () are approximately normal. and | | |
| **Linear Regression** | | |
| **Least squares method**: Ensure sum of squared deviations is minimised | | |
| **Approximations**: | | |
|  | | |
| **Sum of squared residuals (SSR)** = , where  is the approximated value of | | |
| **Goodness of fit** | | |
| **Mean Absolute Error (MAE)** = | | |
| **Root Mean Square Error (RMSE)** = | | |
| **Bayesian Inference** | | |
| **Hypothesis ()**: A statement we wish to accept or reject | | |
| **Prior** : What we believe about the hypothesis without any evidence | | |
| **Likelihood** : Likelihood of the hypothesis | | |
| **Posterior** : Updated belief after seeing evidence | | |
| **Conditional independence**:      , where is the parameter | | |
| **Updating normal distributions**:  If n datapoints are drawn from N(), where is known,      Note also that , and | | |
| **Frequentist Inference** | | |
| **Type I error**: Reject when is true  **Type II error**: Do not reject when is true | | |
| **Significance level ()** = | | |
| **Power of a test** = – note that this requires the knowledge of the distribution of our random variable under as well | | |
| **Variance approximation**: if is unknown, we approximate it as  (i.e. 1/(n-1) \* sample variance | | |
| **Conversion to T-distribution**: Given with unknown , | | |
| **t-test for 2 samples**: take pooled sample variance () as | | |
| **Maximum Likelihood Estimation** | | |
| **MLE**: Consider data drawn from some distribution with an unknown parameter . The MLE estimation of is the value of that maximises the likelihood | | |
| **Geometric and Exponential**: | | |
| **Bernoulli and Binomial**: | | |
| **Uniform**: , | | |
| **Normal**: Sample mean and sample variance | | |
| **Confidence Interval**: Consider data drawn from some distribution with an unknown, fixed value . The interval estimator is called a confidence interval if , where is the confidence level | | |
| **Bias of point estimators**: is an unbiased estimator if , and biased otherwise | | |
| **Classification** | | |
| **Naïve Bayes’**: Assume that all features are mutually independent. | | |
| **Laplacian smoothing**: When the likelihood of any feature is 0, add 1 for each class and feature value | | |
|  | **Prediction (1/0)** | |
| **Ground truth (1/0)** | True Positive (TP) | False Negative (FN) |
| False Positive (FP) | True Negatve (TN) |
| **Accuracy**: | | |
| **Precision**: | | |
| **Recall**: | | |
| **F-score**: Harmonic mean of precision and recall | | |