**Finding Nearest Neighbours**

| **S No.** | **Name of the paper** | **Author and Year** | **Summary** |
| --- | --- | --- | --- |
| 1 | An algorithm for finding nearest neighbours in (approximately) constant average time | Enrique Vidal Ruiz et al.  Year: 1986 | The main features of the proposed algorithm are: (1) The average number of distance computations is in general tiny and, when the number of prototypes is large enough, tends to be independent of this number. (2) The average number of distance computations depends on the average distance from the test samples to their nearest prototypes. (3) The algorithm's performance decreases exponentially with the 'course of dimensionality'. However, this behaviour is mainly due to the increasing average distances from test samples to their nearest prototypes when the space dimension increases. |
| 2 | On finding p-th nearest neighbours of scattered points in two dimensions for small p | George Goodsell et al.  Year: 2000 | Given a large set of scattered points in the plane, a new and efficient algorithm is searched, for each point, the subset of p closest points, using the Dirichlet tessellation of the set of points, for small values of p. This problem has applications to interpolation and contouring, for example, in the field of Geographic Information Systems. |
| 3 | Finding and evaluating sets of nearest neighbours | Julie Weeds et al.  Year: 2002 | The paper observes that the frequency of a neighbour influences how much weight it has in the pseudo-disambiguation task. Low-frequency neighbours will have occurred with fewer verbs and, therefore, contribute to fewer decisions. One consequence is that more neighbours will generally need to be considered to conclude. A second consequence is that to obtain high performance on the pseudo-disambiguation task, one can sacrifice some semantic similarity to find lower frequency potential neighbours, safe in the knowledge that these neighbours, if wrong, will affect fewer test instance outcomes. |
| 4 | A Method For Determining K-Nearest Neighbours | Josef Kittler et al.  Year: 1978 | The method for determining k-nearest neighbours advocated is implemented in the paper. It has been found that calculating city block distances required for pre-processing is about four times faster than evaluating the Euclidean metric. The method was tested on a pattern recognition problem in 10-dimensional space. It is valid only for optimally decomposed data sets. Optimising the number of nodes and the average population of the final subgroups would increase the overall time needed to implement the method. Finally, the procedure can be used with advantage also in the k-nearest neighbour method of probability density function estimation. |
| 5 | Finding Nearest Neighbours in Growth-restricted Metrics | Matthias Ruhlet al.  Year: 2002 | Most research on nearest neighbour algorithms in the literature has been focused on the Euclidean case.In this paper, we develop an efficient dynamic data structure for nearest neighbour queries in growth-constrained metrics. These metrics satisfy the property that for any point q and number r the ratio between numbers of points in balls of radius 2r and r is bounded by a constant. |
| 6 | What Is a Good Nearest Neighbours Algorithm for Finding Similar Patches in Images? | Shree Nayaret al.  Year: 2008 | Many computer vision algorithms require searching a set of images for similar patches, which is a very expensive operation. In this work, we compare and evaluate a number of nearest neighbours algorithms for speeding up this task. Since image patches follow very different distributions from the uniform and Gaussian distributions that are typically used to evaluate nearest neighbours methods, we determine the method with the best performance via extensive experimentation on real images. Furthermore, we take advantage of the inherent structure and properties of images to achieve highly efficient implementations of these algorithms. Our results indicate that vantage point trees, which are not well known in the vision community, generally offer the best performance. |
| 7 | A Fast Algorithm for Finding k-Nearest Neighbours with Non-metric Dissimilarity | Bin Zhang et al.  Year: 2002 | Fast nearest neighbour (NN) finding has been extensively studied. While some fast NN algorithms using metrics rely on the essential properties of metric spaces. In this paper, a fast hierarchical search algorithm is proposed to find k-NNs using a non-metric measure in a binary feature space. Experiments with handwritten digit recognition show that the new algorithm reduces on average dissimilarity computations by more than 90% while losing the accuracy by less than 0:1%, with a 10% increase in memory. |
| 8 | Customer segmentation using K-means clustering | Kansal et. al.  Year: 2018 | In this paper the author opted for internal clustering validation rather than external clustering validation, which depends on some external data like labels. Internal cluster validation can be used for choosing a clustering algorithm which best suits the dataset and can correctly cluster data into its opposite cluster. |