**SIMILARITY SEARCH :**

**Turi Use case :** <https://en.wikipedia.org/wiki/MNIST_database>

**Similarity search** is the most general term used for a range of mechanisms which share the principle of searching (typically, very large) spaces of objects where the only available comparator is the similarity between any pair of objects. This is becoming increasingly important in an age of large information repositories where the objects contained do not possess any natural order, for example large collections of images, sounds and other sophisticated digital objects.

Similarity search is ubiquitous in the following fields of computer science, and these fields form the core of various everyday **applications**:

* Machine learning
* Computer vision
* Theory
* Databases
* Information retrieval
* Web application
* Collaborative filtering
* Scientific computing

**Examples** for the ***application*** of **similarity search** on **business process models, data compression, pattern recognition, machine learning, biomedical databases, statistical data analysis, malware detection,** and **data mining**.

# **Use Case : Machine learning of patient similarity: A case study on predicting survival in cancer patient after locoregional chemotherapy**

Identifying historical records of patients who are similar to the new patient could help to retrieve similar reference cases for predicting the clinical outcome of the new patient. Amongst different potential applications, this study illustrates use of patient similarity in predicting survival of patients suffering from hepatocellular carcinoma (HCC) treated with locoregional chemotherapy. This study used 14 similarity measures derived from relevant clinical and imaging parameters to classify the HCC patient pairs into two classes, namely the difference between their survival time being longer or no longer than 12 months. Furthermore, this paper proposes and presents a patient similarity algorithm for the classification, named SimSVM. With the 14 similarity measures as input, SimSVM outputs the predicted class and the degree of similarity or dissimilarity. A dataset was collected from 30 patients, forming 300 and 135 patient pairs as training and test datasets respectively. The trained SimSVM with linear kernel gave the best accuracy (66.7%), sensitivity (64.8%) and specificity (67.9%) on the test dataset.

**Useful Links :** <http://www.skytree.net/2015/09/04/learning-with-similarity-search/>