

PNN Documentation

Module name: PNN
Description: Probabilistic Neural Network (PNN)
Author: Gad Getz (Broad Institute) gp-help@broad.mit.edu

Summary: Probabilistic Neural Network (PNN) calculates the probability that a new sample belongs to a given set of classes. The contribution of each of the training samples to the class of the queried sample follows a Gaussian distribution. This algorithm can be considered as a Gaussian-weighted kNN classifier – close training samples have a greater influence on the predicted class.

In detail, the class posterior probability is calculated as follows:

$$P(c | \mathbf{x}) = \frac{P(\mathbf{x} | c)P(c)}{\sum_{c'} P(\mathbf{x} | c')P(c')} = \frac{\frac{P(c)}{n_c} \sum_{i: \bar{y}_i \in c} \exp(-D(\mathbf{x}, \mathbf{y}_i)^2 / 2\sigma^2)}{\sum_{c'} \left[\frac{P(c')}{n_{c'}} \sum_{i: \bar{y}_i \in c'} \exp(-D(\mathbf{x}, \mathbf{y}_i)^2 / 2\sigma^2) \right]}$$

where \mathbf{x} is the predicted sample and c is the class for which the posterior probability is calculated. The training set samples are \mathbf{y}_i , n_c is the number of samples of class c in the training set, and $D(\mathbf{x}, \mathbf{y}_i)$ is the distance between the predicted sample and training sample i . Note that the first step is derived using Bayes rule which allows to incorporate a prior probability for each class, $P(c)$.

Parameters

Name	Description
train.filename	training data file - .gct
train.cls.filename	class file for training data - .cls
test.filename	test data file - .gct
test.cls.filename	class file for test data - .cls
num.features	number of most significant features to select
test.type	statistical test to determine feature significance
min.std	minimal standard deviation to use in variance-thresholded t-test
sigma.type	method to determine sigma scale
sigma	value of sigma
prior.type	method to determine class prior probabilities
prior	prior probability to use for all classes (ignored if prior type is empirical)
distance.type	method to calculate distance
load.params	whether parameters should be loaded from a file
params.filename	file containing input parameters - .odf
pred.results.file	name of output file for prediction results - .odf

GenePattern

pred.full.results.file	name of output file for full prediction results - .odf
------------------------	---

References:

1. Lu, Getz, Miska, et al. "MicroRNA Expression Profiles Classify Human Cancers," Nature 435, 834-838 (9 June 2005)
2. Specht, D. F. Probabilistic Neural Networks, Neural Networks. Elsevier Science Ltd., St. Louis 3, 109-118 (1990).
3. Duda, R. O., Hart, P. E. & Stork, D. G. Pattern Classification, 2nd Ed. Wiley-Interscience, Hoboken (2000).

Return Value:

1. prediction results.
2. full prediction results

Platform dependencies:

Task type:	Prediction
CPU type:	x86
OS:	Windows
Language:	MATLAB