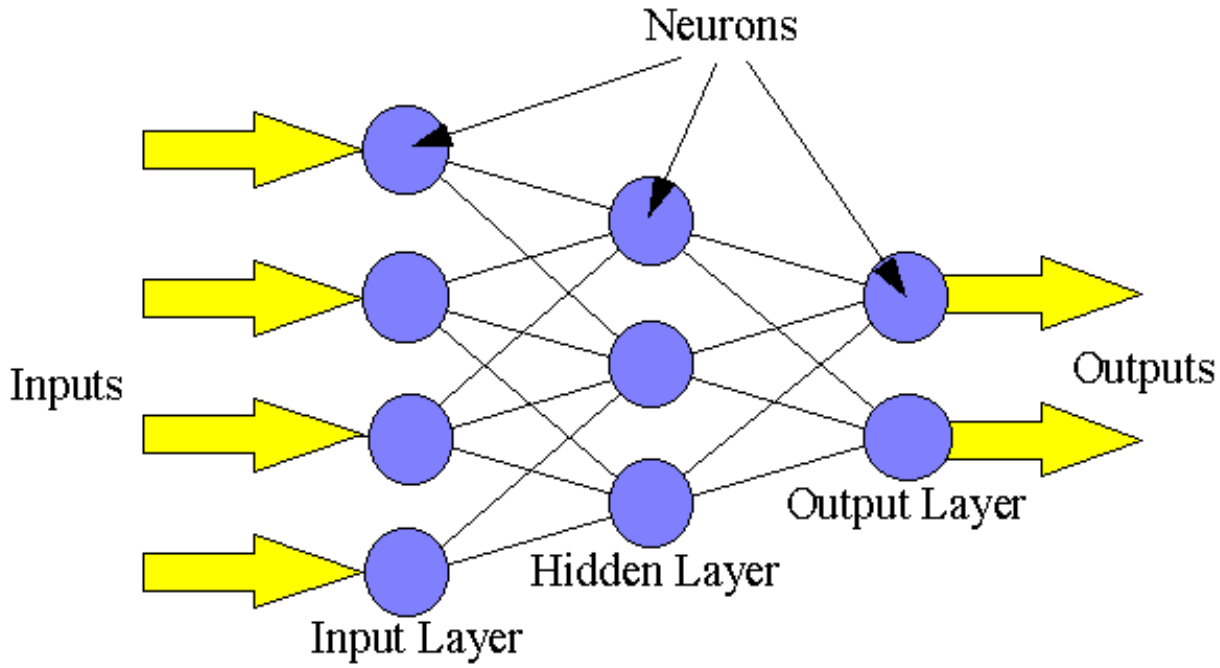
"The leader in the field of artificial intelligence
will become the ruler of the world."
VVP

Machine Learning: Bayesian and Neural Network statistical models

A.Chilingarian

AANL (YerPhI) master courses

Key words: Machine learning, Neural Networks, Nonparametric methods, Bayesian statistics, Pattern Recognition, Monte Carlo Statistical Inference, Multivariate Data Analysis, Learning Algorithms, Applied Programming, Software-Hardware Combined System Development, Trigger Applications, High-Energy Astrophysics, Genome Analysis.



Yerevan 2019

Data analysis and Statistical Inference (12 hours)

Estimating the unknown parameters of the population based on the sample results.
Hypothesis Testing
Picturing Distributions, probability density function, Normal Distributions
Methods for organizing, displaying and describing data by using tables, graphs, and summary measures.
Inferential Statistics: describing the population based on the sampling results.
Probability Theory: population and finite sample
Sampling technique: Representative and random samples.
Picturing Distributions, probability density function, Normal Distributions
Measures of Central Tendency and Variability.
Histogram and scatter plot, correlation and causal relation.
Direct & inverse problems
How to deal with uncertainty entering Statistical inference
Bayesian approach to model-based analysis
Parametric and nonparametric probability density estimation
K-nearest neighbors density estimation
Comparison of different methods of nonparametric probability density estimation on standard N(0,1) and N(1,1) problem.
Bayes law and risk.
Posterior and prior information, experimental (decisive) information.
Selection of the best feature and best pair of features for classification purposes.
MAP (maximum a posteriori probability) method.
Visualization of uncertainty.
Methods to assign prior knowledge.
Butstrapisation.
Parametric classification (parameter fitting, overtraining)
ANI modes of Bayesian classification

Optimization in multivariate spaces (8 hours)

Measures on Metric spaces;
Distances in metric spaces between clusters and populations
Euclidian and Mahalonobis metrics
Parametric and non-parametric distances
Distances based on ranks
Kolmogorov and Mann-Whitney distances
Decision making, Quality function and losses
Optimal strategies of minimizing losses
Newman-Pearson methodology
Family of Random Search methods
Random search with return on unsuccessful step
Committee principle, escaping from local minimum
Median method

Random Search methods in ANI package

Mathematical models of Neural Networks (16 hours)

Monte Carlo Statistical Inference
Biological inspiration
Brain and von-Neumann architecture
Parallel computing
Learning and generalization
The curse of dimensionality
Overfitting and model complexity
Architecture of feed-forward neural networks
Selection of number of nodes and layers
Overall concept of NN learning
Random search learning algorithm
Neural classification
Neural estimation (learning regression function)
Learning (NN training) algorithms
Evolutionary algorithms, committee of networks, ensembles of networks
Resolving the mixture of analytic models
Estimation of the generalization error in neural classification to multiple categories
The cross-validation procedure for final prediction error (FPE) estimation
Estimates of FPE as an net architecture selection tool
Comparison of the different algorithms of NN training for the problem of classification
Deep Learning
Feature selection under concept of training without teacher
Hardware NN accelerators
ANI package modes of Neural Training

Applied problems of Machine learning in high-energy astrophysics, genome analysis and pattern recognition (16 hours)

Monte Carlo statistical Inference
CORSIKA platform
KASCADE and MAKET experiments
Extensive Air Showers (EAS) and primary mass classification
The “purification” procedure
Statistical Techniques in background rejection for Atmospheric Cherenkov Telescope
Colored Nuclear Maps
Estimation of the generalization error in neural classification to multiple categories
A priori methods of background rejection in Very high-energy Cherenkov Imaging technique
Investigation of Interpolation Possibilities by Neural Networks
Multistart Random Search with Early Stop as tool for selection of sets of differently expressed genes

Application for medical diagnostics: two colon cancer cell lines

Literature

Ashot Chilingarian, Nonparametric Methods of Data Analysis in Cosmic Ray Astrophysics, An Applied Theory of Monte Carlo Statistical Inference, International Science and Technology C enter, Moscow, 2007.
ANI program package and ANI manual, http://crd.yerphi.am/ANI_User_Guide_Introduction