**Course: Advance Bio Informatics**

**Module Title: Classification**

**Module No: 124**

**Classification**

Samples are annotated with a qualitative label

* Cancer/Non-cancer patients
* Samples of seed tissue for which it is known if the seed germinated or not
* Age of the sample

Datasets as classification problems, and generate prediction models from the data. Not just as classification problems

* Clustering/Biclustering
* Association Rule Mining
* Regression

**Data Properties**

Domain experts interestin predictions

Biomarker identification

* Identify key variables
* Most strongly associated to each outcome
* Presenting higher prediction capacity

Identify interactions b/w variables

* By presenting very high (anti)correlation between them
* By acting together to generate predictions

Corpus used for informationextraction in biomedical domain.Information extraction is arelationship of genes, proteins, and RNA from scientific publications.Named entities and their relationships is exploitingalong with dependency analysis of sentence syntax.Ontologies defining types of entities & relationships annotated in the corpus.

1100 sentences from abstracts of biomedical research articles.