Feature Selection from gene expression

**Aim:** To select a subset of genes that can maximize the discrimination between various phenotypes.

**Given:** A set of gene expression profiles and the corresponding phenotypes.

**Method:**
- Each gene expression profile forms a data item with the corresponding phenotype as the label.
- The features are the genes. The corresponding gene expression denotes the value of feature.
- Apply feature selection using maximum discrimination to obtain the set of features that maximize discrimination.

**Flowchart:**
- Choose a feature $f_i$ from $F$ which maximizes the discrimination between the 2 classes.
- $F = F - f_i$
- $S = S \cup \{f_i\}$
- Calculate the new divergence between the 2 classes.
- Is increase in divergence very small?
  - Output $S$
  - No

**Feature Selection using Maximum Discrimination**

**Aim:** To select the minimal set of features which maximize the discrimination between classes.

**Given:**
- $F = \{f_1, f_2, \ldots, f_n\}$
- $S = \{\}$

**Selected features**

**Flowchart:**
- Choose a feature $f_i$ from $F$ which maximizes the discrimination between the 2 classes.
- $F = F - f_i$
- $S = S \cup \{f_i\}$
- Calculate the new divergence between the 2 classes.
- Is increase in divergence very small?
  - Output $S$
  - No

**Jensen Shannon Divergence Minimization**

**Aim:**
- To select the minimal set of features which maximize the discrimination between classes.

**Given:**
- $F = \{f_1, f_2, \ldots, f_n\}$
- $S = \{\}$

**Selected features**

**Flowchart:**
- Choose a feature $f_i$ from $F$ which maximizes the discrimination between the 2 classes.
- $F = F - f_i$
- $S = S \cup \{f_i\}$
- Calculate the new divergence between the 2 classes.
- Is increase in divergence very small?
  - Output $S$
  - No

**Applications of Non-extensive Information Theory**

**q-Gaussian distribution**
- Generalization of normal distribution
- q-Gaussians applied to Function Smoothening

**q-Gaussian Smoothed**
- Gaussian Smoothed
- q-Gaussian Smoothed

**Properties of Tsallis Divergence Minimization**

Exponential distribution obtained from KL divergence minimization
- Power law distribution obtained from q-divergence minimization

**Our Work:**
- Generalize the important properties of Kullback-Leibler divergence minimization, such as Subset Independence, Pythagorean property for Tsallis divergence minimization.

**Label Ranking using Minimum Description Length Principle**

**Aim:** To rank the labels according to their relevance to a given data item.

**Intuition:**
- More the regularity in data, more it can be compressed.
- Learning $\leftrightarrow$ Finding regularity in data.

**Algorithm**

- In the algorithm for feature selection using maximum discrimination, Jensen Shannon divergence can be used to calculate divergence between the classes.