COMPSIG NITT

In This Issue...

- **Illustration:** F-score for feature selection
- **On-going research:** Current research works in PRCI Lab.

Dear friends! COMPSIG NITT is a monthly newsletter to share the research work done in the Pattern recognition and computational intelligence laboratory, Department of Electronics and Communication Engineering, National Institute of Technology Trichy.

Concepts, Ideas pertaining to Computational intelligence, Pattern recognition and Signal processing are also included in this newsletter.

We expect the feedback, comments and articles from you all.

Issue 4-3: March 2018

Team members

1. **Dr. E.S.Gopi, Co-ordinator.**
2. **G.JayaBrindha, Ph.D. Scholar.**
3. **Neema. M, Ph.D. Scholar.**
4. **Rajasekharreddy Poreddy, Ph.D Scholar.**
5. **Vineetha Yogesh, M.Tech, Communication systems.**
6. **Shaik Mahammad, M.Tech, Communication systems.**

F - score for feature selection

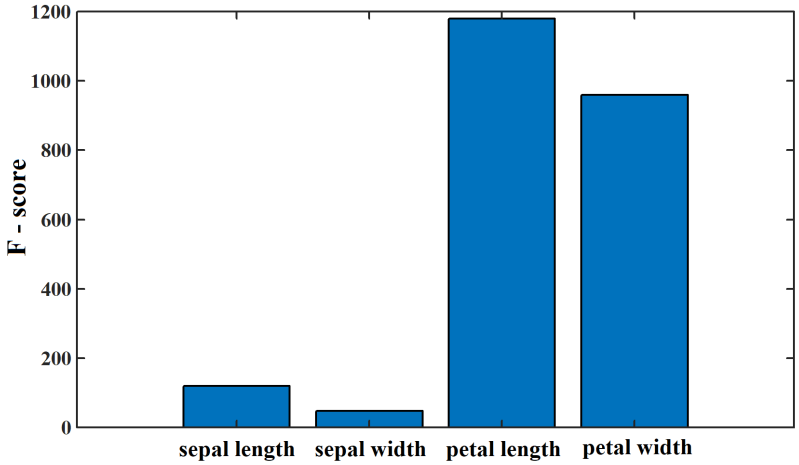

Fig. 1: F - score of 4 features in Iris Data

Statistical analysis of the features in discriminating the data is computed using the F - score. It gives a quantitative measure of discrimination contributed by each feature. Larger F - score implies better discrimination of the data while using that feature. The steps in computing the F - score are,

1. Compute between class variance (S_B^2).

$$S_B^2 = \frac{1}{k-1}\sum_{i=1}^{k} n_i(\overline{X_i} - \overline{X})^2$$

   where $\overline{X_i}$ is the mean of i^{th} class, $\overline{X}$ is the overall mean and n_i is the number of samples in the i^{th} class.
2. Compute within class variance (S_W^2).

$$S_W^2 = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X_i})^2$$

   where X_{ij} is the j^{th} sample in i^{th} class.
3. F - score is the ratio of between class variance (S_B^2) and within class variance (S_W^2).

$$Fscore = \frac{S_B^2}{S_W^2}$$

This ratio follows F distribution and $k-1$ & $N-k$ are the corresponding degrees of freedom of S_B^2 & S_W^2 respectively. For illustration, F - score is computed for 4 features in Iris dataset (refer Fig.1). The dataset consists of 3 classes, each with 50 samples. From the figure, it is seen that petal length and petal width are the most significant features in discriminating the data.

Link to the dataset: Iris dataset

Back to Contents

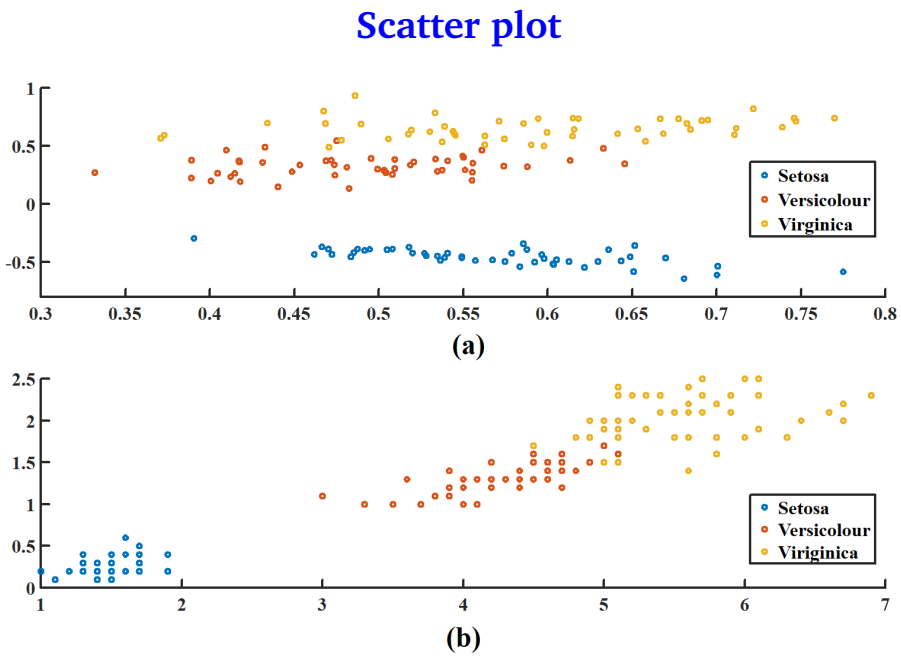

Fig. 2: Scatter plot of Iris data

Fig. 2(a) shows the scatter plot of the dimensionally reduced Iris data, where the significant features are obtained using LDA. Fig. 2(b) shows the scatter plot of the Iris data with only two features (petal length and petal width) that are selected based on the F - score. These significant features can be directly identified using the F - score instead of collecting all the features and proceeding for dimensionality reduction. At instances where more effort is required for collecting the features, the significant features alone can be collected based on the F - score, which reduces the effort in feature extraction.

Link to the m-file: Fscore

Back to Contents

We are happy to share that, our recently published article titled "Ant Colony Technique for Optimizing the Order of Cascaded SVM Classifier for Sunflower Seed Classification" in IEEE Transactions on Emerging Topics in Computational Intelligence stands as 4th most popular article amongst the articles published in the journal.

Link to the article: Popular article, IEEE TETCI

Back to Contents

Quotes

"Creativity is seeing the same thing but thinking differently." — Dr. A.P.J.Abdul Kalam

On-going Research

- Constructing a Sunflower plant database and perform off-type identification using morphological features
- Application of machine learning techniques in next generation wireless communication
- Classification of Music composition styles using probabilistic generative model
- Computational Intelligence for channel estimation of Massive MIMO
- Investigation of Empirical Match Algorithm for latent sequence estimation in HMM

Back to Contents

Feedback

COMPSIG NITT invites articles and innovative ideas from readers for the Reader's Space column. We expect feedback and comments to monthly newsletter COMPSIG NITT . Readers can share their views in our facebook page, "COMPSIG-NITT". Those who are interested can be a part of the facebook group.

Back to Contents

Contact Information:
Pattern Recognition and Computational Intelligence Laboratory,
Department of Electronics and Communication Engineering,
National Institute of Technology Trichy - 620015
E-mail:esgopi@nitt.edu