Description of the program "fselq":

Automatic selection of informative features providing minimum of probability of classification errors using the quadratic discriminatior

The program "fselq" accomplishes the following operations:

1) the automatic stepwise selecting of the most informative features providing the least error probability P(k) for the classification based on the given set of learning data;

2) calculating of a function *P(k), k = 1,2,...,p* which is a theoretic total probability classification errors in depend on an amount k of features used for classification;

3) calculating of the value k_0 for which the function *P(k)* attains its minimum ($k_0 = argminP(k)$);

4) plotting of the calculated function *P(k)* on the screen using the standard UNIX routine "plotxy" with displaying on this plot the numbers and labels of features chosen at every selection step.

The value *P(k)* is calculated by the formula:

$$P_{er}(k) = \frac{1}{2}[1 + F_0(\frac{m_1(k)}{\sqrt{\sigma_1^2(k)}}) - F_0(\frac{m_2(k)}{\sqrt{\sigma_2^2(k)}})], \tag{A1}$$

where: $F_0(z)$ is the cumulative function of the standard Gaussian probability distribution;

$$m_1(k) = \frac{1}{2}\{ pk_1 - k_2(\theta_1 + \theta_2) - \frac{p}{n_2 - p} + \alpha\left[\theta_6 + p\,ln\frac{n_2 - 1}{n_1 - 1} + \theta_5\right]\};$$

$$m_2(k) = \frac{1}{2}\left\{k_1(\omega_1 + \omega_2) - pk_2 + \frac{p}{n_1 - p} + \alpha[\ \theta_6 + p\,ln\frac{n_2 - 1}{n_1 - 1} + \theta_5\ ]\right\};$$

$$\sigma_1^2(k) = \frac{1}{2}\left\{pk_1^2 - 2k_1k_2\theta_1 + k_2^2\theta_4 + 2k_2^2(\frac{\theta_1}{n_2} + \theta_3) + p^2\frac{k_1^2}{(n_1 - p)} + \theta_1^2\frac{k_2^2}{(n_2 - p)} + \left[\theta_7 + \theta_8\right]\right\};$$

$$\sigma_2^2(k) = \frac{1}{2}\left\{pk_2^2 - 2k_1k_2\omega_1 + k_1^2\omega_4 + 2k_1^2(\frac{\omega_1}{n_1} + \omega_3) + p^2\frac{k_2^2}{(n_2 - p)} + \omega_1^2\frac{k_1^2}{(n_1 - p)} + \left[\theta_7 + \theta_8\right]\right\};$$

$$\theta_1 = \sum_{i=1}^{p}\frac{s_{ii}^{(1)}}{s_{ii}^{(2)}}; \quad \omega_1 = \sum_{i=1}^{p}\frac{s_{ii}^{(2)}}{s_{ii}^{(1)}}; \quad \theta_2 = \sum_{i=1}^{p}\frac{(\overline{x}_i^{(2)} - \overline{x}_i^{(1)})^2}{s_{ii}^{(2)}}; \quad \omega_2 = \sum_{i=1}^{p}\frac{(\overline{x}_i^{(2)} - \overline{x}_i^{(1)})^2}{s_{ii}^{(1)}};$$

$$\theta_3 = \sum_{i=1}^{p}\frac{(\overline{x}_i^{(2)} - \overline{x}_i^{(1)})^2 s_{ii}^{(1)}}{(s_{ii}^{(2)})^2}; \quad \omega_3 = \sum_{i=1}^{p}\frac{(\overline{x}_i^{(2)} - \overline{x}_i^{(1)})^2 s_{ii}^{(2)}}{(s_{ii}^{(1)})^2}; \quad \theta_4 = \sum_{i=1}^{p}(\frac{s_{ii}^{(1)}}{s_{ii}^{(2)}})^2; \quad \omega_4 = \sum_{i=1}^{p}(\frac{s_{ii}^{(2)}}{s_{ii}^{(1)}})^2;$$

$$\theta_5 = \sum_{i=1}^{p} ln(\frac{s_{ii}^{(1)}}{s_{ii}^{(2)}}); \quad \theta_6 = \sum_{i=1}^{p} ln(\frac{n_1 - i}{n_2 - i}); \quad \theta_7 = \sum_{i=1}^{p-1}\frac{1}{n_2 - i}; \quad \theta_8 = \sum_{i=1}^{p-1}\frac{1}{n_1 - i};$$

$\overline{x}_i^{(1)}$ is i-th component of the sampling mean vector $\overline{X}^{(1)}$ of the first class;

$\overline{x}_i^{(2)}$ is i-th component of the sampling mean vector $\overline{X}^{(2)}$ of the second class;

$s_{ii}^{(1)}$ is i-th diagonal element of the covariance matrix $\boldsymbol{S}^{(1)}$ of the first class;

$s_{ii}^{(2)}$ is i-th diagonal element of the covariance matrix $\boldsymbol{S}^{(2)}$ of the second class; (i=1,2,...,p);

α is the coefficient with the property: $\alpha = 1$, if the term $ln\frac{\left|S^{(1)}\right|}{\left|S^{(2)}\right|}$ is taken into account in the equation for quadratic discriminant function QD and $\alpha = 0$ in the opposite case ($\alpha = 1$ in this version).

$F_0(z)$ - is the cumulative function of the standard Gaussian probability distribution;

Formula (A1) was derived via an asymptotic expansion of the distribution function of the conventional quadratic discriminator under the assumption that the number or features p and numbers or learning vectors n_1, n_2 for both the classes are simultaneously increasing with the same rate (Kolmogorov's asymptotic).

At the first step of the selecting procedure p values of the $P(1)$ probability are calculated for every feature. The minimum of these p values is attained at some $j(1)$ feature which thus is selected. At the second step $(p-1)$ values of the $P(2)$ probability are calculated for the pairs of features: the first member of these pairs is always the previously selected feature $j(1)$, the second member - is an arbitrary feature from the remaining ones. Then the second feature $j(2)$ is selected which ensures the minimum of these $(p-1)$ values of the $P(2)$ probability. At the k-th step $(k=1,...,p)$ of this selecting procedure $(p-k+1)$ values of the $P(k)$ probability are calculated for the k-dimensional feature vectors. The first $k-1$ components of these vectors are the features $j(1),j(2),...j(k-1)$ which were selected at the previous $k-1$ steps, the k-th component is an arbitrary feature from the remaining ones. The k-th feature $j(k)$ is selected which ensures the minimum of these $(p-k+1)$ values. On the each step $k = 1,2,...,p$ of the selecting procedure the value $P_{min}(k)$, number $j(k)$ and label of the feature selected are stored.

The program "fselq" then defines a number k_0 of that selecting step for which a minimum of function $P_{min}(k)$, $k = 1,...,p$, is attained: $k_0 = argmin\ P_{min}(k)$. Thus the optimal set of features with the numbers j$(1),j(2),...j(k_0)$ become determined. These features provide the minimal estimate of the theoretic total error probability based on the quadratic discrimination function.

Program input parameters

The all input parameters of the program are to be contained in the file "fselq.inp". An example of this input file is given below

```
******************************************************************************
****INPUT FILE FOR PROGRAM "FSELQ": standard****
NAME OF FILE FOR LEARNING MEAN VECTORS AND COVARIANCE MATRIX
data/troitsky/resku.dat
NAME OF FILE FOR GRAPHIC RESULTS:
```

```
plot/troitsky/mygr.gr
******************************************************************************
```