1. \(a(x,y)\) and \(\theta(x,y)\) are two low spatial frequency images that we might fuse into a new image, which should also have low spatial frequencies.
   (a) Assuming that \(a(x,y)\) and \(\theta(x,y)\) are fused into a complex Pseudospectrum (PS) \(s(x,y)\) which is to be compared to the ith reference PS, \(s_i(x,y)\). Give the real cross-correlation function \(c_i(x,y)\) that can be used for this.
   (b) Assume that \(a(x,y)\) and \(\theta(x,y)\) are not to be fused. Instead, their cross-correlations with reference images \(s_{ia}(x,y)\) and \(s_{i\theta}(x,y)\) are to be fused. Give the two cross-correlations \(c_{ai}(x,y)\) and \(c_{i\theta}(x,y)\). Give the final fused cross-correlation \(c_i(x,y)\) in terms of the symbols \(c_{ai}(x,y)\) and \(c_{i\theta}(x,y)\).
   (c) Assume that \(\theta(x,y)\) has much more useful information than does \(a(x,y)\). How can we change \(c_i(x,y)\) in part (b) in terms of the symbols \(c_{ai}(x,y)\) and \(c_{i\theta}(x,y)\) ?

2. Let \(d_{ik}\) denote the ith class type \(B_3\) discriminant for the kth classifier, where \(k\) varies from 1 to \(K\). Let the nth discriminant be crisp, so that \(d_{in}\) equals 0 or 1.
   (a) Can we successfully fuse the \(K\) discriminants using the max or min operators? Why?
   (b) Repeat (a) for the product operator.
   (c) Repeat part (a) for the averaging operator.

3. Center vectors for a nearest neighbor classifier can be stored as \(\{m_{ik}\}\) for \(1 \leq i \leq N_c\), \(1 \leq k \leq K(i)\). The total number of clusters is \(K = \sum K(i)\). In order to calculate all \(N_c\) class discriminants without error, we can apply the partial distance (PD) method independently for each class.
   (a) For the ith class, how can we initialize the minimum distance \(d_i\) for that class at the beginning of its PD algorithm?
   (b) Keeping the data in the \(\{m_{ik}\}\) form, we want to modify the PD approach in part (a) to reduce the number of multiplications. How should we initialize the minimum distance for cluster number 1. How should we initialize the minimum distance for the ith cluster?
   (c) We want to make the non-minimum discriminants from part (b) larger so that we can more easily form case \(B_3\) discriminants from them. Suppose that \(d_i\) is not the minimum \(d_i\). Some of the distances \(d(x, m_{ik})\) used for the ith class were complete (denoted by \(d'(x, m_{ik})\) where all \(N\) features \(x(n)\) were used) and some were partial (denoted by \(d^p(x, m_{ik})\)). How can we make \(d_i\) larger, so that we can more easily form case \(B_3\) discriminants from the \(d(x, m_{ik})\)?
4. In a pattern recognition problem, there are $N_c$ classes and $N$ features.
(a) In a nonlinear transformation approach to feature selection, what is the
minimum value that can be used for $N_1$, if $N_c$ perfect case $B_3$ discriminants are
available which sum up to 1?
(b) If the discriminants are imperfect and don’t add up to 1, what is the new
minimum value for $N_1$, assuming that the $B_3$ discriminants must be used as
features, without modification?
(c) Continuing part (b), what nonlinear operator can be applied to the imperfect
discriminants that solves the problem mentioned in part (b)?
(d) Assume that naïve Bayes Gaussian discriminants are used to form the
transformed features $z_j$, where $1 \leq j \leq N_1$. Give an expression for $f(x|j)$ in terms
of $m_i(n)$ and $\sigma_i^2(n)$. Give expressions for $f(x)$ in terms of the symbol $f(x|j)$, and
the feature $z_j$ in terms of the symbols $f(x|j), f(x)$, and the probabilities $P_i$. For
this case, do we use the $N_1$ value of part (a) or the $N_1$, value from part (b)?

5. Given training data of the form $\{x_p, i_c(p)\}$, we want to find a good feature subset
of size $N_1$ for $N_1$ between 1 and $N$-1. Sequential Forward Selection (SFS) and
Sequential Backward Selection (SBS) require the same numbers of evaluations of
the J function, for this task. However, that does not prove that both algorithms are
equally efficient. Let $J(N_1)$ denote a J function evaluation for one subset of size $N_1$.
(a) How many $J(1)$ and $J(2)$ evaluations are used in the first and second steps of
SFS? Give $N_{SFS}(N_1)$, the number of $J(N_1)$ evaluations used by SFS.
(b) How many $J(N-1)$ and $J(N-2)$ evaluations are used in the first and second steps
of SBS? Find $N_{SBS}(N_1)$, the number of $J(N_1)$ evaluations used by SBS.
(c) If each $J(N_1)$ evaluation requires $(N_1)^3$ multiplications, how many multiplies are
required to evaluate $J(N_1)$ for all relevant values of $N_1$ in SFS? Repeat this for
SBS. (Hint: Sums are ok. Closed forms are not required)

6. Consider the three FLN type subset evaluation functions (SEFs) described in the
appendix. We want to use them in branch and bound feature selection, where $J_K(x^K) = J_{FLN}(U)$
(a) Describe the set $U$ in terms of all the features $x^N$ and the set of $K$ discarded
features $x^K$.
(b) If $J_K(x^K)$ is formed from $J_{FLN1}(x)$, give the monotonicity condition that must be
satisfied in terms of $J_K(x^K)$ and $J_{K+n}(x^{K+n})$, where $x^K \subset x^{K+n}$. Is it satisfied?
(c) If $J_K(x^K)$ is formed from $J_{FLN3}(x)$, give the monotonicity condition that must be
satisfied in terms of $J_K(x^K)$ and $J_{K+n}(x^{K+n})$, where $x^K \subset x^{K+n}$. Is it satisfied?
7. We want to transform the feature vector $\mathbf{x}$ into a smaller feature vector $\mathbf{z}$, as $\mathbf{z} = \mathbf{A} \cdot \mathbf{x}$, where $\mathbf{A}$ is $N_1$ by $N$. The class separability in the $N_1$-dimensional $\mathbf{z}$ space can be measured by 

$$J_1 = \text{tr}(S_{wz}^{-1} S_{bz})$$

where $S_{wz} = \mathbf{A} \mathbf{S}_{wx} \mathbf{A}^T$ and $S_{bz} = \mathbf{A} \mathbf{S}_{bx} \mathbf{A}^T$. Equating

$$\frac{\partial J_1}{\partial \mathbf{A}^T} = -2S_{wz}^{-1} \mathbf{A}^T S_{wz}^{-1} S_{bz} S_{wz}^{-1} + 2S_{bz} \mathbf{A}^T S_{wz}^{-1}$$

to a matrix of zeroes, we want to find an equation of the form $\mathbf{B} \cdot \mathbf{A}^T = \mathbf{A}^T \cdot \mathbf{D}$.

(a) Give matrices $\mathbf{B}$ and $\mathbf{D}$ in terms of $S_{wz}$, $S_{bz}$, $S_{wx}$, and $S_{bx}$.

(b) What assumption about the scatter matrices allows us to develop your solution in part (a)?
Gaussian Probability Density

\[ f_{x}(x) = \frac{1}{(2\pi)^{N/2} |C|^{1/2}} e^{-1/2(x-m)^T A (x-m)} \]

FLN-Based SEFs

(1) Design the FLN as before, for \( x^{N_1} \). Pass through the training data once, and use \( J_{FLN1}(x^{N_1}) = P_e \) for the classifier.

(2) Design the FLN as before, for \( x^{N_1} \). \( J_{FLN2}(x^{N_1}) \) is calculated during a pass through the data as

\[ J_{FLN2}(x^{N_1}) = \sum_{n \times x(n) \in x^{N_1}} \sum_{p=1}^{N_x} \sum_{i=1}^{N_c} \left| \frac{\partial y_{pi}}{\partial x_{pn}} \right|. \]

(3) Given the feature subset \( x^{N_1} \), delete unnecessary rows and columns from \( C \) and \( R \), and design the FLN. Calculate the FLN’s MSE as

\[ J_{FLN3}(x^{N_1}) = \sum_{i=1}^{N_c} E(i), \]

\[ E(i) = \frac{1}{N_v} \sum_{p=1}^{N_v} \left[ t_{pi} - \sum_{n=1}^{L} w(i, n) X_p(n) \right]^2 \]

\[ t_{pi} = \delta(i - i_c(p)) \]

Softmax Operator

\[ d_{i}' = \frac{d_{i}'}{\sum_{n=1}^{N_c} d_{n}'} \]