Journal of Multivariate Analysis 101 (2010) 68–76

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Energy discriminant analysis, quantum logic, and fuzzy sets

Grigorii Melnichenko

Informatics Department, Vilnius Pedagogical University, Studentu Street 39, LT-08106 Vilnius, Lithuania

ARTICLE INFO

Article history:
Received 2 January 2008
Available online 3 May 2009

AMS subject classifications:
68T10
03G12
03E72
94D05
91B06

Keywords:
Recognition
Quantum logic
Discriminant function
Fuzzy set
von Neumann density matrix
Membership functions
Subspace classifier
Quality functional
Quantum decision

ABSTRACT

In this paper, we show that quantum logic of linear subspaces can be used for recognition of random signals by a Bayesian energy discriminant classifier. The energy distribution on linear subspaces is described by the correlation matrix of the probability distribution. We show that the correlation matrix corresponds to von Neumann density matrix in quantum theory. We suggest the interpretation of quantum logic as a fuzzy logic of fuzzy sets. The use of quantum logic for recognition is based on the fact that the probability distribution of each class lies approximately in a lower-dimensional subspace of feature space. We offer the interpretation of discriminant functions as membership functions of fuzzy sets. Also, we offer the quality functional for optimal choice of discriminant functions for recognition from some class of discriminant functions.

© 2009 Published by Elsevier Inc.

1. Introduction

A Bayesian probabilistic discriminant classifier is based on a classical probability theory using algebra of subsets. The decision rule of the probabilistic classifier maximizes the probability of “correct” recognition. A Bayesian energy discriminant classifier was briefly presented in [1]. The algebra of linear subspaces (quantum logic) is used instead of algebra of subsets. The decision rule of energy classifier maximizes the energy of “correct” recognition. The recognition of two classes is considered in detail. The use of quantum logic for recognition of signals is considered in [2].

The use of linear subspaces is based on the assumption that the distribution of each class lies approximately in a lower-dimensional subspace of feature space. These spaces can be found by principal components analysis carried out individually on each class. An input vector from the unknown class is classified according to the greatest projection to the subspaces, each of which represents one class.

The subspace classifier was suggested by Watanabe (method CLAFIC [3,4]). This method, however, has drawbacks: a priori probabilities of classes are not used; subspaces of classes can overlap. T. Kohonen has offered the Learning Subspace Method (LSM) [5,3]. During the training LSM decreases the number of vectors that are included in subspaces of different classes. The recognition of handwritten signs by the subspace classifier is considered in [4]. The subspace classifier is applied to phonemes recognition in [6] and to speaker recognition in [7].

Eldar and Oppenheim [8] draw a parallel between quantum measurements and algorithms in signal processing. They propose to exploit the rich mathematical structure of quantum theory in signal processing without realization of quantum

E-mail addresses: gmelnicenko@gmail.com, grimel@vpu.lt.

0047-259X/$ – see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.jmva.2009.04.011

processes. We suggest to consider energy processes instead of quantum processes because nature spends some energy to create any signal.

2. Quantum logic as an example of fuzzy logic

Let H be a Hilbert space that can be complex or real, finite- or infinite-dimensional in this section. In the following sections we consider only real and finite-dimensional Hilbert spaces. A fuzzy set A of H is a pair $\{x, \mu_A(x): x \in H\}$, where $\mu_A(x): H \to [0, \infty)$ is called a membership function of the fuzzy set A, i.e., for each $x \in H$, $\mu_A(x)$ is the degree of membership of x. The fuzzy set and its membership function often do not differ. Suppose $\mu_A(x)$ be non necessarily normal: $\sup \mu_A(x) \neq 1, x \in$ H. A set of membership functions is a partially ordered set equipped with a partial order relation: $\mu_A(x) \leq \mu_B(x)$ for all $x \in H$. The result of operations

$$\mu_A(x) \wedge \mu_B(x) = \inf\big(\mu_A(x), \mu_B(x)\big), \qquad \mu_A(x) \vee \mu_B(x) = \sup\big(\mu_A(x), \mu_B(x)\big)$$

is defined pointwise and the result is again a nonnegative function. Hence, the set of membership functions is a lattice.

Each closed linear subspace $M \subset H$ corresponds to an elementary logical proposition of quantum logic. Each linear subspace M has an orthogonal projection P_M onto M. So a proposition of quantum logic can be associated with the orthogonal projection. The set of all orthogonal projections is a lattice equipped with a partial order relation: $P \leq R$ if $\langle Px, x\rangle \leq \langle Rx, x\rangle$ for all $x \in H$. Hence every pair of projections P, R has a unique supremum (least upper bound) and a unique infimum (greatest lower bound):

$$P \wedge R = \inf(P, R), \qquad P \vee R = \sup(P, R).$$

Operations $P \wedge R$, $P \vee R$, and $P^{\perp} = I - P$ are conjunction, disjunction, and negation of quantum logic, respectively.

Each projection P_M on the subspace M can be viewed as a filter [2] and it passes some energy $\mu_M(x) = \langle P_M x, x\rangle = \|P_M x\|^2$ of signal x (in quantum theory, a projection passes some quantum probability). This energy evaluates the value of membership of signal x to subspace M. So each linear subspace $M \subset H$ can be associated with the fuzzy set:

$$A_M = \big\{x, \mu_M(x): x \in H,\ M \subset H\big\}, \quad \text{where } \mu_M(x) = \langle P_M x, x\rangle.$$

A set of all membership functions $\{\mu_M(x), M \subset H\}$ is a lattice equipped with a partial order relation: $\langle P_M x, x\rangle \leq \langle P_N x, x\rangle$ for all $x \in H$. So operations supremum and infimum of that lattice can be used as a fuzzy logic conjunction and disjunction of fuzzy sets $\{A_M, M \in H\}$. A fuzzy logic negation of fuzzy set A_M with membership function $\mu_M(x)$ can be defined as a fuzzy set $A_{M^{\perp}}$ using the following membership function: $\mu_{M^{\perp}}(x) = \langle P_{M^{\perp}} x, x\rangle = \langle P_M^{\perp} x, x\rangle = \langle (I - P_M)x, x\rangle$, where a subspace $M^{\perp}$ is the orthogonal complement of subspace M. Thus fuzzy sets $\{A_M, M \subset H\}$ form a fuzzy logic.

3. Discriminant functions as membership functions

If an object of recognition is described as a vector $x = (x_1, x_2, \ldots, x_n)$, where each $x_i \in \mathbb{R}$, $i = 1 \ldots l$, is a feature, then the vector x is the pattern of the object in the feature space $\mathbb{R}^n$. A membership of object to some class S_i, $i = 1 \ldots l$, is an additional feature, which can be defined as the index i of the class, where $i \in I = \{1, 2, \ldots, l\}$.

We use discriminant functions for the classifier of recognition. Discriminant functions are functions $g_i(x)$, $i = 1 \ldots l$, that determine the membership of the object with the pattern x to class S_i according to the following decision rule: if the object with the pattern x satisfies $g_i(x) > g_j(x)$ for all $j \neq i$, then the object having the pattern x is assigned the class S_i.

Discriminant functions split the feature space $\mathbb{R}^n$ into disjoint sets:

$$A_i = \big\{x: g_i(x) > g_j(x), j = 1 \ldots l, j \neq i\big\}.$$

Thus, if $x \in A_i$, then the object having the pattern x is assigned the class S_i. However, there are sets $\{x: g_i(x) = g_j(x),\ j \neq i\}$, $i = 1 \ldots l$, whose elements it is impossible to include in some set A_i, $i = 1 \ldots l$. Usually these sets redefine A_i, $i = 1 \ldots l$ to allow equality.

Using discriminant functions, the classifier determines the value about the membership of the object with the pattern x to some class S_i. Thus the discriminant functions $g_i(x)$, $i = 1 \ldots l$, can be understood as membership functions. In the following, we assume that discriminant functions are non-negative and need not satisfy $\sup g_i(x) \neq 1, x \in \mathbb{R}, i = 1 \ldots l$.

4. Quality functional for a choice of optimal decision rule

We shall use a probabilistic model for recognition. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space where a sample space Ω is a set of recognition objects. It is evident that the set of recognition classes $S_1, S_2, \ldots, S_l$ are a partition of Ω: $S_1 \cup S_2 \cup \cdots \cup S_l = \Omega$ where $S_i \cap S_j = \emptyset$ for all $i \neq j$.

Following Zadeh [9], a fuzzy set A is called a fuzzy event if the corresponding membership function $\mu_A(\omega): \Omega \to [0, \infty)$ is $\mathcal{A}$-measurable. The probability of a fuzzy event is defined as

$$\mathbf{P}(A) = \mathbf{E}\mu_A = \int_{\Omega} \mu_A \mathrm{d}\mathbf{P}. \tag{1}$$

Suppose that an object ω is described by the vector $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ of features where each $\xi_i(\omega)$: $\Omega \to \mathbb{R}$, $i = 1 \dots n$, is $\mathcal{A}$-measurable random variable. Since each object ω has the pattern x in the feature space $\mathbb{R}^n = H$, there is a map $\xi(\omega)$: $\Omega \to H$. We define an integer-valued random variable γ such that $\gamma(\omega) = i$ for all $\omega \in S_i$, where $i \in I=\{1, 2, \dots, l\}$. The sample space Ω of the objects is usually not accessible to immediate observation, therefore it is necessary to deal with the feature space H. However, Ω can be considered as $I \times H$.

We use a Bayesian method which needs a prior probabilities $p_i = \mathbf{P}(S_i)$, $i = 1 \dots l$, and conditional distributions $\mu_i(A) = \mathbf{P}(\xi \in A|S_i)$, $i = 1 \dots l$. Since $\mathbf{P}(S_i) = \mathbf{P}(\gamma = i)$, it follows that p_i, $i = 1 \dots l$, is the probability distribution of the random variable γ.

We denote $\mu(B, A) = \mathbf{P}(\gamma \in B, \xi \in A)$, where $B = \{i_1, i_2, \dots, i_m\} \subset I$ and $A \subset$ H. The measure μ is the probability distribution on $I \times H$. Since $\mu_i(A) = \mathbf{P}(\xi \in A|S_i)$ and $S_i = (\gamma = i)$, where $i = 1 \dots l$, we get

$$\mu(\{i\}, A) = \mathbf{P}(\gamma = i, \xi \in A) = \mathbf{P}(\xi \in A|\gamma = i)\mathbf{P}(\gamma = i) = \mathbf{P}(\xi \in A|S_i)\mathbf{P}(S_i) = p_i\mu_i(A).$$

Let us put $\mu_1(\{i\}) = p_i$ and $\mu_2^1(i, A) = \mu_i(A)$. We have

$$\begin{aligned}\mu(B, A) &= \mathbf{P}\left(\sum_{k=1}^{m}(\gamma = i_k) \cap (\xi \in A)\right) = \sum_{k=1}^{m}\mathbf{P}(\xi \in A \mid \gamma = i_k)\mathbf{P}(\gamma = i_k)\\ &= \sum_{k=1}^{m}\frac{\mu(\{i_k\}, A)}{p_{i_k}}p_{i_k} = \sum_{k=1}^{m}\mu_2^1(i_k, A)\mu_1(\{i_k\}) = \int_B \mu_2^1(i, A)\mu_1(\mathrm{d}\, i).\end{aligned}$$

It follows that $\mu_2^1(i, A) = \mu_i(A)$ is the transition probability on $I \times \mathcal{B}$ [10], where $\mathcal{B}$ is a σ-algebra of Borel subsets of feature space $H = R^n$.

Discriminant functions $g_i(x)$, $i = 1 \dots l$, define a random variable $g_\gamma(\xi)$. We put $g(\gamma, \xi) = g_\gamma(\xi)$. Since $\mu_2^1(i, A) = \mu_i(A)$ is the transition probability on $I \times \mathcal{B}$, we have [10]

$$\mathbf{E}g(\gamma, \xi) = \int_I \mu_1(\mathrm{d}i)\int_H g(i, x)\mu_2^1(i, \mathrm{d}x) = \sum_{i=1}^{l}p_i\int_H g_i(x)\mu_i(\mathrm{d}x). \tag{2}$$

Suppose $H = A_1 \cup A_2 \cup \dots \cup A_l$, where A_i, $i = 1 \dots l$, are disjoint sets. Let Φ be a class of discriminant functions which contain only indicator functions:

$$g_i(x) = 1_{A_i}(x) = \begin{cases}1 & \text{if } x \in A_i,\\ 0 & \text{if } x \notin A_i.\end{cases}$$

It is evident that $g(\gamma(\omega), \xi(\omega)) = g_{\gamma(\omega)}(\xi(\omega))$ is the indicator function with a support:

$$G = \sum_{i=1}^{l}(\xi \in A_i) \cap (\gamma = i) = \sum_{i=1}^{l}(\xi \in A_i) \cap S_i.$$

We can say that the indicator function $1_G = g(\gamma, \xi)$ is the membership function of "correct" recognition, where G is a ordinary event (crisp event) of "correct" recognition. By (2), we have

$$\mathbf{P}(G) = \mathbf{E}g(\gamma, \xi) = \sum_{i=1}^{l}p_i\int_H g(i, x)\mu_i(\mathrm{d}x) = \sum_{i=1}^{l}\mathbf{P}(\xi \in A_i|S_i)\mathbf{P}(S_i). \tag{3}$$

If we use discriminant functions that are indicator functions, then Bayesian discriminant classifier splits the feature space H into disjoint sets $H = A_1 \cup A_2 \cup \dots \cup A_l$ such that the probability (3) for the ordinary event G of "correct" recognition would be maximal.

Let $g_i(x)$, $i = 1 \dots l$, be discriminant functions from some class Φ, where each function $g_i(x)$: $H \to [0, \infty)$ is a Borel-measurable membership function of class S_i. Then the random variable $g_i(\xi(\omega))$, $i = 1 \dots l$ on Ω is a membership function such that the value $g_i(\xi(\omega))$ is a membership degree of object ω to a class S_i. We define fuzzy events $G_1, G_2, \dots, G_l$ as $G_i = \{\omega, g_i(\xi(\omega)) \colon \omega \in \Omega\}$ for all $i = 1 \dots l$.

Let us define the membership function:

$$\mu_j(i, \omega) = 1_{S_j}(\omega)g_i\big(\xi(\omega)\big) = \begin{cases}g_i\big(\xi(\omega)\big) & \text{if } \omega \in S_j,\\ 0 & \text{if } \omega \notin S_j.\end{cases}$$

This membership function defines the fuzzy event $S_jG_i = \{\omega, \mu_j(i, \omega) \colon \omega \in \Omega\}$, which is a product [9] of events G_j and S_i. The value $\mu_j(i, \omega)$ is the membership degree of the object ω to the class S_i if the statement $\omega \in S_j$ is true. There can be two cases. First, if $j = i$, then $\mu_i(i, \omega)$ is the membership degree of the object ω to the class S_i when the object ω belongs to its own class S_i. We call the value $\mu_i(i, \omega)$ a "correct" degree of membership; we call the fuzzy event S_iG_i a fuzzy event of "correct" recognition. Second, if $j \neq i$, then $\mu_j(i, \omega)$ is the membership degree of the object ω to the class S_i when the object ω belongs to the other class S_j. We call the value $\mu_j(i, \omega)$, $j \neq i$, an "error" degree of membership; we call the fuzzy event S_jG_i, $j \neq i$, a fuzzy event of "error" recognition.

Since $g(\gamma,\xi)=g_\gamma(\xi)$ and $1_{S_i}=1_{(\gamma=i)}$ for all $i=1\ldots l$, we get

$$g(\gamma,\xi)=\sum_{i=1}^{l}1_{(\gamma=i)}g(\gamma,\xi)=\sum_{i=1}^{l}1_{(\gamma=i)}g_i(\xi)=\sum_{i=1}^{l}1_{S_i}g_i(\xi)=\sum_{i=1}^{l}\mu_i(i,\omega).$$

This membership function defines a degree of "correct" membership for all objects $\omega\in\Omega$. We call the random variable $g(\gamma,\xi)$ as a membership function of "correct" recognition and the fuzzy set $G=\{\omega,g(\gamma(\omega),\xi(\omega))\colon\omega\in\Omega\}$ as a fuzzy event of "correct" recognition.

It is natural to choose discriminant functions $g_i(x), i=1\ldots l$ from the class Φ such that the probability of the fuzzy event G of "correct" recognition would be maximal. From (1) and (2), we have that the probability of the fuzzy event G is defined as

$$\mathbf{P}(G)=\mathbf{E}g(\gamma,\xi)=\sum_{i=1}^{l}p_i\int_H g(i,x)\mu_i(\mathrm{d}x). \tag{4}$$

Also (4) defines a quality functional for choice of discriminant functions from the class Φ.

Let us show another interpretation of the quality functional (4). We define

$$1_{(k=j)}(k)=\begin{cases}1 & \text{if } k=j,\\ 0 & \text{if } k\neq j.\end{cases}$$

Since $1_{S_i}=1_{(\gamma=j)}$, it follows that [10]

$$\begin{aligned}\mathbf{E}\big(1_{S_j}g_i(\xi)\big)&=\mathbf{E}\big(1_{(\gamma=j)}g_i(\xi)\big)\\&=\iint_{I\times H}1_{(k=j)}g_i(x)\mu(\mathrm{d}k,\mathrm{d}x)=\int_I\mu_1(\mathrm{d}k)1_{(k=j)}\int_H g_i(x)\mu_2^1(k,\mathrm{d}x)\\&=\sum_{k=1}^{l}\mu_1(\{k\})1_{(k=j)}\int_H g_i(x)\mu_2^1(k,\mathrm{d}x)=p_j\int_H g_i(x)\mu_j(\mathrm{d}x).\end{aligned}$$

Then the probability of the fuzzy event $S_jG_i=\{\omega,1_{S_j}(\omega)g_i(\xi(\omega))\colon\omega\in\Omega\}$ is defined as

$$r_j(i)=\mathbf{P}(S_jG_i)=\mathbf{E}\big(1_{S_j}g_i(\xi)\big)=p_j\int_H g_i(x)\mu_j(\mathrm{d}x). \tag{5}$$

We call the value $r_j(i)$ a "correct" probability of recognition if $i=j$ and an "error" probability of recognition if $i\neq j$. By (4) and (5), the total sum of all the "correct" probabilities of recognition is equal to $\mathbf{P}(G)$, so we have another interpretation of the quality functional (4).

By definition of the conditional expectation, we have $\mathbf{E}\big(g_i(\xi)|S_i\big)=\mathbf{E}(g_i(\xi)1_{S_i})/\mathbf{P}(S_i)$, $i=1\ldots l$. Hence we get one more interpretation of the quality functional (4):

$$\mathbf{P}(G)=\mathbf{E}g(\gamma,\xi)=\mathbf{E}\left(\sum_{i=1}^{l}1_{(\gamma=i)}g(\gamma,\xi)\right)=\sum_{i=1}^{l}\mathbf{E}\big(1_{S_i}g_i(\xi)\big)=\sum_{i=1}^{l}\mathbf{E}\big(g_i(\xi)|S_i\big)\mathbf{P}(S_i).$$

5. Basic formula

Let $(\Omega,\mathcal{A},\mathbf{P})$ be a probability space and H is a finite-dimensional Hilbert space over the real numbers with the inner product $\langle\cdot,\cdot\rangle$, norm $\|\cdot\|$, and σ-algebra $\mathcal{B}_H$ of Borel sets. A map $\xi\colon\Omega\to H$ is called a random element if it is of $(\mathcal{A},\mathcal{B}_H)$-measurable. The measure ν given by $\nu=\mathbf{P}(\xi^{-1}(A)), A\in\mathcal{B}_H$ is called a probability distribution of the random element ξ. If Ω is a set of recognition objects, then the vector value $x=\xi(\omega)$ can be understood as the pattern of the object ω.

Assuming that $\mathbf{E}\|\xi\|^2<\infty$, we can define one linear form and two bilinear forms

$$\langle m,y\rangle=\mathbf{E}\langle\xi,y\rangle=\int_H\langle x,y\rangle\nu(\mathrm{d}x),$$

$$\langle Ky,z\rangle=\mathbf{E}\Big(\langle\xi,y\rangle\langle\xi,z\rangle\Big)=\int_H\langle x,y\rangle\langle x,z\rangle\nu(\mathrm{d}x), \tag{6}$$

$$\langle Ry,z\rangle=\mathbf{E}\Big(\langle\xi-m,y\rangle\langle\xi-m,z\rangle\Big)=\int_H\langle x-m,y\rangle\langle x-m,z\rangle\nu(\mathrm{d}x). \tag{7}$$

Under the assumption that the random element ξ takes values in Hilbert space, the non-random vector m, operator K, and operator R are usually called as mathematical expectation, a correlation operator, and a covariance operator, respectively.

From (6) and (7), we have $\langle Ky, z\rangle = \langle Ry, z\rangle + \langle m, y\rangle\langle m, z\rangle$. Then $\langle Ry, z\rangle + \langle m, y\rangle\langle m, z\rangle = \langle (R + p_m)y, z\rangle$, where $p_m y = \langle y, m\rangle m$ is a rank-one operator. It is evident that $p_m y = \|m\|^2 p_{\bar{m}} y$, where $\bar{m} = m/\|m\|$ and $p_{\bar{m}} y = \langle y, \bar{m}\rangle \bar{m}$ is a one-dimensional projection. Then

$$K = R + p_m = R + \|m\|^2 p_{\bar{m}}. \tag{8}$$

If $\mathbf{P}(\xi = a) = 1$, where a is a non-random vector, then the covariance operator R is the zero operator. From (9) it follows that the correlation K operator is the rank-one operator. If $\|a\| = 1$ in addition, in quantum theory it can be said that a quantum system is prepared in the pure state.

An affine structure of Hilbert space H is used when realizations of random element are considered as points. Using a vector-space structure of H, it is possible to interpret a value $\|x\|^2$ as a physical value, for example, as energy, power, or intensity. If the random element is random signal, then the value $\|x\|^2$ is a measure of deviation of signal from the zero vector, and nature uses some energy for this deviation. In the following, we refer to $\|x\|^2$ as energy.

Let $\langle A\xi, \xi\rangle$ be a bilinear form, where A is a linear operator. Then

$$\mathbf{E}\langle A\xi, \xi\rangle = \int_H \langle Ax, x\rangle \nu(\mathrm{d}x) = \int_H \langle x, Ax\rangle \nu(\mathrm{d}x) = \operatorname{tr} KA = \operatorname{tr} AK. \tag{9}$$

If P is an orthogonal projection, then $\langle P\xi, \xi\rangle$ is the membership function. We can define a fuzzy event $A_P = \{\omega, \langle P\xi(\omega), \xi(\omega)\rangle : \omega \in \Omega\}$. According to (1) and (9), the probability of the fuzzy event A_P is equal to

$$\mathbf{P}(A_P) = \mathbf{E}\langle P\xi, \xi\rangle = \int_H \langle Px, x\rangle \nu(\mathrm{d}x) = \operatorname{tr} PK = \operatorname{tr} KP.$$

We now prove formula (9). Let $\{e_i\}$, $i = 1 \dots n$, be an orthonormal basis in H. Using definitions of trace and correlation operator (6), we have

$$\begin{aligned}\operatorname{tr} KA &= \sum_{i=1}^n \langle KAe_i, e_i\rangle = \sum_{i=1}^n \int_H \langle x, Ae_i\rangle\langle x, e_i\rangle \nu(\mathrm{d}x) = \int_H \sum_{i=1}^n \left\langle A^*x, \langle x, e_i\rangle e_i\right\rangle \nu(\mathrm{d}x)\\ &= \int_H \left\langle A^*x, \sum_{i=1}^n \langle x, e_i\rangle e_i\right\rangle \nu(\mathrm{d}x) = \int_H \langle A^*x, x\rangle \nu(\mathrm{d}x) = \int_H \langle x, Ax\rangle \nu(\mathrm{d}x).\end{aligned}$$

Since the scalar product is symmetric in a real Hilbert space, $\langle x, y\rangle = \langle y, x\rangle$, we get $\langle Ax, x\rangle = \langle x, Ax\rangle$. Then

$$\begin{aligned}\operatorname{tr} AK &= \sum_{i=1}^n \langle AKe_i, e_i\rangle = \sum_{i=1}^n \langle Ke_i, A^*e_i\rangle = \sum_{i=1}^n \int_H \langle x, e_i\rangle\langle x, A^*e_i\rangle \nu(\mathrm{d}x)\\ &= \int_H \left\langle \sum_{i=1}^n \langle x, e_i\rangle x, A^*e_i\right\rangle \nu(\mathrm{d}x) = \int_H \langle x, Ax\rangle \nu(\mathrm{d}x) = \int_H \langle Ax, x\rangle \nu(\mathrm{d}x) = \mathbf{E}\langle A\xi, \xi\rangle.\end{aligned}$$

Statistical states of quantum system are described by the von Neumann density matrix [11]. In fact, the von Neumann density matrix is the correlation matrix of the discrete probability distribution. The formula (9) enables us to describe statistical states of quantum system with continuous probability distributions.

6. Recognition of two signal classes

Helstrom was first who considered recognition of two classes in the quantum theory [11]. We apply Helstrom's result for recognition of two classes of random signals; we only consider an energy distribution instead of quantum probability distribution on projections.

If each recognition object $\omega \in \Omega$ has a pattern $x \in H$, then we have map $\xi : \Omega \to H$, i.e., ξ is a random element. Here and in the sequel, we will use a term "signal" instead of the term "element". Generally a random signal is of physical nature, therefore it carries energy.

Assume that the object ω of recognition belongs to one of the classes S_i, $i = 1, 2$ and the pattern of object is the signal $x = \xi(\omega)$. Suppose that each class S_i, $i = 1, 2$, is matched with the orthogonal projection P_i, $i = 1, 2$, where $P_1 + P_2 = I$. Then the value $\langle P_i x, x\rangle = \langle P_i\xi(\omega), \xi(\omega)\rangle = g_i(\xi(\omega))$ is the membership of object ω to the class S_i, $i = 1, 2$. Therefore, the projections P_i, $i = 1, 2$, define a class Φ of discriminant functions $g_i(x) = \langle P_i x, x\rangle$, $i = 1, 2$.

Let $p_i = \mathbf{P}(S_i)$, $i = 1, 2$ be a priori probabilities of classes and let the conditional distributions $\nu_i(A) = \mathbf{P}(\xi \in A|S_i)$, $i = 1, 2$, have the correlation operators K_i, $i = 1, 2$. We define a fuzzy event $G = \{\omega, g(\gamma(\omega), \xi(\omega)) : \omega \in \Omega\}$, where $g(\gamma, \xi) = \langle P_\gamma \xi, \xi\rangle$. By (4), we must maximize the probability of the fuzzy event G:

$$\mathbf{P}(G) = \mathbf{E}g(\gamma, \xi) = p_1 \int_H \langle P_1 x, x\rangle \nu_1(\mathrm{d}x) + p_2 \int_H \langle P_2 x, x\rangle \nu_2(\mathrm{d}x). \tag{10}$$

Let us suggest an energy interpretation of formula (10). Using (5) and (10), we have

$$r_j(i) = \mathbf{E}\big(1_{S_j}\langle P_i\xi, \xi\rangle\big) = p_j \int_H \langle P_i x, x\rangle \nu_j(\mathrm{d}x) = p_j \mathrm{tr}\, P_i K_j.$$

By the assumption, the projection P_i corresponds to the class S_i. If $\omega \in S_i$, then the projection P_i (as a filter) passes an energy of signals of its own class S_i; the average value of such energy is $r_i(i)$. If $\omega \in S_j, j \neq i$, then the projection P_i passes an energy of signals of the other class $S_j, j \neq i$; the average value of such energy is $r_j(i), j \neq i$. We call energy $r_j(i)$ a "correct" energy if $i = j$ and an "error" energy if $i \neq j$. We also call the total "correct" energy, which passes projections of all classes, as an energy of "correct" recognition. This energy is defined as

$$\mathrm{Enr}_{\mathrm{C}}(P_1, P_2) = r_1(1) + r_2(2) = p_1 \mathrm{tr}\, P_1 K_1 + p_2 \mathrm{tr}\, P_2 K_2. \tag{11}$$

It is clear that we must find projections P_1, P_2 so that the value $\mathrm{Enr}_{\mathrm{C}}(P_1, P_2)$ would be the largest. In other words, projections P_1, P_2 together must pass the energy of signals from their own classes as much as possible.

Since $P_2 = I - P_1$, we have

$$\mathrm{Enr}_{\mathrm{C}}(P_1, P_2) = p_2 \mathrm{tr}\, K_2 + \mathrm{tr}\, P_1(p_1K_1 - p_2K_2).$$

Here the first value is constant, but the second value depends only on the projection P_1. Hence we must find the projection P_1 such that the second value was the largest. Assume that $\lambda_i, i = 1 \dots n$, are eigenvalues and $y_i, i = 1 \dots n$, are the eigenvectors of the operator $p_1K_1 - p_2K_2$. Then

$$\begin{aligned} \mathrm{tr}\, P_1(p_1K_1 - p_2K_2) &= \sum_{i=1}^{n} \langle P_1(p_1K_1 - p_2K_2)y_i, y_i\rangle = \sum_{i=1}^{n} \langle P_1 \lambda_i y_i, y_i\rangle \\ &= \sum_{i=1}^{n} \lambda_i \|P_1 y_i\|^2 = \sum_{\lambda_i > 0} \lambda_i \|P_1 y_i\|^2 + \sum_{\lambda_i \le 0} \lambda_i \|P_1 y_i\|^2 = d_1 + d_2, \end{aligned}$$

where $\|P_1y_i\|^2 \le \|y_i\|^2$ for all $i = 1 \dots n, d_1 > 0, d_2 \le 0$. Let P_1 be a projection onto a subspace spanned by the eigenvectors with positive eigenvalues. Then $\|P_1y_i\|^2 = \|y_i\|^2$ if $\lambda_i > 0$ and $\|P_1y_i\|^2 = 0$ if $\lambda_i \le 0$. It follows that d_1 will be the largest and $d_2 = 0$. Hence the required projection P_1 is found and $P_2 = I - P_1$.

Comment 1. It is possible to minimize the energy of "error" recognition. The energy of "error" recognition is the following sum:

$$\mathrm{Enr}_{\mathrm{E}}(P_1, P_2) = p_1 r_1(2) + p_2 r_2(1) = p_1 \mathrm{tr}\, P_2 K_1 + p_2 \mathrm{tr}\, P_1 K_2.$$

If the projections P_1, P_2 maximize the energy of "correct" recognition, then they must minimize energy of "error" recognition. Indeed, we have

$$\begin{aligned} \mathrm{Enr}_{\mathrm{E}}(P_1, P_2) &= p_1 \mathrm{tr}(P_2K_1) + p_2 \mathrm{tr}(P_1K_2) = p_1 \mathrm{tr}(I - P_1)K_1 + p_2 \mathrm{tr}(I - P_2)K_2 \\ &= p_1 \mathrm{tr}\, K_1 + p_2 \mathrm{tr}(K_2) - p_1 \mathrm{tr}\, P_1K_1 - p_2 \mathrm{tr}(P_2K_2) \\ &= p_1 \mathrm{tr}\, K_1 + p_2 \mathrm{tr}(K_2) - \mathrm{Enr}_{\mathrm{C}}(P_1, P_2). \end{aligned} \tag{12}$$

There the values $p_1 \mathrm{tr}\, K_1$ and $p_2 \mathrm{tr}\, K_2$ are constant. Hence the value $\mathrm{Enr}_{\mathrm{E}}(P_1, P_2)$ will be the least if the value $\mathrm{Enr}_{\mathrm{R}}(P_1, P_2)$ is the greatest.

From (12) it follows that the sum energy of "correct" recognition and "error" recognition is a constant. Thus, increasing the energy of "correct" recognition, we decrease the energy of "error" recognition and vice versa.

7. Decision rule for recognition

Suppose there are two classes of objects $S_i, i = 1, 2$, and the signal $x = \xi(\omega)$ is the pattern of the object ω. If we use a probabilistic Bayesian classifier, then the feature space H is divided into the disjoint subsets: $L_1, L_2, L_1 \cup L_2 = \mathrm{H}$, where the subset L_1 correspond to the class S_1 and the subset L_2 corresponds to the class S_2. The decision rule that determines unambiguously to which class S_1 or S_2 belongs the object ω, is defined as follows: $\omega \in S_1$ if $x \in L_1$ and $\omega \in S_2$ if $x \in L_2$.

However, the situation is different when quantum logic is used. Suppose each class $S_i, i = 1, 2$, is matched with the orthogonal projection $P_i, i = 1, 2$, where $P_1 + P_2 = I$. Denote $L_1 = P_1H, L_2 = P_2H$, where $L_1 \oplus L_2 = H$. Then the pattern of the object $x = \xi(\omega)$ is a sum of two signals: $x = P_1x + P_2x$, where $P_1x \in L_1, P_2x \in L_2$. It is natural to accept that $\omega \in S_1$ if $P_1x = x$ (hence $P_2x = 0$), $\omega \in S_2$ if $P_2x = x$ (hence $P_1x = 0$). Suppose $P_1x \neq 0$ and $P_2x \neq 0$, then $0 \neq P_1x \in L_1$ and $0 \neq P_2x \in L_2$ simultaneously. It is as if the pattern x belongs to above subspaces L_1 and L_2 simultaneously. For instance, let projections P_1, P_2 be resonators that select different frequencies from a signal. If the spectrum of the signal has frequencies of both resonators, then both resonators oscillate simultaneously. Hence we can not decide to which class belongs the object using subspaces of quantum logic.

Therefore, we must use discriminant functions $g_i(x) = \langle P_i x, x\rangle$, $i = 1, 2$, which unambiguously gives the decision about the membership of the object to one of the classes: S_1 or S_2. By (11), we can find discriminant functions $g_1(x) = \langle P_1 x, x\rangle$ and $g_2(x) = \langle P_2 x, x\rangle$ such that they maximize the energy of "correct" recognition. Thus we have the following decision rule:

$$\omega \in S_1 \quad \text{if } \langle P_1 x, x\rangle > \langle P_2 x, x\rangle \quad \text{and} \quad \omega \in S_2 \quad \text{otherwise.} \tag{13}$$

When the decision rule (13) is applied, the feature space H is divided into disjoint sets: $A_1 = \{x: \langle P_1 x, x\rangle > \langle P_2 x, x\rangle\}$ and $A_2 = \{x: \langle P_2 x, x\rangle \geq \langle P_1 x, x\rangle\}$. We put

$$\mathrm{Enr}_{\mathrm{C}}(A_1, A_2) = p_1 \int_{A_1} \langle P_1 x, x\rangle \nu_1(\mathrm{d}x) + p_2 \int_{A_2} \langle P_2 x, x\rangle \nu_2(\mathrm{d}x).$$

It is evident that

$$\mathrm{Enr}_{\mathrm{C}}(P_1, P_2) = \mathrm{Enr}_{\mathrm{C}}(A_1, A_2) + p_1 \int_{A_2} \langle P_1 x, x\rangle \nu_1(\mathrm{d}x) + p_2 \int_{A_1} \langle P_2 x, x\rangle \nu_2(\mathrm{d}x). \tag{14}$$

The object ω of recognition is chosen in a random way but we hope that the value of the discriminant function $g_i(x)$ of class S_i is maximal if statement $\omega \in S_i$ is true. Also it is natural to hope that $\mathrm{Enr}_{\mathrm{C}}(P_1, P_2)$ is approximately equal to $\mathrm{Enr}_{\mathrm{C}}(A_1, A_2)$. If we apply the decision rule (13), then $\langle P_1 x, x\rangle \leq \langle P_2 x, x\rangle$ on A_2 and $\langle P_2 x, x\rangle < \langle P_1 x, x\rangle$ on A_1. So we get

$$p_1 \int_{A_2} \langle P_1 x, x\rangle \nu_1(\mathrm{d}x) \leq p_1 \int_{A_2} \langle P_2 x, x\rangle \nu_1(\mathrm{d}x) \leq p_1 \int_{H} \langle P_2 x, x\rangle \nu_1(\mathrm{d}x) = p_1 \mathrm{tr}\, P_2 K_1,$$
$$p_2 \int_{A_1} \langle P_2 x, x\rangle \nu_2(\mathrm{d}x) \leq p_2 \int_{A_1} \langle P_1 x, x\rangle \nu_2(\mathrm{d}x) \leq p_2 \int_{H} \langle P_1 x, x\rangle \nu_2(\mathrm{d}x) = p_2 \mathrm{tr}\, P_1 K_2.$$

From (14) it follows that

$$0 \leq \mathrm{Enr}_{\mathrm{C}}(P_1, P_2) - \mathrm{Enr}_{\mathrm{C}}(A_1, A_2) \leq p_1 \mathrm{tr}\, P_2 K_1 + p_2 \mathrm{tr}\, P_1 K_2 = \mathrm{Enr}_{\mathrm{E}}(P_1, P_2). \tag{15}$$

If projections P_1, P_2 maximize the energy $\mathrm{Enr}_{\mathrm{C}}(P_1, P_2)$ of "correct" recognition, then from Comment 1 it follows that projections P_1, P_2 minimize the energy $\mathrm{Enr}_{\mathrm{E}}(P_1, P_2)$ of "error" recognition. If we have good recognition with projections P_1, P_2, then the value $\mathrm{Enr}_{\mathrm{E}}(P_1, P_2)$ is small. Therefore from (15) it follows that $\mathrm{Enr}_{\mathrm{C}}(P_1, P_2)$ is approximately equal to $\mathrm{Enr}_{\mathrm{C}}(A_1, A_2)$.

Example 1. Suppose the object of recognition ω belongs to one of the classes S_i, $i = 1, 2$. Assume that a priori probabilities of classes are equal $p_1 = p_2 = 1/2$; the conditional distributions $\nu_i(A) = \mathbf{P}(\xi \in A|S_i)$, $i = 1, 2$, have the identical covariance matrices equal to R and mathematical expectations m_1, m_2 are orthogonal as vectors.

We choose the orthonormal basis e_i, $i = 1 \ldots n$, in H such that $e_1 = m_1/\|m_1\|$, $e_n = m_2/\|m_2\|$. We get from (8) that $K_1 = R + \|m_1\|^2 p_1$, $K_2 = R + \|m_2\|^2 p_2$, where $p_1 x = \langle x, e_1\rangle e_1$, $p_2 x = \langle x, e_n\rangle e_n$. In the chosen basis, the matrix $p_1 K_1 - p_2 K_2 = 1/2(K_1 - K_2)$ is diagonal with eigenvalues $\|m_1\|^2/2, 0, \ldots, 0, -\|m_2\|^2/2$. Then $P_1 x = \langle x, m_1\rangle/\|m_1\|$, $P_2 x = \langle x, m_2\rangle/\|m_2\|$. If $x = \xi(\omega)$ is the pattern of the object ω, then by (13) we have the following decision rule: $\omega \in S_1$ if $\langle m_1, x\rangle^2/\|m_1\|^2 > \langle m_2, x\rangle^2/\|m_2\|^2$ and $\omega \in S_2$ otherwise.

8. Normalization by trace

Suppose $x = \xi(\omega)$ is the pattern of the object ω and $\mathbf{E}(\langle P\xi, \xi\rangle|S_i) = \mathrm{tr}\, PK_i$, $i = 1, 2$, are conditional energy distributions on projections. The conditional energy distributions on projections of different classes are not equivalent if the traces of the correlation operators K_i, $i = 1, 2$, are not equal. It is possible to normalize the conditional energy distribution on projections by normalizing the pattern of objects of each class as follows: $\eta_i = \xi/\sqrt{\mathrm{tr}\, K_i}$, $i = 1, 2$. Then the correlation operators will be normalized as follows: $\bar{K}_1 = K_1/\mathrm{tr}\, K_1$, $\bar{K}_2 = K_2/\mathrm{tr}\, K_2$, where $\mathrm{tr}\, \bar{K}_1 = \mathrm{tr}\, \bar{K}_2 = 1$. Also it is necessary to normalize the object patterns $x = \xi(\omega)$ in the decision rule (13). So, we have the following decision rule: $\omega \in S_1$ if $\langle P_1 x, x\rangle/\mathrm{tr}\, K_1 > \langle P_2 x, x\rangle/\mathrm{tr}\, K_2$ and $\omega \in S_2$ otherwise.

Example 2. We consider a classical recognition task of two classes: the class S_1 is a random signal $\xi = a + \eta$, where a is a known non-random signal and η is a white noise; the class S_2 is a white noise η. Suppose $p_1 = p_2 = 1/2$.

The correlation matrix of white noise η is $\sigma^2 I$, where σ^2 is a constant and I is an identity matrix. The mathematical expectations of the random signals of classes S_i, $i = 1, 2$, are respectively $m_1 = a$, $m_2 = 0$. Applying the decision rule of example 1, the classifier always decides that all objects $\omega \in S_1$.

We normalize the correlation matrices of both classes by their trace. From (8), we have $K_1 = \sigma^2 I + \|a\|^2 p_{\bar{a}}$, where $p_{\bar{a}} x = \langle x, \bar{a}\rangle \bar{a}$, $\bar{a} = a/\|a\|$; we also have $K_2 = \sigma^2 I$. Then $\mathrm{tr}\, K_1 = \sigma^2 \mathrm{tr}\, I + \|a\|^2 \mathrm{tr}\, p_{\bar{a}} = n\sigma^2 + \|a\|^2$ and $\mathrm{tr}\, K_2 = n\sigma^2$. Since covariance matrices of both classes are $\sigma^2 I$, they are diagonal in any basis. We choose a basis in H such that $e_1 = \bar{a}$. Then the matrix $p_1\bar{K}_1 - p_2\bar{K}_2 = 1/2(K_1/\operatorname{tr} K_1 - K_2/\operatorname{tr} K_2)$ is diagonal in the chosen basis with following eigenvalues:

$$\frac{(n-1)\|a\|^2}{2n(n\sigma^2+\|a\|^2)}, \dots, -\frac{\|a\|^2}{2n(n\sigma^2+\|a\|^2)}, -\frac{\|a\|^2}{2n(n\sigma^2+\|a\|^2)}.$$

Here the first eigenvalue is positive and the last $n-1$ eigenvalues are negative. So the projection P_1 is a one-dimensional projection: $P_1x = \langle x, e_1\rangle e_1$. Then $\langle P_1x, x\rangle = \langle x, a\rangle^2/\|a\|^2$ and $\langle P_2x, x\rangle = \langle (I-P_1)x, x\rangle = \langle x, x\rangle - \langle x, a\rangle^2/\|a\|^2$.

We denote the signal-to-noise ratio as $\mathrm{SNR} = \|a\|^2/\sigma^2$. Normalizing the object pattern by the trace, we get from (13) the following decision rule: $\omega \in S_1$ if $\langle x, a\rangle^2/(1+\mathrm{SNR}/n) > \|x\|^2\|a\|^2 - \langle x, a\rangle^2$ and $\omega \in S_2$ otherwise.

We have $\operatorname{tr} P_2K_1 = \sigma^2 \operatorname{tr} P_2 = (n-1)\sigma^2$ and $\operatorname{tr} P_1K_2 = \sigma^2 \operatorname{tr} P_1 = \sigma^2$. Then

$$\mathrm{Enr}_{\mathrm{E}}(P_1, P_2) = \frac{p_1}{\operatorname{tr} K_1}\operatorname{tr} P_2K_1 + \frac{p_2}{\operatorname{tr} K_2}\operatorname{tr} P_1K_2 = \frac{(n-1)\sigma^2}{2(n\sigma^2+\|a\|^2)} + \frac{\sigma^2}{2n\sigma^2} = \frac{n-1}{2(n+\mathrm{SNR})} + \frac{1}{2n}.$$

Thus the energy of "error" recognition depends only on the dimensionality n of the feature space and SNR. The energy of "error" recognition is approximately equal to $1/2n$ if SNR is high enough.

9. Normalization by signal norm

We can normalize object patterns by normalizing each signal $x = \xi(\omega)$ to have the unit length. In that case, ends of normalized random vectors are located on a unit sphere. Suppose $\mathbf{P}(\xi = O) = 0$. Putting $\eta = \xi/\|\xi\|$, we have

$$\mathbf{E}\langle \eta, \eta\rangle = \mathbf{E}\big(\langle \xi, \xi\rangle/\|\xi\|^2\big) = \mathbf{E}\big(\|\xi\|^2/\|\xi\|^2\big) = 1. \tag{16}$$

Let $\bar{K}$ be the correlation operator of the normalized random signal η. From (9) and (16), we have $\operatorname{tr} \bar{K} = 1$. Hence, the energy distribution on projections is normalized.

If $x = \xi(\omega)/\|\xi(\omega)\|$, i.e., $\|x\| = 1$, then $g_i(x) = \langle P_ix, x\rangle \le 1$, $i = 1, 2$. This yields that $\sup g_i(x) = 1$, $i = 1, 2$. So the discriminant functions $g_i(x)$, $i = 1, 2$ are classical membership functions [9].

Vectors x and λx for any $\lambda > 0$ describe the same physical state in quantum mechanics. It means that states of quantum systems are rays, i.e. points of projective space. Due to this fact, we can consider states with unit norm $\|x\| = 1$ only.

The same holds for sound signals and monochrome images. In fact, the sound signals x and λx for any $\lambda > 0$ differ in loudness only. The monochrome images can be described as a set of $l = nm$ real numbers corresponding to the intensity of the light in each pixel. Hence the space of the monochrome images can be described as a vector space of dimension $l = nm$. All the intensities of the monochrome image can be multiplied by a number $\lambda > 0$, but that does not change monochrome image.

10. Subtraction of mean

The following hypothesis is accepted in the recognition theory: the distribution of the patterns of a class is concentrated in a compact area of a feature space. It is natural to assume that distribution of patterns is grouped around the mean (mathematical expectation) of this distribution. Then each object pattern $x = \xi(\omega)$ can be written as the sum $x = y + a$, where a is the mean and y is the random vector from the compact area such that its beginning is the end of the mean a.

If $\mathbf{P}(\xi = a) = 1$, where a is a non-random signal, then $\mathbf{E}\langle A\xi, \xi\rangle = \langle Aa, a\rangle$, the correlation operator is the rank-one operator p_a, and the covariance operator is the zero operator. Using (9), we have $\langle Aa, a\rangle = \operatorname{tr} Ap_a = \operatorname{tr} p_aA$. Hence, applying (9) and taking into account that the orthogonal projection is a self-adjoint operator we have

$$\mathbf{E}\langle P\xi, \xi\rangle = \int_H \langle Px, x\rangle \nu(\mathrm{d}x) = \operatorname{tr} RA + \|Pm\|^2 = \operatorname{tr} AR + \|Pm\|^2. \tag{17}$$

Therefore, we can use the formula (17) for the energy distribution on projections if we know the mean m, and the covariance operator R of a random signal.

Suppose the conditional distributions $\nu_i(A) = \mathbf{P}(\xi \in A|S_i)$, $i = 1, 2$, have the covariance operators R_1, R_2 and means m_1, m_2. Then from (11) and (17) it follows that we can maximize

$$\mathrm{Enr}_{\mathrm{R}}(P_1, P_2) = p_1 \operatorname{tr} P_1R_1 + p_2 \operatorname{tr} P_2R_2 + \|P_1m_1\|^2 + \|P_2m_2\|^2$$

instead of (11).

References

[1] G. Melnichenko, Raspoznavanie signalov s razlichnymi korreliacionymi matricami pri pomoschi kvantovoj logiki, Liet. Mat. Rink. 45 (2005) 557–561 (special issue) (in Russian).
[2] V.P. Belavkin, V.P. Maslov, Design of Optimal Dynamic Analyzers: Mathematical Aspects of Wave Pattern Recognition. http://arxiv.org/abs/quant-ph/0412031, 1988.
[3] E. Oja, Subspace Methods of Pattern Recognition, Research Studies Press, Letchworth, 1983.
[4] J. Laaksonen, Subspace Classifiers in Recognition of Handwritten Digits, Doctoral Thesis, Acta Polytechnica Scandinavica., Ma 84, 1997.
[5] T. Kohonen, Self-Organization and Associative Memory, Springer Verlag, Berlin, 1989.
[6] Z. Cvetkovic, B. Beferull-Lozano, A. Buja, Robust phoneme discrimination using acoustic waveforms. in: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, Orlando, FL, vol. 1, 2002.
[7] M. Nishida, Y. Ariki, Speaker verification by integrating dynamic and static features using subspace method. in: Proceedings of the 6th International Conference on Spoken Language Processing, ICSLP, vol. 3, 2000.
[8] Y.C. Eldar, A.V. Oppenheim, Quantum signal processing, Signal Process. Mag. 19 (2002) 12–32.
[9] L.A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968) 421–427.
[10] J. Neveu, Mathematical Foundations of the Calculus of Probability, Holden-Day, San Francisco, 1965.
[11] C.W. Helstrom, Quantum Detection and Estimation Theory, Academic Press, New York, 1976.