MATHEMATICA — REVUE D'ANALYSE NUMÉRIQUE
ET DE THÉORIE DE L'APPROXIMATION
L'ANALYSE NUMÉRIQUE ET LA THÉORIE DE L'APPROXIMATION
Tome 11, N° 1—2, 1982, pp. 175—183

INEQUALITIES CONCERNING A RANDOM COMPUTATIONAL LENGTH OF PATTERN RECOGNIZERS

by

BRUNO VISCOLANI
(Padova)

Abstract

The computational length of an algorithm for pattern recognition by absolute comparison is a random variable, whose features depend on the probability distribution $\pi \in \mathbf{R}^k$ of the k classes to be discriminated. The Euclidean distance from the uniform probability distribution to any other distribution π is proportional to the greatest lower bound of the total variation of the mean computational length of algorithms used to recognize classes with distribution π. This result is reached by first finding a suitable basis of $\mathbf{R}^k$ which allows simple representations of probability distributions and of the functions under study. Furthermore, using the same basis, the Schwartz inequality easily gives an upper bound of the total variation of the mean computational length.

1. Introduction

The computational length ν of an algorithm for pattern recognition by absolute comparison, as introduced in another paper [1], is a random variable which takes up integral values in $\{1, 2, \ldots k\}$, where $k \geqslant 3$ is the number of classes to be discriminated. The mean computational length $E(\nu)$ has a smallest value $E_{\min}$ and a greatest value $E_{\max}$ such that $E_{\min} + E_{\max} = k + 1$, furthermore they depend on the probability distribution $\pi \equiv (p_1, p_2, \ldots, p_k)$ of the classes to be discriminated.

The *total variation* $V(\pi) = E_{\max} - E_{\min}$ of the mean computational length is

(1) $$V(\pi) = k + 1 - 2\sum_{i=1}^{k} p_i \alpha_i^{-1}$$

where $\alpha \in G_k$ is a permutation such that

(2) $$i < j \Rightarrow p_{\alpha_i} \geqslant p_{\alpha_j}, \text{ for all } i, j \in \{1, 2, \dots k\}.$$

In the present paper we point out two inequalities concerning $V(\pi)$, which are independent of any permutation $\alpha \in G_k$.

2. Domain of V

Let $\Omega \subset \mathbf{R}^k$ be the set of the probability distributions of k classes. If $\pi \equiv (p_1, p_2, \dots p_k) \in \Omega$ is known and with reference to any permutation $\sigma \in G_k$, we define a new probability distribution $\sigma * \pi$ in the following way

(3) $$\sigma * \pi \equiv (p_{\sigma_1}, p_{\sigma_2}, \dots p_{\sigma_k}).$$

For a fixed $\pi \in \Omega$ let α be such a permutation that (2) holds and let

$$\delta = \alpha * \pi \equiv (\delta_1, \delta_2, \dots \delta_k),$$

then

$$\delta_i \geqslant \delta_{i+1}, \quad (i = 1, 2, \dots k - 1).$$

Definition (3) implies that

$$\gamma * (\sigma * \pi) = (\gamma * \sigma) * \pi, \text{ for all } \gamma, \ \sigma \in G_k, \ \pi \in \Omega,$$

and in particular

$$(\alpha\sigma^{-1}) * (\sigma * \pi) = (\alpha\sigma^{-1}\sigma) * \pi = \alpha * \pi, \text{ for all } \sigma \in G_k, \ \pi \in \Omega.$$

From the above results and from definition (1) of V, it follows

(4) $$V(\sigma * \pi) = V(\pi), \text{ for all } \sigma \in G_k, \ \pi \in \Omega.$$

Therefore we may start analyzing the function $V(\pi)$ in the domain $\Omega^* \subset \Omega$ of the non-increasing probability distributions:

$$\Omega^* = \{\pi \equiv (p_1, p_2, \dots p_k) \in \Omega : p_1 \geqslant p_2 \geqslant \dots \geqslant p_k\}.$$

Now, definition (1) becomes

(5) $$V(\pi) = k + 1 - 2\sum_{i=1}^{k} i p_i, \text{ for all } \pi \in \Omega^*,$$

which is simpler than (1), because (5) is a permutation-independent expression.

Let $\Gamma = \{\gamma^{(1)}, \gamma^{(2)}, \dots \gamma^{(k)}\}$ be a family of k vectors of $\mathbf{R}^k$ defined in the following way:

(6a) $$\gamma^{(j)} = (\gamma_1^{(j)}, \gamma_2^{(j)} \dots \gamma_k^{(j)}),$$

(6b) $$\gamma_i^{(k)} = \frac{1}{\sqrt{k}}, \quad (i = 1, 2, \dots k),$$

(6c) $$\gamma_i^{(j)} = \begin{cases} 0 & \text{if } i < j, \\ (k - j)v_j^{-1} & \text{if } i = j, \\ -v_j^{-1} & \text{if } i > j, \ (i = 1, 2, \dots k; j = 1, 2, \dots k - 1), \end{cases}$$

(6d) $$v_j = \sqrt{(k - j)(k - j + 1)}, \quad (j = 1, 2, \dots k).$$

We can easily verify that Γ is an ortonormal basis of $\mathbf{R}^k$. Every non-increasing probability distribution $\pi \in \Omega^*$ is represented as a linear function of the vector in the basis Γ:

(7) $$\pi = \sum_{j=1}^{k} \tau_j \gamma^{(j)} = \pi(\tau),$$

where $\tau \equiv (\tau_1, \tau_2, \dots \tau_k) \in \mathbf{R}^k$.

From (6) we obtain

$$\sum_{i=1}^{k} \gamma_i^{(j)} = \begin{cases} 0 & \text{if } j < k, \\ \sqrt{k} & \text{if } j = k. \end{cases}$$

Then, $(p_1, p_2, \dots p_k) = \pi(\tau)$ being a probability distribution, it holds

$$1 = \sum_{i=1}^{k} p_i = \sum_{i=1}^{k}\sum_{j=1}^{k} \tau_j \gamma_i^{(j)} = \sum_{j=1}^{k} \tau_j \sum_{i=1}^{k} \gamma_i^{(j)} = \tau_k \sqrt{k},$$

so that

(8) $$\tau_k = \frac{1}{\sqrt{k}}.$$

For the other parameters it must first hold

(9) $$\tau_j \geqslant 0, \quad (j = 1, 2, \dots k - 1).$$

In fact, if (9) does not hold and e.g. $(p_1, \dots p_k) = \pi(\tau)$ with

$$\tau_h < 0, \ \tau_{h+1} > 0,$$

for a particular index $h < k$, then we verify that

$$p_h < p_{h+1} \text{ and } \pi(\tau) \notin \Omega^*.$$

Secondly the condition $p_i \geqslant p_{i+1}$, by (6) and (7), implies

$$\frac{1}{k} - \sum_{j=1}^{i-1} v_j^{-1}\tau_j + (k - i)v_i^{-1}\tau_i \geqslant \frac{1}{k} - \sum_{j=1}^{i} v_j^{-1}\tau_j + (k - i - 1)v_{i+1}^{-1}\tau_{i+1},$$
$$(i = 1, 2, \dots k - 2),$$

which is equivalent to

$$v_{i+1}\tau_{i+1} \leqslant v_i\tau_i, \quad (i = 1, 2, \dots k-2). \tag{10}$$

Finally the nonnegativity condition, $p_i \geqslant 0$, becomes

$$\sum_{j=1}^{k-1} v_j^{-1}\tau_j \leqslant \frac{1}{k}. \tag{11}$$

Conversely every vector $\pi = \pi(\tau) \in \mathbf{R}^k$, defined by (7) as a function of a vector $\tau \in \mathbf{R}^k$ for which conditions (8) to (11) hold, is a non-increasing probability distribution, $\pi \in \Omega^*$.

Let us indicate by T the set of all vectors $\tau \in \mathbf{R}^k$ such that conditions (8) to (11) hold.

We notice that, if $\tau \in T$, by (10) it holds

$$h < k, \ \tau_h = 0 \Rightarrow \tau_{h+j} = 0, \ (j = 1, \dots k - h - 1); \tag{12}$$

therefore, if $\pi = (p_1, p_2 \dots p_k)$, we obtain

$$h < k, \ \tau_h = 0, \ \pi = \pi(\tau) \Rightarrow p_h = p_{h+1} = \dots = p_k. \tag{13}$$

3. Permutation-independent lower bound

We prove that

$$V(\pi) \geqslant V_1(d(\pi, u\ )), \text{ for all } \pi \in \Omega, \tag{14}$$

where

$$V_1(t) = v_1 t, \quad t \in \mathbf{R}, \tag{15}$$

$$u_k = \frac{1}{\sqrt{k}}\gamma^{(k)} \in \Omega^*, \tag{16}$$

d is the Euclidean distance in $\mathbf{R}^k$ and $\gamma^{(k)}$, v_1 are defined by (6). u_k is the uniform probability distribution. We first prove (14) for $\pi \in \Omega^* \subset \Omega$, i.e. for the vectors $\pi = \pi(\tau)$ with $\tau \in T$. Computation of $V(\pi(\tau))$, by (5) and (6), gives:

$$\begin{aligned} V(\pi(\tau)) &= k + 1 - 2\sum_{i=1}^{k} i \sum_{j=1}^{k} \tau_j\gamma_i^{(j)} = \\ &= k + 1 - 2\left\{\sum_{j=1}^{k-1}\tau_j\sum_{i=j}^{k} i\gamma_i^{(j)} + \tau_k\sum_{i=1}^{k} i\gamma_i^{(k)}\right\} = \\ &= -2\sum_{j=1}^{k=1}\tau_j v_j^{-1}\left[(k-j)j - \sum_{i=j+1}^{k} i\right] = \\ &= \sum_{l=1}^{k-1} v_j\tau_j, \text{ for all } \tau \in T. \end{aligned} \tag{17}$$

On the other hand, from (7) and (16) we derive

$$\pi(\tau) = u_k + \sum_{j=1}^{k-1}\tau_j\gamma^{(j)}, \ \tau \in T;$$

therefore, since the basis Γ is an ortonormal one, it holds

$$d(\pi(\tau), u_k) = \| \pi(\tau) - u_k \| = \left\| \sum_{j=1}^{k-1} \tau_j\gamma^{(j)} \right\| = \sqrt{\sum_{j=1}^{k-1}\tau_j^2} \text{ for all } \tau \in T. \tag{18}$$

Now, since both $V(\pi)$ and $V_1(d(\pi, u_k))$ assume nonnegative values for every $\pi \in \Omega^*$, we have only to prove the inequality

$$V^2(\pi) \geqslant V_1^2(d(\pi, u_k)), \text{ for all } \pi \in \Omega^*. \tag{19}$$

Setting

$$\begin{aligned} Z(\tau) &= V^2(\pi(\tau)) - V_1^2(d(\pi(\tau), u_k)) = \\ &= \left[\sum_{j=1}^{k-1} v_j\tau_j\right]^2 - v_1^2\sum_{j=1}^{k-1}\tau_j^2, \end{aligned} \tag{20}$$

(19) becomes

$$Z(\tau) \geqslant 0, \text{ for all } \tau \in T. \tag{19'}$$

The quadratic form $Z(\tau)$ is not positive semidefinite, then the inequality (19′) does not hold for all $\tau \in \mathbf{R}^k$ and $Z(\tau)$ must be examined in detail in the domain T.

The special case for which $\tau_2 = 0$ (i.e. $\pi(\tau) = (p_1, \dots p_k)$ with $p_1 \geqslant p_2 = p_3 = \dots = p_k$) yields

$$Z(\tau) = Z\left(\tau_1, 0, \dots 0, \frac{1}{\sqrt{k}}\right) = 0,$$

for all $\tau_1 \geqslant 0$ feasible with conditions (8)—(11), i.e. for all $\tau_1 \in \left[0, \sqrt{\frac{k-1}{k}}\right]$.

In this case, the inequality (19′) is trivially verified. We note incidentally that $\tau_1 = d\left(\pi\left(\tau_1, 0, \dots 0, \frac{1}{\sqrt{k}}\right), u_k\right)$. In general, if $\tau_h > 0$ and $h = k-1$ or $\tau_j = 0$ for $h < j \leqslant k-1$, then by (10) and (20) it holds

$$\begin{aligned} Z(\tau) &= \sum_{j=2}^{h}(v_j^2 - v_1^2)\tau_j^2 + 2\sum_{j=2}^{h}\left[\sum_{i=1}^{j-1} v_i\tau_i\right]v_j\tau_j \geqslant \\ &\geqslant \sum_{j=2}^{h}(1 - v_1^2 v_j^{-2})v_j^2\tau_j^2 + 2\sum_{j=2}^{h}\sum_{i=1}^{j-1} v_j^2\tau_j^2 = \sum_{j=2}^{h} c_j v_j^2\tau_j^2, \end{aligned} \tag{21}$$

where

$$c_j = 2j - 1 - v_1^2 v_j^{-2}, \ j = 2, 3, \dots k-1. \tag{22}$$

If $k = 3$, then $c_2 = 0$ and from (21) we obtain

(23) $$k = 3 \Rightarrow Z(\tau) \geqslant 0, \text{ for all } \tau \in T.$$

Upon definition of the third degree polynomial

(24) $$Q(x) = 2\left(x - \frac{1}{2}\right) \cdot (x - k) \cdot (x - (k + 1)),$$

the coefficients c_j assume the form

(25) $$c_j = [Q(j) - k(k-1)]v_j^{-2}, \quad (j = 2, 3, \ldots k - 1).$$

It is easily verified that

(26) $$Q(2) > k(k-1), \text{ for all } k > 3,$$

and consequently:

(26′) $$c_2 > 0, \text{ for all } k > 3.$$

By the fact that $Q(k) = 0$, $Q(x) > 0$, for $\frac{1}{2} < x < k$, and by (26), we see that there exists a real number $x^* = x^*(k)$ such that

$$Q(x^*) = k(k-1), \quad 2 < x^* < k,$$

and x^* is unique.
Therefore by (25) we obtain

(27) $$\begin{cases} c_j \geqslant 0 & \text{if } j \leqslant x^*(k), \\ c_j < 0 & \text{if } j > x^*(k). \end{cases}$$

Let us define the numbers

(28) $$s_j = \sum_{i=2}^{j} c_i, \quad (j = 2, 3, \ldots k - 1),$$

then from (26′) and (27) we obtain

(29) $$\begin{cases} s_2 > 0, \\ s_j \geqslant s_{j-1}, & \text{if } j \leqslant x^*(k), \\ s_j < s_{j-1}, & \text{if } j > x^*(k). \end{cases}$$

Moreover it holds

(30) $$s_{k-1} = 0, \text{ for all } k \geqslant 3.$$

Indeed, (28) and (22) give

(31) $$s_{k-1} = k(k-1)\left[\frac{k-2}{k-1} - G(k)\right],$$

where

(32) $$G(k) = \sum_{i=2}^{k-1} v_i^{-2} = \sum_{i=2}^{k-1} \frac{1}{i(i-1)}.$$

Now $G(3) = \frac{1}{2} = \frac{3-2}{3-1}$ and nothing that, by (32),

$$G(k+1) = G(k) + \frac{1}{k(k-1)},$$

it holds

$$G(k) = \frac{k-2}{k-1} \Rightarrow G(k+1) = \frac{k-2}{k-1} + \frac{1}{k(k-1)} = \frac{(k+1)-2}{(k+1)-1}.$$

Therefore

(33) $$G(k) = \frac{k-2}{k-1}, \text{ for all } k \geqslant 3.$$

Finally, (31) and (33) imply the truth of (30). From the relations (29) and (30) it follows

(34) $$s_j > 0, \quad (j = 2, \ldots k - 2).$$

We are now able to prove inequality (19′). By (21), if $\tau_h > 0$ and $\tau_j = 0$ $(h < j < k)$ or $h = k - 1$, we have

(35) $$Z(\tau) = s_2 v_2^2 \tau_2^2 + \sum_{j=3}^{h} c_j v_j^2 \tau_j^2 \geqslant s_3 v_3^2 \tau_3^2 + \sum_{j=4}^{h} c_j v_j^2 \tau_j^2,$$

because $s_2 > 0$ and (10) imply $s_2 v_2^2 \tau_2^2 \geqslant s_2 v_3^2 \tau_3^2$.
If we assume

$$Z(\tau) \geqslant s_{i-1} v_{i-1}^2 \tau_{i-1}^2 + \sum_{j=i}^{h} c_j v_j^2 \tau_j^2,$$

then the relations (34) and (10) yield

(36) $$Z(\tau) \geqslant s_i v_i^2 \tau_i^2 + \sum_{j=i+1}^{h} c_j v_j^2 \tau_j^2,$$

for all $i < h - 1$.
From (35) and (36) we can conclude that, if $\tau_h > 0$ and $\tau_j = 0$ $(h < j < k)$ or $h = k - 1$, it holds

$$Z(\tau) \geqslant s_h v_h^2 \tau_h^2 \geqslant 0$$

and (19′) is proved, i.e. (14) is true for all $\pi \in \Omega^*$.
In order to extend the result all over Ω, we have the following proof.

If $\alpha = \alpha(\pi) \in G_k$ is such a permutation that (2) holds, then

$$\alpha(\pi) * \pi \in \Omega^*, \text{ for all } \pi \in \Omega,$$

and, by (4) and (19), we obtain

$$V(\pi) = V(\alpha(\pi) * \pi) \geqslant V_1[d(\alpha(\pi) * \pi, u_k)] = = V_1(d(\pi, u_k)), \text{ for all } \pi \in \Omega.$$

The proof of (14) is completed.

4. Permutation-independent upper bound

We prove that

(37) $$V(\pi) \leqslant V_2(\pi), \text{ for all } \pi \in \Omega,$$

where

(38) $$V_2(\pi) = k\left(p^* - \frac{1}{k}\right) + \sqrt{\frac{k(k-1)(k-2)}{3}}\sqrt{d^2(\pi, u_k) - \frac{k}{k-1}\left(p^* - \frac{1}{k}\right)^2}$$

and

(39) $$p^* = p^*(\pi) = \max\{p_1, p_2, \dots p_k\}.$$

Taking $\pi = \pi(\tau)$, $\tau \in T$, it holds

$$p^* - \frac{1}{k} = (k-1)v_1^{-1}\tau_1 = \sqrt{\frac{k-1}{k}}\tau_1$$

and, upon substitution in (38), we find

(40) $$V_2(\pi(\tau)) = v_1\left\{\tau_1 + \sqrt{\frac{k-2}{3}}\sqrt{d^2(\pi(\tau), u_k) - \tau_1^2}\right\},$$

for all $\pi = \pi(\tau)$, $\tau \in T$, i.e. $\pi \in \Omega^*$.
By (17) we can write

$$V(\pi(\tau)) = V_1\tau_1 + \sum_{j=2}^{k-1} v_j\tau_j$$

and the Schwartz inequality gives

(41) $$V(\pi(\tau)) \leqslant v_1\sigma_1 + \sqrt{\sum_{j=2}^{k-1} v_j^2}\sqrt{\sum_{j=2}^{k-1}\tau_j^2} = = v_1\tau_1 + v_1\sqrt{\frac{k-2}{3}}\sqrt{d^2(\pi(\tau), u_k) - \tau_1^2}, \text{ for all } \tau \in T.$$

The last equality follows from (18).
The relations (40) and (41) say that (37) is true for all $\pi \in \Omega^*$. Finally (37) holds for all $\pi \in \Omega$ because of (4) and of the existence of a permutation $\alpha = \alpha(\pi)$ such that $\alpha(\pi) * \pi \in \Omega^*$, for all $\pi \in \Omega$.
The proof of (37) is completed.

REFERENCES

[1] B. Viscolani, *Computational lenght in pattern recognizers*, (unpublished).

Received 2.VIII.1981.

Seminario Matematico dell'Università di Padova
Via Belzoni 7 — 35100 *Padova — Italie*