LARS Information Note 10-10-66

Purdue University

Feature Space Dimensionality Reduction Using Orthogonal Polynomials

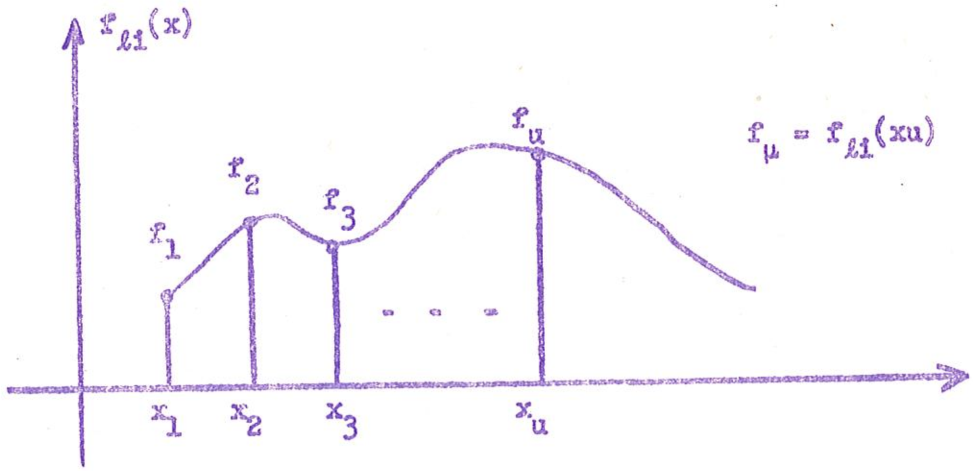
Let us consider a pattern recognition problem with j classes:

$$\omega_i \qquad i = 1, \ldots, j$$

Let $X_{\ell i}$ be the ℓ^{th} sample from class ω_i. Each sample is assumed to have m components resulting from an m-tuple of measurements:

$$X_{\ell i} \longleftrightarrow (f_1, \ldots, f_m)$$

We may consider the components of $X_{\ell i}$ as having resulted from the sampling of a sample curve $f_{\ell i}(x)$.



SAMPLE CURVE

Thus, our basic set of features $(f_1, \ldots, f_m)$ span a space of dimension m.

If $f_{\ell i}$ is well sampled, m may be large, leading to data storage and processing problems. It is then advantageous to reduce the dimension of the feature space.

We consider feature space dimensionality reduction by curve fitting using orthogonal polynomials.

Suppose we desire to fit the sample points of $f_{\ell i}(x)$, (f_μ vs x_μ) by a polynomial:

$$y_k(x) = s_o p_o(x) + \cdots + s_k p_k(x) \qquad k < m$$

where

$$p_o(x) = 1$$
$$p_1(x) = x p_o(x) - \alpha_1$$
$$p_2(x) = x p_1(x) - \alpha_2 p_1 - \beta_1 p_o(x)$$
$$\vdots$$
$$p_{i+1}(x) = x p_i(x) - \alpha_{i+1} p_i(x) - \beta_i p_{i-1}(x)$$

and, in addition, the polynomials are pointwise orthogonal; i.e.:

$$\sum_{\mu=1}^{m} p_i(x_\mu)\, p_j(x_\mu) = 0 \qquad i \neq j$$

Forsythe[1] has shown that the coefficients α_i, β_i, s_i can be determined in a simple, iterative fashion.

If we now let our sample curve be represented by $y_k(x)$, then we obtain a new set of features given by the coefficients $(s_o, \ldots, s_k)$, and the dimension of our new feature space is k + 1.

A further restriction on the polynomial $y_k(x)$ is that k be restricted such taht $k + 1 < m$; this then insures a reduction in dimensionality.

To apply this to the pattern recognition problem we may proceed as follows:

Find a number k_i such that all members of the i^{th} pattern class may be represented to within a prescribed error by a polynomial of degree k_i. Also, find a number k_{max} such that $k_{max} \geq k_i \quad i = 1, \ldots, j$.

Thus we have

$$f_{\ell i} \doteq s_{\ell o}p_o + \cdots + s_{\ell ki}p_{k_i}$$

or

$$f_{\ell i} \overset{S}{\longleftrightarrow} (s_{\ell o}, \ldots, s_{\ell k_i})$$

As an example, suppose we have $\omega_1, \omega_2, \omega_3$ and we find that for any sample in class ω_1 and a certain prespecified error, we can represent all elements in ω_1 by polynomials of proper degree $\leq k_1$. Here we may limit ourselves to $k_{max} = 25$, then k_1 may be 16
k_2 may be 18
k_3 may be 23

This does not mean that <u>all</u> elements of ω_1 require polynomials of degree 16, and hence for $f_{\ell i}$ some s_{ij} may be 0; i.e.

$$f_{\ell i} \doteq \sum_{j=0}^{k_i} s_{ij}p_j \quad \text{where some } s_{ij} \text{ may be 0 for some members of}$$

class ω_i. Thus our new set of features for the pattern recognition problem are to coefficients: $s_{\ell o}, \ldots, s_{\ell ki}$ which are in feature space of dimension $k_i \leq k_{max} \leq m$.

Some Experimental Results

The procedure described above was applied to 100 samples of Dk-2 Spectrophotometer Reflectance curves taken from the leaves of Oats. The variety within the species was not restricted in any way. The data was normalized to an absolute scale 0 to 1, corresponding to 0 to 100% reflectance.

(I) Average Mean Square Error

Figure I is a plot of the average m.s. error obtained for a given polynomial degree. The average is taken over all 100 samples. (The m.s. error is actually smaller; four place accuracy was used and other calculations give a m.s. error of .000064 at 26.)

(II) Maximum Polynomial Degree

It was found that no sample required a polynomial of degree greater than 30 for a m.s. error less than 6.4×10^{-5}.

(III) Maximum Polynomial Deviation

Although the m.s. error calculations themselves do not explicitly show it, inspection of plots of the difference between the actual curve and the fitted curve has shown a maximum deviation of 2% reflectance. Furthermore the deviation is largest where the curve amplitude is largest.

(IV) "Convergence" Properties

It was also found that, for a given m.s. error, most samples required the same polynomial degree. Fig II illustrates this for a m.s. error = 2.5×10^{-4}.

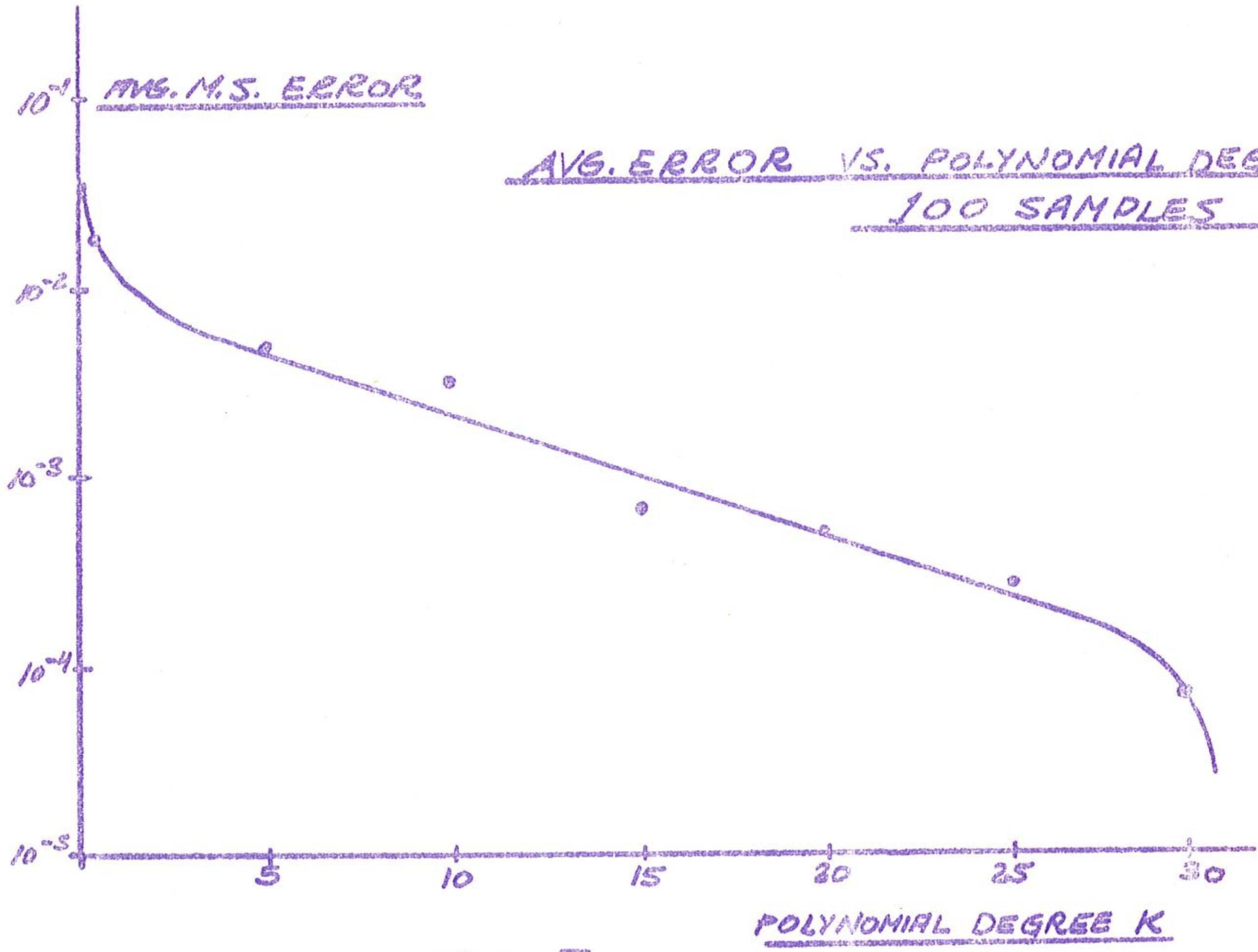

AVG. M.S. ERROR
AVG. ERROR VS. POLYNOMIAL DEGREE
100 SAMPLES
10^-1
10^-2
10^-3
10^-4
10^-5
5
10
15
20
25
30
POLYNOMIAL DEGREE K

FIG. I

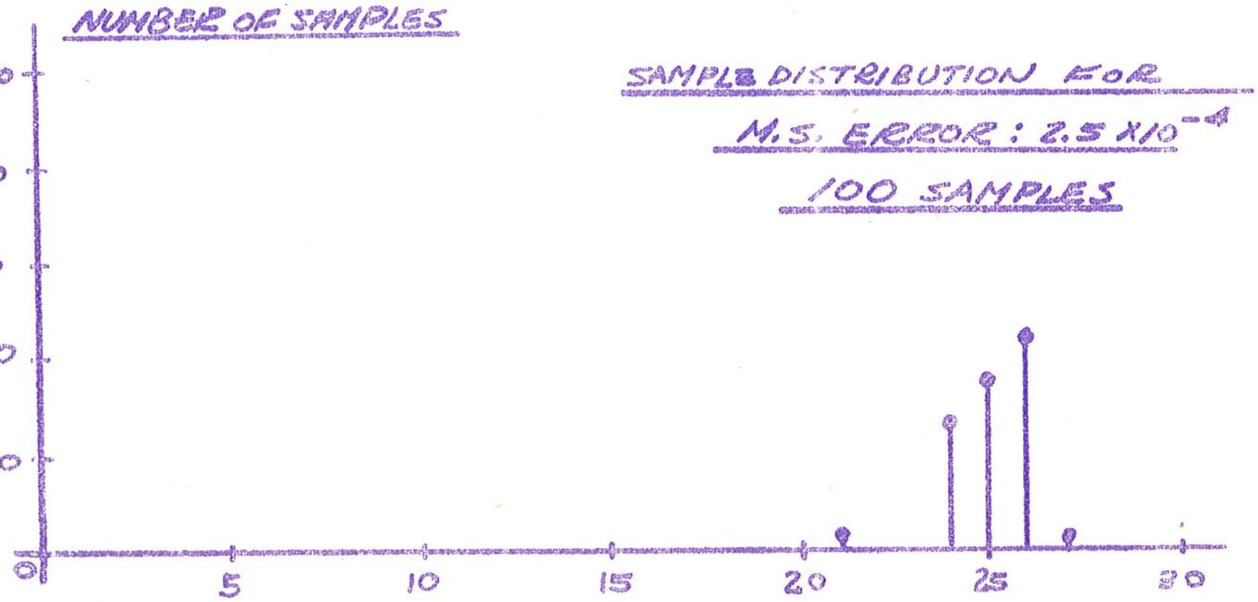

NUMBER OF SAMPLES
SAMPLE DISTRIBUTION FOR
M.S. ERROR : 2.5 X10^-4
100 SAMPLES
100
80
60
40
20
0
5
10
15
20
25
30
POLYNOMIAL DEGREE K

FIG. II

Two subsequent procedures may now be taken after we have accomplished the above:

(1) Feature Improvement Method

We may rank the features $(f_1, \ldots, f_m)$ by some criterion such as the Divergence Criterion of Marill and Green.[2] The features in m-space then have a certain rank. By data fitting we transform these m dimensional features into k_i dimensional features for ω_i. We can now compare the rank of our transformed features.

(2) Separability Improvement

Suppose we take n samples for each of the classes ω_i and ω_j. These samples now occupy a portion of the m-dimensional space, and, furthermore, the two classes may overlap. If we now apply the feature reduction method we may be able to reduce this overlap.

REFERENCES

[1] Forsythe, G.E., "Generation and Use of Orthogonal Polynomials For Data-Fitting with a Digital Computer," J. SIAM, June, 1957, pp. 75-88.

[2] Marill, T. and Green, D.M., "On the Effectiveness of Receptors in Recognition Systems," IEEE PGIT, Jan., 1963, pp. 11-17.

J. R. Sergo
LARS