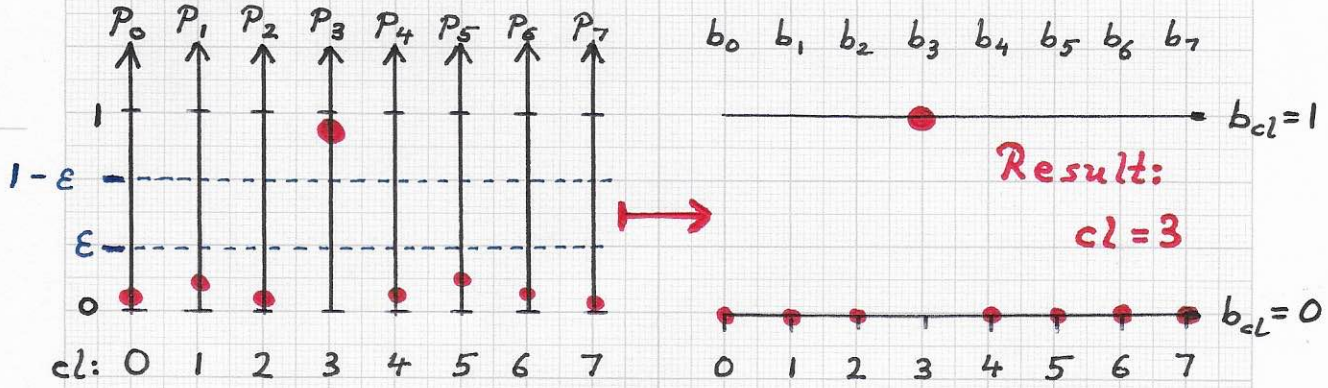
■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• Laplacian eigenfunctions and neural networks:... The mapping of the real-valued $\mathbb{P}$-tuple $(P_0, P_1, \ldots, P_c)$ of an unclassified material (segment) to a binary-valued tuple $(b_0, b_1, \ldots, b_c)$ is the most important processing step in the final phase of the material classification pipeline. Using a "parallel coordinates" plot to visualize a $\mathbb{P}$-tuple, one can think of the mapping as follows:



Mapping real P_{cl}-values to binary b_{cl}-values.

The example shown in this figure considers eight material classes, cl-0, ..., cl-7. Here, $P_{cl} \leq \varepsilon$ for $cl \neq 3$; $P_3 \geq 1-\varepsilon$. In this simple case, a threshold-based mapping approach suffices to produce the sketched result, i.e., $b_{cl} = 0$ for $cl \neq 3$; $b_3 = 1$:

$$b_{cl} = \begin{cases} 0, & \text{if } P_{cl} \leq \varepsilon \\ 1, & \text{if } P_{cl} \geq 1-\varepsilon \\ ???, & \text{otherwise} \end{cases}$$

This mapping formula, of course, is not general enough, as it does not map P_{cl}-values in the interval $(\varepsilon, 1-\varepsilon)$.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• **Laplacian eigenfunctions and neural networks: ...** Considering the overall complexity of the calculation of P_{cl}-values, it is "reasonable" to assume that they support an extremely high-quality / high-performance material classification. Nevertheless, for a multitude of reasons, it is possible that unusual, very rare, never-seen-before or abnormal classification cases arise. We can use the simplistic threshold-based example presented on the previous page to motivate the need for a more advanced approach:

(i) All P_{cl}-values are smaller than the threshold, i.e., $P_{cl} \leq \varepsilon$ for $cl = 0, \ldots, C$.

(ii) The unclassified material generates two or even more P_{cl}-values in the interval $[1-\varepsilon, 1]$.

(iii) The unclassified material generates several P_{cl}-values in the "uncertainty interval" $(\varepsilon, 1-\varepsilon)$.

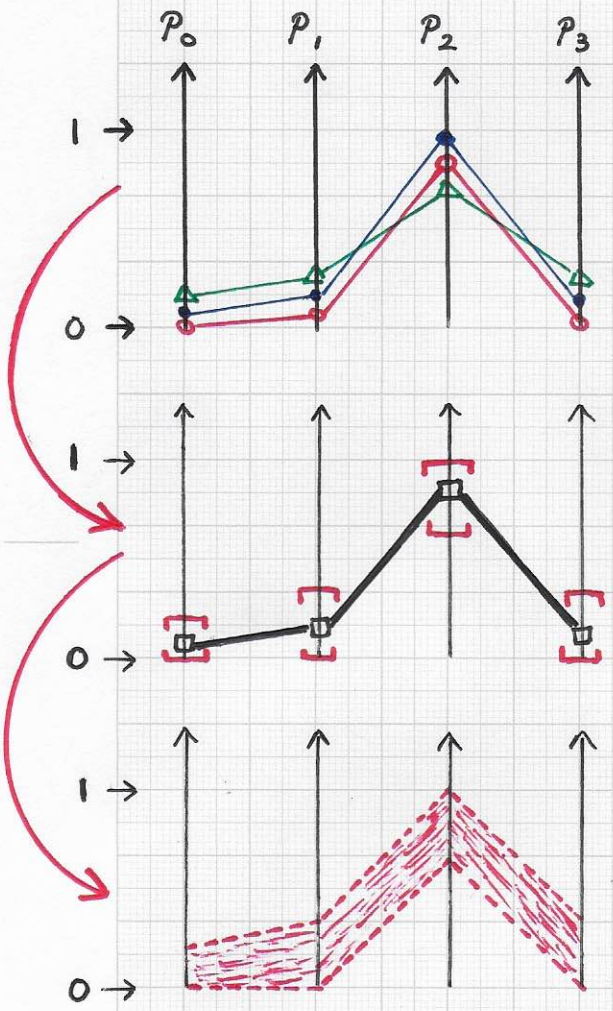
(iv) The unclassified material is classified as a material of class cl – but this classification is incorrect.

These four scenarios are merely exemplary. Other challenging situations can happen – theoretically and practically. It is therefore necessary to devise and continually improve a "clever" algorithmic method for this final mapping step. This method should be based on a/the optimization goal for classification system performance.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• Laplacian eigenfunctions and neural networks:...



Parallel coordinates clustering

The figures (left) visualize one possibility to think about the goal of optimizing classification system performance:

i) Having implemented the classification system as described, including the system's necessary training via representative material samples (stored in a sample database), we can evaluate the system's performance.

ii) For example, we use three class-2 material segments as input for the system – designed to recognize four classes (0,...,3) as shown in the figure.

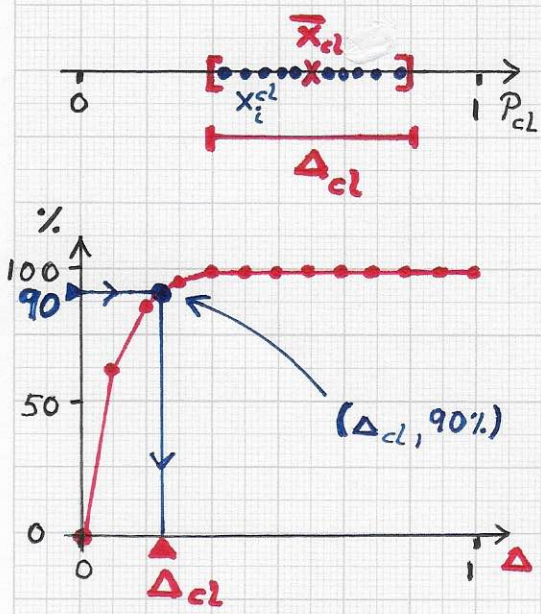
iii) The system generates the three $\mathbb{P}$-tuples that are represented via a parallel coordinates plot in the top of the three images above.

iv) The generated P_{cl}-values / $\mathbb{P}$-tuple polylines induce certain average P_{cl}-values, defining the average (cluster center) polyline shown in the middle image.

v) The average P_{cl}-values have associated "confidence intervals (middle image), defining the "confidence band" (bottom image). ...

OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• Laplacian eigenfunctions and neural networks:...



Top: Concerning the P_{cl}-axis, the average value of all recorded x_i^{cl}-values is $\bar{x}_{cL}$. This value defines the midpoint of an interval that must contain 90% of all observed x_i^{cl}-values. One must estimate the length of this interval.

Bottom: For several interval length values Δ, one determines the percentages of x_i^{cl}-values lying inside the associated interval. Thus, one obtains tuples $[\Delta, \%]$ that define the piecewise linear function/polyline shown in the figure. One can now calculate the value of Δ_{cL} by determining where on the Δ-axis this function takes on the value 90%. This example establishes a 90% "confidence interval" for P_{cl}.

For our purposes — and at this late stage of the material classification pipeline — it is possible to employ a simple, practical method for defining a "confidence band" (point v) on the previous page). A simple method could be designed and executed as follows (left figure):

→ We observe the classification system's behavior/response by using "a large number of class-2 material segments that the system is expected to classify as class-2 segments.

→ Eventually, the system generates value sets $\{x_i^{cl}\}$ for each P_{cl} variable.

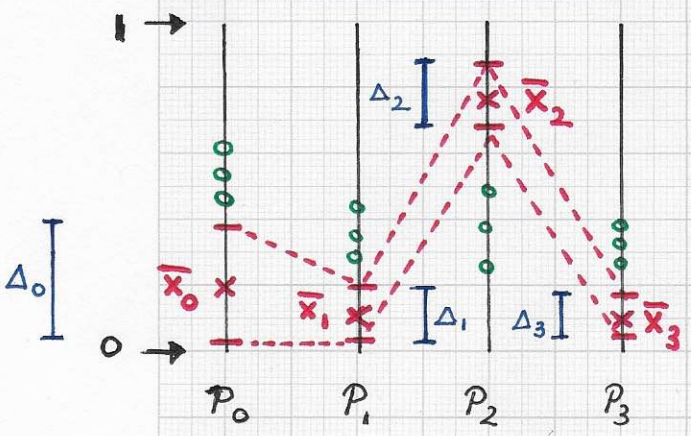
→ For each P_{cl} variable, we determine the average $\bar{x}_{cL}$ of its associated set of x_i^{cl}-values.

→ The goal is to define the length Δ_{cl} of a P_{cl}-axis interval that, for example, contains 90% of the x_i^{cl}-values and has $\bar{x}_{cl}$ as its midpoint.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks:...



Parallel coordinates plot of probabilities $P_0, \ldots, P_3$ with overlaid "confidence intervals." Intervals have midpoints $\bar{x}_{cl}$ and lengths Δ_{cl}. For each P_{cl}-axis, 90% of all generated P_{cl}-values lie inside the associated "confidence interval"; a small number of values lie outside, shown as 'o' data.

One can define a "confidence interval" as

$$I_{cl} = [\bar{x}_{cl} - \Delta_{cl}/2, \; \bar{x}_{cl} + \Delta_{cl}/2],$$

where one also might have to clip the lower and upper bounds to the values 0 and 1, respectively.

• **MEANING:**

When providing the system with a set of class-2 material data as input, the system will respond with P_{cl}-value data that satisfy $P_{cl} \in I_{cl}$ in 90% of the time.

The figure (left) illustrates a hypothetical result when performing system response characterization when using only class-2 data as input (to be classified as class-2 data). Again, 90% of the values generated on a P_{cl}-axis lie inside the interval with length Δ_{cl} and midpoint $\bar{x}_{cl}$. (For the purpose of keeping the discussion focused, we do not consider, at this point, the possibility that this interval extends beyond the "parent interval" $[0,1]$.)

If the system "performs well/as expected," then one will assume that these conditions are satisfied:

i) $\bar{x}_2 > \frac{1}{2} \wedge \bar{x}_{cl} < \frac{1}{2}$ for $cl \neq 2$.

ii) $\frac{1}{2} < MIN_2 \leq \bar{x}_2 - \Delta_2/2 \wedge \frac{1}{2} > MAX_{cl} \geq \bar{x}_{cl} + \Delta_{cl}/2$.

(If $\bar{x}_2 + \Delta_2/2 > 1$, the upper bound of the interval should be set to 1. If $\bar{x}_{cl} - \Delta_{cl}/2 < 0$, the lower bound of the interval should be set to 0.)

...