■ GEOMETRIC MODELING AND DENSITY - Cont'd.

• **"Quadratics":** We consider quadratic functions, using monomial basis functions, to discuss the best approximation of a finite set of discrete data defining density values. The density values are ϱ_i, given at locations $\mathbf{x}_i$ in a D-dimsional (feature) space ($i = 1 \ldots N$), must be approximated over $\mathbf{x}$-space, $\mathbf{x} = (x_1, \ldots, x_D)^T$. In monomial form, the approximating quadratic function is written as

$$\varrho(\mathbf{x}) = \varrho(x_1, \ldots, x_D) = \sum_{\|\mathbf{i}\| \le 2} c_{\mathbf{i}} \, \mathbf{x}^{\mathbf{i}}$$

$$= c_{0,\ldots,0} + c_{1,0,\ldots,0} x_1 + \ldots + c_{0,\ldots,0,1} x_D + c_{1,1,0,\ldots,0} x_1 x_2 + \ldots + c_{0,\ldots,0,1,1} x_{D-1} x_D + c_{2,0,\ldots,0} x_1^2 + \ldots + c_{0,\ldots,0,2} x_D^2$$

$$= \sum_{i_1,\ldots,i_D \ge 0 \;\wedge\; i_1+\ldots+i_D \le 2} c_{i_1,\ldots,i_D} x_1^{i_1} \cdot \ldots \cdot x_D^{i_D} .$$

D	no. of coeffs. $c_{\mathbf{i}}$
1	3
2	6
3	10
4	15
5	21
6	28
7	36
8	45
9	55

$\Rightarrow$ no. of coeffs. (D)
$= 1 + 2 + \ldots + D + D + 1$
$= \frac{1}{2}(D+1)(D+2)$
$= \frac{1}{2}(D^2 + 3D + 2)$
$\Rightarrow \mathcal{O}(D^2)$ storage complexity

The quadratic function $\varrho(\mathbf{x})$ is defined over the entire $\mathbf{x}$-space - without boundary. (To restrict it to a finite and 'simple' region in $\mathbf{x}$-space one can, for example, use a simplex in the D-dimensional $\mathbf{x}$-space.)

Despite the fact that the number of coefficients $c_{\mathbf{i}}$ grows quadratically with respect to dimension D, it is assumed that $D^2 \ll N$ and 'data reduction' results.

■ GEOMETRIC MODELING AND DENSITY - Cont'd.

• **'Quadratics':** The best quadratic function $\varrho(\mathbf{x})$ is defined by the given N discrete density values ϱ_j:

$$\varrho(\mathbf{x}_j) = \sum_{\|\mathbf{i}\| \le 2} c_{\mathbf{i}} \, \mathbf{x}_j^{\mathbf{i}} = \varrho_j \quad , \; j = 1 \ldots N$$

In matrix form, the set of equations is

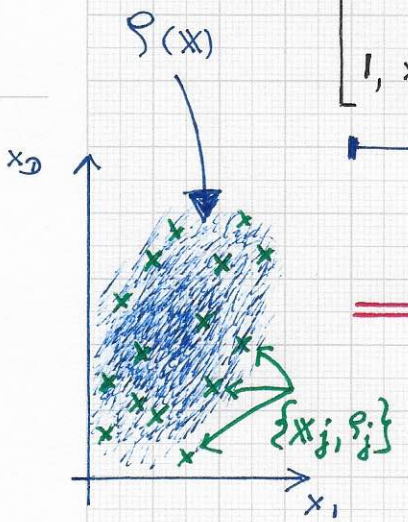
$$\begin{bmatrix} 1, x_1^1, \ldots, x_D^1, x_1^1 x_2^1, \ldots, x_{D-1}^1 x_D^1, (x_1^1)^2, \ldots, (x_D^1)^2 \\ \vdots \\ 1, x_1^N, \ldots, x_D^N, x_1^N x_2^N, \ldots, x_{D-1}^N x_D^N, (x_1^N)^2, \ldots, (x_D^N)^2 \end{bmatrix} \begin{bmatrix} c_{0,\ldots,0} \\ c_{1,0,\ldots,0} \\ c_{0,\ldots,0,1} \\ c_{1,1,0,\ldots,0} \\ c_{0,\ldots,0,1,1} \\ c_{2,0,\ldots,0} \\ c_{0,\ldots,0,2} \end{bmatrix} = \begin{bmatrix} \varrho_1 \\ \vdots \\ \varrho_N \end{bmatrix}$$

$\frac{1}{2}(D+1)(D+2)$ columns

$$\Leftrightarrow \quad X \, \mathbb{C} = \mathbb{P}$$

Least squares

$$\mathbb{C} = (X^T X)^{-1} X^T \mathbb{P} \quad ,$$



Computation of density $\varrho(\mathbf{x})$

where $\mathbf{x}_j = (x_1^j, \ldots, x_D^j)^T$.

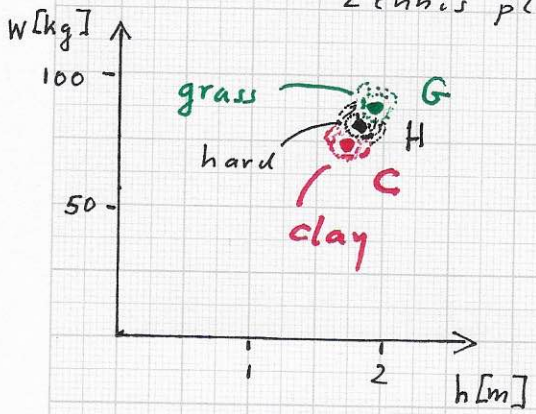
• **Ex.:** $D = 99$ (99-dim. feature space) $\Rightarrow$ 5050 coeffs.
$\Rightarrow$ desired: $N \gg 5050$ samples (feature points/vectors) in 99-dim. $\mathbf{x}$-space!

• **Error:** The resulting best approximation has error values $E_{RMS} = \left(\frac{1}{N} \sum_{j=1}^{N} [\varrho_j - \varrho(\mathbf{x}_j)]^2 \right)^{1/2}$ and

$$E_{MAX} = \max \left\{ |\varrho_j - \varrho(\mathbf{x}_j)| \right\}_{j=1}^{N} \quad .$$

■ GEOMETRIC MODELING AND DENSITY - Cont'd.

• 'Quadratics': We consider a 3-class classification problem. A tennis player performs differently on clay (C), hard (H) and grass (G) courts, based on her/his physical attributes - in a statistical sense. The top players for C, H and G define height-weight distributions (and densities) in the plane spanned by the height (h) and weight (w) axes. To classify a player as a C-, H- or G-player approximations of C-, H- and G-densities can be used.



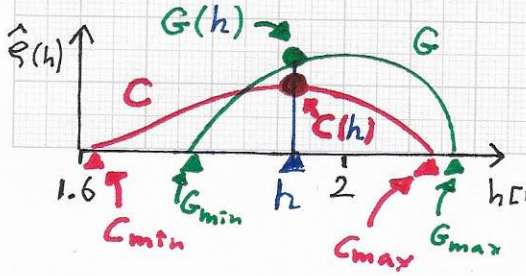
Clusters for clay, hard and grass courts

⇒ Needed: normalized C-, H-, G-densities that satisfy E_{RMS} and/or E_{MAX} requirements

⇒ Error-limited normalized densities are needed.

(i) Normalization: For our purpose (classification), a density function $\varrho(x)$ has a finite domain; outside this domain $\varrho(x)$ is zero, while $\varrho(x)$ is positive inside. Denoting the domain by Ω, one can compute the standard L_2 norm for $\varrho(x)$: $\|\varrho(x)\| = \sqrt{\langle \varrho, \varrho \rangle} = I = \int_\Omega \varrho(x) \cdot \varrho(x)\, dx$. Thus, the resulting normalized density $\hat{\varrho}(x)$ is defined as $\hat{\varrho}(x) = \varrho(x)/I$.

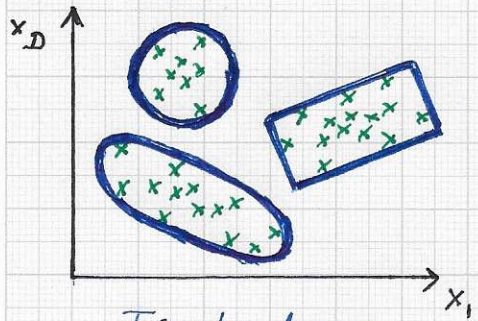
(Note: The following example demonstrates the need for 'density normalization'.



$G = G(h)$ is the (normalized) density of height for the top grass court players; $C = C(h)$... for the top clay court players. $c_{min} \le h < g_{min}$ ⇒ clay player; $c_{max} < h \le g_{max}$ ⇒ grass player; $g_{min} \le h \le c_{max}$ ⇒ probability C vs. G: $C(h) : G(h)$.)

GEOMETRIC MODELING AND DENSITY - Cont'd.

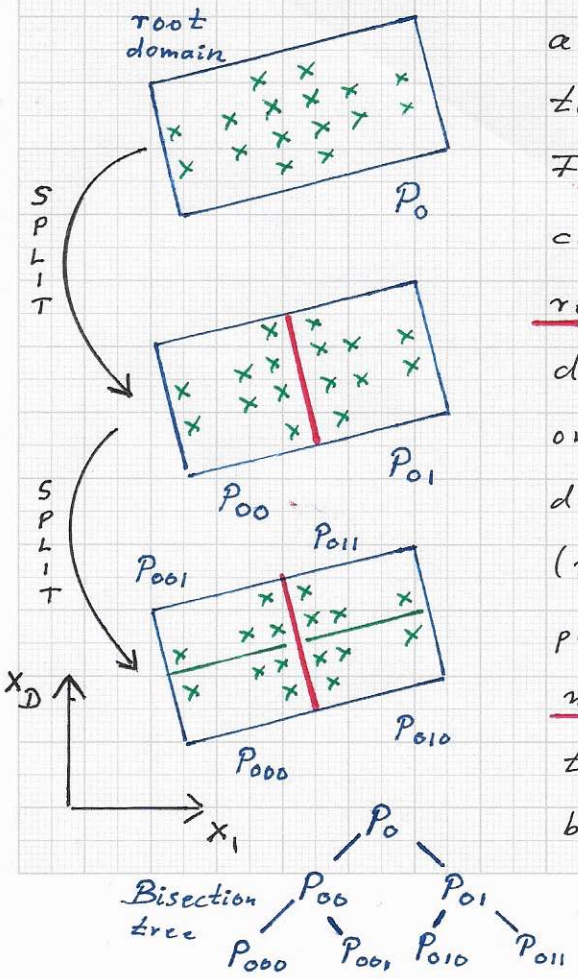
(i) Normalization with respect to a simple finite domain:



Finite domains for sample sets

The boundary of the finite domain of the best density approximation should be simple - to compute and represent. Three possibilities are a hyper-sphere with small/minimal radius, a hyper-ellipsoid with small/minimal radii, or a small/minimal parallelepiped (bounding all sample points).

(ii) Subdivision: If the error values E_{RMS} and/or E_{MAX} exceed a threshold value ε, it will be necessary to improve a density approximation. For example, a hyper-parallelepiped can be subdivided (recursively) via repeated bisection of a sample set's domain. After each bisection step one must compute (new) best quadratic density approximations, one for each (new) sub-domain. (Note: For the purpose of classification it is not necessary to ensure continuity of two approximations on a shared boundary interface. Only the reduction of the E_{RMS} and/or E_{MAX} values is of importance. Bisection stops when the error values are smaller than ε.)

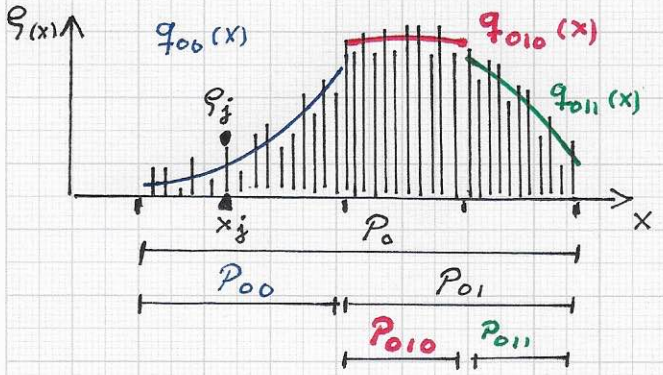


■ GEOMETRIC MODELING AND DENSITY - Cont'd.

(ii) Subdivision: → When one sub-domain is bisected its associated polynomial coefficients are 'replaced' by the two sets of coefficients computed for the two (new) child sub-domains.

→ The final best density approximation is a piecewise, discontinuous quadratic function. The domain of this piecewise quadratic function is the root domain. The final normalization step of this piecewise quadratic function is performed with respect to the original root domain.

→ If the root domain in $\mathcal{D}$-dimensional (feature) space is divided into S sub-domains, the number of all quadratic polynomial coefficients is $\frac{S}{2}(\mathcal{D}+1)(\mathcal{D}+2)$.



Discontinuous density approx.

$$\rho(x) = \begin{cases} q_{00}(x), & x \in P_{00} \\ q_{010}(x), & x \in P_{010} \\ q_{011}(x), & x \in P_{011} \\ 0, & \text{otherwise} \end{cases}$$

→ Ex. (left): Density tuples (x_j, ρ_j) are approximated by a function consisting of 3 polynomial segments — best quadratic polynomials — resulting after repeated bisection of the root domain P_0.

→ Once the piecewise quadratic MODEL has been established, one can use it for the generation of synthetic data (discrete) with the same distribution / density.

~ BH