Algorithm 1 PC-SMOTE (Percentile-Controlled SMOTE) — binario + multiclase en un único esquema

Entrada: $X \in \mathbb{R}^{n\times d}$, y (binario o multiclase).
Hiperparámetros: k, radio r, p_{dist}, p_{den} (opc), p_{ent} (opc), criterio $\in$ {entropía, proporción}, modo $d \in$ {2D,3D}, *multiclase*: factor_equilibrio $\in (0,1]$, *max_total_multiplier*, *max_sinteticas_por_clase* (opc).
Salida: (X', y').

Parte A — auxiliares

Riesgo en x_i: k-NN global en $X \Rightarrow$ vecindad $\mathcal{N}_i^{all}$; $r_i \leftarrow \frac{1}{k}\sum_{j\in\mathcal{N}_i^{all}} \mathbf{1}[y_j = \text{mayoritaria}]$.
Densidad (minoritaria): k-NN en $X_{\text{mín}} \Rightarrow \mathcal{N}_i^{min}$; $d_i \leftarrow \#\{j \in \mathcal{N}_i^{min} : \|x_i - x_j\|_2 \le 2r\}/|\mathcal{N}_i^{min}|$ ▷ en 3D usar primeras 3 coords
Pureza:
Entropía $\Rightarrow ent_i = -\sum_c p_{c,i}\log_2 p_{c,i}$ en $\mathcal{N}_i^{all}$; umbral $u_{\text{ent}} = \text{percentil}(ent, p_{\text{ent}})$ (o 1.0 si no se da); *mask_pureza*= $[\,ent_i \le u_{\text{ent}}\,]$.
Proporción $\Rightarrow q_i = \frac{1}{k}\sum_{j\in\mathcal{N}_i^{all}} \mathbf{1}[y_j = \text{minoritaria}]$; *mask_pureza*= $[\,0.4 \le q_i \le 0.6\,]$.

Parte B — núcleo binario `PC-SMOTE-BINARIO`(X, y)

$X_{\text{mín}} \leftarrow X[y=1]$, $X \leftarrow X[y=0]$.
if $|X_{\text{mín}}| < k+1$ **then**
  return (X, y)
end if
Calcular $\mathcal{N}_i^{all}$, r_i; y $\mathcal{N}_i^{min}$, d_i.
Umbral de densidad: si p_{den} dado $\Rightarrow u_{\text{den}} = \text{percentil}(d, p_{\text{den}})$ y *mask_dens*= $[\,d_i \ge u_{\text{den}}\,]$; si no, *mask_dens*= $[\,d_i > 0\,]$.
Candidatos $S = \{x_i \in X_{\text{mín}} : \textit{mask\_pureza} \wedge \textit{mask\_dens}\}$.
if $|S| < k+1$ **then**
  return (X, y)
end if
$n_{\text{sint}} \leftarrow \text{máx}(0, |X| - |X_{\text{mín}}|)$ ▷ si no se fija otro valor
$X_{\text{sint}} \leftarrow \emptyset$
while $|X_{\text{sint}}| < n_{\text{sint}}$ **do**
  Elegir $x_i \in S$; obtener $\mathcal{N}_i^{all}$ y distancias d_{ij} (usar 3D si $d = $ 3D).
  $\tau_i = \text{percentil}(d_{ij}, p_{\text{dist}})$; $\mathcal{V}_i = \{j \in \mathcal{N}_i^{all} : d_{ij} \le \tau_i\}$.
  if $|\mathcal{V}_i| = 0$ **then**
  end if
  Elegir $z \in \mathcal{V}_i$; $x_z \leftarrow X[z]$.
  Seleccionar δ según r_i:

$$\delta \sim \begin{cases} \mathcal{U}(0.6, 0.8) & \text{si } 0.4 \le r_i < 0.5, \\ \mathcal{U}(0.3, 0.5) & \text{si } 0.5 \le r_i \le 0.6, \\ \mathcal{U}(0.4, 0.6) & \text{otro caso.} \end{cases}$$

  $x_{\text{new}} = x_i + \delta(x_z - x_i)$; añadir x_{new} a X_{sint} (etiqueta 1).
end while
return $\big([X; X_{\text{sint}}], [y; \mathbf{1}]\big)$

Parte C — multiclase `PC-SMOTE-MC`(X, y)

$X' \leftarrow X$, $y' \leftarrow y$; $\mathcal{C} \leftarrow$ clases únicas; $n_0 = |y|$; $n_{\text{máx}} = \text{máx}_{c\in\mathcal{C}} |\{i : y_i = c\}|$.
for all $c \in \mathcal{C}$ **do**
  $y^{(c)} = \mathbf{1}[y = c]$; $n_{\text{act}} = |\{i : y_i = c\}|$; $n_{\text{obj}} = \lfloor n_{\text{máx}} \cdot \text{factor\_equilibrio} \rfloor$; $faltante = \text{máx}(0, n_{\text{obj}} - n_{\text{act}})$.
  Tope por clase: si *max_sinteticas_por_clase* $\Rightarrow faltante = \text{mín}(faltante, \textit{max\_sinteticas\_por\_clase})$.
  Tope global: si *max_total_multiplier* $\Rightarrow n_{\text{máx}\_total} = \lfloor n_0 \cdot \textit{max\_total\_multiplier} \rfloor$ y $faltante = \text{mín}\big(faltante,\, n_{\text{máx}\_total} - |y'|\big)$.
  if $faltante \le 0$ **then**
  end if
  Ejecutar `PC-SMOTE-BINARIO` sobre $(X, y^{(c)})$ con $n_{\text{sint}} = faltante$.
  Tomar las últimas $faltante$ sintéticas $\hat{X}$, etiquetarlas con c; concatenar: $X' = [X'; \hat{X}]$, $y' = [y'; c\mathbf{1}]$.
end for
return (X', y')