Να δειχθεί ότι η κλαση των ομόκεντρων μπαλών (δίσκων) γύρω από το $(0,0)$ στον $\mathbb{R}^2$ είναι PAC-learnable.

Λύση: Η κλάση που επιθυμούμε να μάθουμε είναι η $C = \{B_r : r > 0\}$, όπου $B_r = \{x \in \mathbb{R}^2 : \|x\|_2 \leq r\}$ είναι η μπάλα με κέντρο το $(0,0)$ και ακτίνα r. Θεωρούμε τον εξής αλγόριθμο A: Για δοθέν δείγμα S, ο A επιστρέφει τη μικρότερη μπάλα που περιέχει τα θετικά σημεία του δείγματός μας.

Συγκεκριμένα, αν $c = B_r$ η πραγματική μπάλα που προσπαθούμε να μάθουμε και S το δείγμα, ορίζουμε $S^+ = \{x \in S : c(x) = 1\}$ το σύνολο των θετικών δειγμάτων. Αν θέσουμε $r_* = \max\{\|x\| : x \in S^+\}$, τότε η μπάλα B_{r_*} είναι η μικρότερη μπάλα που περιέχει τα θετικά δείγματα και επομένως $A(S) = h_S = B_{r_*}$. Προφανώς $h_S \subseteq c$ και επομένως το ενδεχόμενο $[h_S \neq c]$ περιέχεται στο ενδεχόμενο $c \setminus h_S$.

Έστω $\varepsilon > 0$, $\delta > 0$ και D κατανομή. Αν $P(c) \leq \varepsilon$, τότε $P[h_S \neq c] \leq P(c) \leq \varepsilon$ για κάθε δείγμα S, επομένως $P_{S\sim D^m}[P_{x\sim D}[h_S \neq c] \leq \varepsilon] = 1 > 1-\delta$ και η ιδιότητα που προσπαθούμε να δείξουμε ισχύει.

Εαν $P(c) > \varepsilon$, θεωρούμε την οικογένεια των δακτυλίων $\Delta_t = \{x \in \mathbb{R}^2 : t \leq \|x\|_2 \leq r\}$. Η οικογένεια $(\Delta_t)_{t=0}^r$ είναι φθίνουσα και nested, με την ιδιότητα ότι $P(\Delta_0) = P(c) > \varepsilon$ και $P(\Delta_r) = 0$. Θέτουμε t_ε να είναι ο μεγαλύτερος δείκτης t για τον οποίον $P(\Delta_t) \geq \varepsilon$,

$$t_\varepsilon = \sup\{t > 0 : P(\Delta_t) \geq \varepsilon\}. \tag{1}$$

Προφανώς θα ισχύει ότι $P(\Delta_{t_\varepsilon}) \geq \varepsilon$. Πράγματι, αν θεωρήσουμε μια γνησίως αύξουσα ακολουθία $(t_n)_n$ η οποία συγκλίνει στο t_ε, η αντίστοιχη ακολουθία δακτυλίων $(\Delta_{t_n})_n$ θα είναι φθίνουσα με $P(\Delta_{t_n}) \geq \varepsilon$ για κάθε n και επιπλέον $\cap_{n=1}^\infty \Delta_{t_n} = \Delta_{t_\varepsilon}$, επομένως

$$P(\Delta_{t_\varepsilon}) = \lim_n P(\Delta_{t_n}) \geq \varepsilon\,. \tag{2}$$

Θέτουμε $A = \Delta_{t_\varepsilon}$. Τονίζουμε σε αυτό το σημείο ότι για την κατασκευή του A δε λάβαμε υπόψιν μας ούτε κάποιο πιθανό δείγμα S, ούτε κάποια συνάρτηση h_S, αλλά βασιστήκαμε μόνο στο αληθινό πρότυπο c, την κατανομή D και το δοθέν $\varepsilon > 0$.

Σκοπός μας είναι να φράξουμε την ποσότητα $P_{S\sim D^m}[P_{x\sim D}[h_S \neq c] > \varepsilon]$. Για να το επιτύχουμε, πρώτα δεσμεύουμε την πιθανότητα αυτή στο ενδεχόμενο το S να τέμνει το A ή όχι:

$$\begin{aligned} P_{S\sim D^m}[P_{x\sim D}[h_S \neq c] > \varepsilon] = {} & P_{S\sim D^m}[P_{x\sim D}[h_S \neq c] > \varepsilon \mid S \cap A \neq \emptyset] \cdot P_{S\sim D^m}[S \cap A \neq \emptyset] \\ & + P_{S\sim D^m}[P_{x\sim D}[h_S \neq c] > \varepsilon \mid S \cap A = \emptyset] \cdot P_{S\sim D^m}[S \cap A = \emptyset] \end{aligned} \tag{3}$$

Για να φράξουμε αποτελεσματικά την (3) αρκεί να φράξουμε τις δύο ποσότητες που χρωματίσαμε. Σημειώνουμε ότι για να υπολογίσουμε την κόκκινη ποσότητα, δε μας ενδιαφέρει η συνάρτηση h_S που επέστρεψε ο αλγόριθμός μας. Είναι μια πιθανότητα που αφορά μονάχα την τυχαία μεταβλητή $S = \{X_1, \ldots, X_m\}$ και το σύνολο A. Λόγω ανεξαρτησίας,

$$
\begin{aligned}
P_{S\sim D^m}[S \cap A = \emptyset] &= P_{S\sim D^m}\left(\cap_{i=1}^m [x_i \notin A]\right) = \prod_{i=1}^m (1 - P_{x\sim D}(A)) \\
&= (1 - P_{x\sim D}(A))^m \le (1-\varepsilon)^m, \qquad (4)
\end{aligned}
$$

επειδή $P(A) \ge \varepsilon$.

Σχετικά με την μπλε ποσότητα, τα πράγματα θα ήταν πολύ απλούστερα αν επιπλέον ίσχυε ότι $P(A) \le \varepsilon$, δηλαδή αν η $P(A)$ ήταν ακριβώς ίση με ε. Πράγματι, σε αυτή την περίπτωση, δεδομένου ότι $S \cap A \neq \emptyset$, το σύνολο h_S θα τέμνει το A και επομένως $c \setminus h_S \subseteq A$, από την οποία προκύπτει ότι $P[c \neq h_s] \le P(A) \le \varepsilon$, δηλαδή ο μπλε όρος ισούται με μηδέν για κάθε τέτοιο S.

Εν γένει όμως δεν είναι σωστό ότι $P(A) = \varepsilon$. Θυμίζουμε ότι από την κατασκευή μας έχουμε εξασφαλίσει ότι $P(A) \ge \varepsilon$, αλλά για το συγκεκριμένο A δεν μπορεί να αποδειχθεί ισότητα χωρίς κάποια επιπλέον υπόθεση για την κατανομή D. Για το λόγο αυτό, θα μικρύνουμε το A ελάχιστα, ώστε να μπορέσουμε να πάρουμε την αντίστροφη ανισότητα, χωρίς να χαλάσει βέβαια το προηγούμενο σκέλος της απόδειξης, στο οποίο το γεγονός ότι $P(A) \ge \varepsilon$ έπαιξε κεντρικό ρόλο.

Διαμερίζουμε το δακτύλιο $A = \Delta_{t_\varepsilon} = \{x \in \mathbb{R}^2 : t_\varepsilon \le \|x\|_2 \le r\}$ στα εξής δύο σύνολα, $A = \tilde{A} \cup L_{t_\varepsilon}$, με $\tilde{A} = \{x \in \mathbb{R}^2 : t_\varepsilon < \|x\|_2 \le r\}$ και $L_{t_\varepsilon} = \{x \in \mathbb{R}^2 : \|x\|_2 = t_\varepsilon\}$. Αφού το σύνολο h_S τέμνει το A, αναγκαστικά θα περιέχει και το σύνολο $L_{t_\varepsilon} = \{x \in \mathbb{R}^2 : \|x\|_2 = t_\varepsilon\}$. Επομένως δεν ισχύει μόνο ότι $c \setminus h_S \subseteq A$, αλλά επιπλέον ότι $c \setminus h_S \subseteq \tilde{A}$.

Αν για το σύνολο $\tilde{A}$ μπορέσουμε να δείξουμε ότι $P(\tilde{A}) \le \varepsilon$, τότε από το επιχείρημα που γράψαμε παραπάνω, θα έχουμε καταφέρει να μηδενίσουμε τον μπλε όρο. Καταρχάς, παρατηρούμε ότι το $\tilde{A}$ μπορεί να γραφεί ως ένωση κλειστών δακτυλίων με τον εξής τρόπο:

$$
\tilde{A} := \{x \in \mathbb{R}^2 : t_\varepsilon < \|x\|_2 \le r\} = \bigcup_{t \in (t_\varepsilon, r]} \{x \in \mathbb{R}^2 : t \le \|x\| \le r\}. \qquad (5)
$$

Πράγματι, όλα τα σύνολα που εμφανίζονται στο δεξί μέλος περιέχονται στο $\tilde{A}$, άρα το ίδιο θα ισχύει και για την ένωσή τους. Αντίστροφα, αν $x \in \tilde{A}$, τότε $t_\varepsilon < \|x\|_2 \le r$ και για $s = \|x\|_2 - \frac{\|x\| - t_\varepsilon}{2}$, έχουμε ότι $x \in \{y \in \mathbb{R}^2 : s \le \|y\|_2 \le r\}$ με $t_\varepsilon < s$, δηλαδή το x ανήκει στην ένωση του δεξιού μέλους.

Έστω τώρα $(t_n)_n$ φθίνουσα ακολουθία στο $(t_\varepsilon, r]$ η οποία συγκλίνει στο t_ε. Η ακολουθία $(\Delta_{t_n})_n$ είναι αύξουσα με $\cup_n \Delta_{t_n} = \tilde{A}$, άρα $P(\tilde{A}) = \lim_n P(\Delta_{t_n})$. Όμως από τον ορισμό του t_ε, για κάθε $t_n > t_\varepsilon$ θα ισχύει ότι $P(\Delta_{t_n}) < \varepsilon$, άρα $P(\tilde{A}) \leq \varepsilon$, όπως θέλαμε.

Ουσιαστικά, δείξαμε ότι

$$P_{S\sim D^m}\left[P_{x\sim D}[h_S \neq c] > \varepsilon \mid S \cap A \neq \emptyset\right] \leq P_{S\sim D^m}\left[P_{x\sim D}[h_S \neq c] > \varepsilon \mid S \cap \tilde{A} \neq \emptyset\right] = 0.$$

Τελικά,

$$P_{S\sim D^m}[P_{x\sim D}[h_S \neq c] > \varepsilon] \leq 0 + (1-\varepsilon)^m \leq e^{-m}, \tag{6}$$

το οποίο είναι μικρότερο του δ για κάθε $m \geq \frac{1}{\varepsilon} \ln \frac{1}{\delta}$.