The Learning algorithm:
L fits a hypothesis (from H) to the training set by choosing r1 to be the distance between the origin (the center of the circle), and the closest data point in D that is positive (i.e., that the point belongs to the concept c), and choosing r2 to be the distance between the origin, and the furthest data point in D that is positive (i.e., that the point belongs to the concept c).

It's also clear that the time complexity is polynomial in terms of the number of samples: finding a maximum distance and a minimum distance is linear in the number of points.

Correctness:
To show correctness of L, we must show that L is a consistent learner (consistent with the concept it's trying to learn, (c). We show that for all points in the training data, p, and for L(D) = h, h(p) = c(p).
Let ϵ>0,δ>0. Consider c∈C, and denote the inner radius of c by r_{c_1} and the outer radius by r_{c_2}.
By the definition of c, all the interior points (inside the annulus) p_{in} are positive, and all the exterior points, p_{out} are negative. Let p_1^* be the closest point to the origin in c, and let p_2^* be the furthest point to the origin in c, i.e. p_1^* and p_2^* are positive, and denote r_1 as the distance of p_1^* from the origin, and r_2 as the distance of p_2^* from the origin.
Recall that then L(D) = h* is the annulus with inner radius r_1 and outer radius r_2. Therefore, for all interior points of c given in the training data, p_{in}, h*(p_{in}) is positive and for all exterior points of c given in the training data, p_{out}, h*(p_{out}) is negative.

- ⇨ *h* is consistent with c*

Sample Complexity:
Given the desired parameters ϵ and δ, the number of training samples m that is required to guarantee the desired error and confidence, is polynomial in $\frac{1}{\epsilon} > 0$, $\frac{1}{\delta} > 0$.
Let $\epsilon > 0, \delta > 0$. Consider $c \in C$, and denote the inner radius of c by r_{c_1} and the outer radius by r_{c_2}.
Now consider the annulus $c^{(\epsilon)}$ for which the inner radius of $c^{(\epsilon)}$, denoted by $r_{c_1^{(\epsilon)}}$ where $r_{c_1^{(\epsilon)}} > r_{c_1}$, and the outer radius of $c^{(\epsilon)}$, denoted by $r_{c_2^{(\epsilon)}}$, where $r_{c_2^{(\epsilon)}} < r_{c_2}$. This annulus $c^{(\epsilon)}$ creates the annulus A_ϵ, with outer radius of $r_{c_1^{(\epsilon)}}$ and inner radius of r_{c_1} that satisfies $\pi(A_\epsilon) = \frac{\epsilon}{2}$, and creates the annulus B_ϵ, with outer radius of r_{c_2} and inner radius of $r_{c_2^{(\epsilon)}}$ that satisfies $\pi(B_\epsilon) = \frac{\epsilon}{2}$.
More formally:
For $s_1 > r_{c_1}$ define the annulus A_s = $\{(x_1, x_2) | \, r_{c_1} \leq d\big((x_1, x_2), (0,0)\big) \leq s_1\}$, and for $s_2 < r_{c_2}$, define the annulus B_s = $\{(x_1, x_2) | \, s_2 \leq d\big((x_1, x_2), (0,0)\big) \leq r_{c_2}\}$.
Now,
$r_{c_1^{(\epsilon)}}$ = $\inf\{s_1 \mid \pi(A_s) \leq \frac{\epsilon}{2}\}$ and $r_{c_2^{(\epsilon)}}$ = $\inf\{s_2 \mid \pi(B_s) \leq \frac{\epsilon}{2}\}$

Now consider training data $D^{(m)}$, where $|D^m|$ = m.
We have 2 cases:

(1) The “bad” case: No points in the training data falls in the two annuluses that we created.

- ⇨ Happens with probability $2\left(1-\frac{\epsilon}{2}\right)^m \leq 2e^{-\frac{m\epsilon}{2}}$, since for a single point $d \in D^m$, we have $\pi(d \notin A_\epsilon) = 1-\frac{\epsilon}{2}$ and $\pi(d \notin B_\epsilon) = 1-\frac{\epsilon}{2}$ and the points are independent.

(2) The "good" case: There exists a point in the training set that falls in the two annuluses that we created.

- ⇨ Happens with probability $1-2\left(1-\frac{\epsilon}{2}\right)^m$.

Consider the good case, that is there exists a point in at least one of the annuluses.
So the error is as follows:

$err(L(D^m), c) = \pi(h(D^m)\Delta c)$ by error definition from the general set up.

Note that, $\pi(h(D^m)\Delta c) \leq \pi(A_\epsilon \cup B_\epsilon) = \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon$ as A_ϵ and B_ϵ are disjoint. This implies that $err(L(D^m), c) \leq \epsilon$.

Finally, assume we can tune m to be so that in the bad case:

- ⇨ $\pi(D^m \; will \; yield \; the \; bad \; case) < \delta$

Therefore:
$\pi(err(L(D^m), c) \leq \epsilon) \geq \pi(D^m \; will \; yield \; the \; good \; case) = 1 -$
$\pi(D^m \; will \; yield \; the \; bad \; case) > 1-\delta$.

To see what m needs to be in order for $\pi(D^m \; will \; yield \; the \; bad \; case) < \delta$ to hold:

- ⇨ $\pi(D^m \; will \; yield \; the \; bad \; case) \leq 2e^{-\frac{m\epsilon}{2}} < \delta$
- ⇨ $ln2 - \frac{m\epsilon}{2} < ln\delta$
- ⇨ $2ln2 - m\epsilon < 2ln\delta$
- ⇨ $-m\epsilon < 2ln\delta - 2ln2$
- ⇨ $-m\epsilon < 2(ln\delta - ln2) > -2\ln\left(\frac{\delta}{2}\right)$
- ⇨ $m > \frac{2}{\epsilon}\ln\left(\frac{2}{\delta}\right)$

QED

c. 95% confidence → $1-\delta = 0.95 \rightarrow \delta = 0.05$ and 5% error → $\epsilon = 0.05$.

- ⇨ $m > \frac{2}{0.05}\ln\left(\frac{2}{0.05}\right) \approx 147.56 = 148$ samples
- ⇨ $m \geq \frac{1}{0.05}(4\log_2\left(\frac{2}{0.05}\right) + 8 \, * 2 * \log_2\left(\frac{13}{0.05}\right) \approx 2992.91 = 2993$ samples

Even though the hypothesis space may be infinite, the fact that we found a finite VC dimension allows for a more precise estimation of the sample complexity, leading to a lower bound compared to the one assuming an infinite hypothesis space.

Type equation here.