1.1

$$h(x; c_1, c_2, r) = \begin{cases} 1 & \text{if } \sqrt{(c_1 - x_1)^2 + (c_2 - x_2)^2} \le r \\ 0 & \text{otherwise} \end{cases}$$

inside the circle

outside the circle

predictor, input, given point, radius

Let A be the algorithm that returns the smallest circle enclosing all positive examples in the training set.
Why is A NOT an ERM?

Answer

- A is not an ERM due to the large probability of it failing outside of the given training set.
- Since the Algorithm returns the smallest circle enclosing all the positive examples in the set, it may also include all negative examples in the set, or a large portion of them, thus increasing the amount of noise and increasing the possibility to make a mistake, while an algorithm that takes the negative examples into account could have decreased that risk.

For example, if we take two circles, the smaller one contains all positive examples and a single negative example, and the larger one contains all positive examples without the negative, that way the larger circle will have a smaller loss than the smaller circle, thus disproving the above statement.

bonus: The minimal geometric shape is a single dot.

1.2

true error of h: $L_D(h) = P_{x\sim D}[h(x) \neq f(x)]$

empirical error of h over S: $L_S(h) = \frac{1}{m}\sum_{i=1}^{m} 1_{[h(x) \neq f(x)]}$

show that the expected value of $L_S(h)$ over the choice of S equals $L_D(h)$, namely: $E_{S\sim D}(L_S(h)) = L_D(h)$

Proof:

$$\underset{S\sim D}{E}[\underbrace{L_S(h)}_{\text{empirical risk}}] \overset{\text{def}}{=} E\left[\frac{1}{m}\sum_{i=1}^{m} 1_{[h(x)\neq f(x)]}\right]$$

$$\overset{\substack{\text{linearity}\\\text{of expected value}}}{=} \frac{1}{m}\sum_{i=1}^{m} E[1_{[h(x)\neq f(x)]}] \overset{\text{linearity}}{=} \frac{1}{m}\sum_{i=1}^{m} \underset{x\sim D}{P}[h(x)\neq f(x)]$$

$$= \frac{1}{\cancel{m}}\cdot\cancel{m}\, \underbrace{\underset{x\sim D}{P}[h(x)\neq f(x)]}_{\text{truerisk}} = \underbrace{\underset{S\sim D}{P}[h(x)\neq f(x)]}_{\text{true risk}} = L_D(h)$$