Homework 1 Part 1

This is an individual assignment.

You may write your solutions on paper and push a scanned copy to your repository.

Question 1 (5 points)

Let $\mathbf{X} = \mathbb{R}^2$ and consider the set of concepts of the form $c = \{(x, y) : x^2 + y^2 \leq r^2\}$ for some real number r. Show that this class can be (ϵ, δ)-PAC-learned from training data of size $n \geq \frac{1}{\epsilon} \ln \frac{1}{\delta}$.

The VC dimension of the class of circles is 3 because any set of any 3 non-colinear points are shattered by the class of circle.

However, a set of 4 points are not shattered. Example: Take a convex hull of the 4 points in a triangle and keep a negative point at the centroid of the tringle, which is formed by the 3 positive points.

$VC(H)=3$. , i.e. any set of 3 points are shattered by the class of circle.

However any set of 4 points are not shattered.

$$\therefore \quad n = O\left[\frac{1}{\varepsilon}\left[VC(H) \ln\frac{1}{\varepsilon} + \ln\frac{1}{\delta}\right]\right]$$

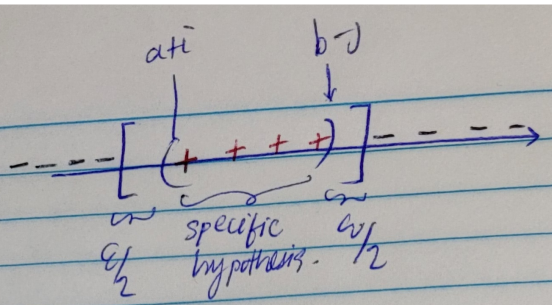
$$= O\left[\frac{1}{\varepsilon}(3)\ln\frac{1}{\varepsilon} + \frac{1}{\varepsilon}\ln\frac{1}{\delta}\right]$$

$$\therefore \quad n \geq \frac{1}{\varepsilon}\ln\frac{1}{\delta}$$

Question 2 (5 points)

Give a PAC-learning algorithm for the concept class $\mathcal{C}$ formed by closed intervals $[a, b]$ with $a, b \in \mathbb{R}$.

Q2.



$$P_{X\sim D^n}[R(R_X)>\varepsilon] \le P_{X\sim D^n}\left[\vee_{i=1}^{2}\{R_X \cap r_i = \emptyset\}\right]$$

$$\le \sum_{i=1}^{2} P_{X\sim D^n}[\{R_X \cap r_i = \emptyset\}]$$

$$\le 2\left(1-\frac{\varepsilon}{2}\right)^n$$

$$\le 2e^{-n\frac{\varepsilon}{2}}$$

$$P_{X\sim D^n}[R(R_X)>\varepsilon] \le \delta$$

$$2e^{-n\frac{\varepsilon}{2}} \le \delta \iff e^{-n\frac{\varepsilon}{2}} \le \frac{\delta}{2}$$

$$\therefore -n\frac{\varepsilon}{2} \le \ln\frac{\delta}{2}$$

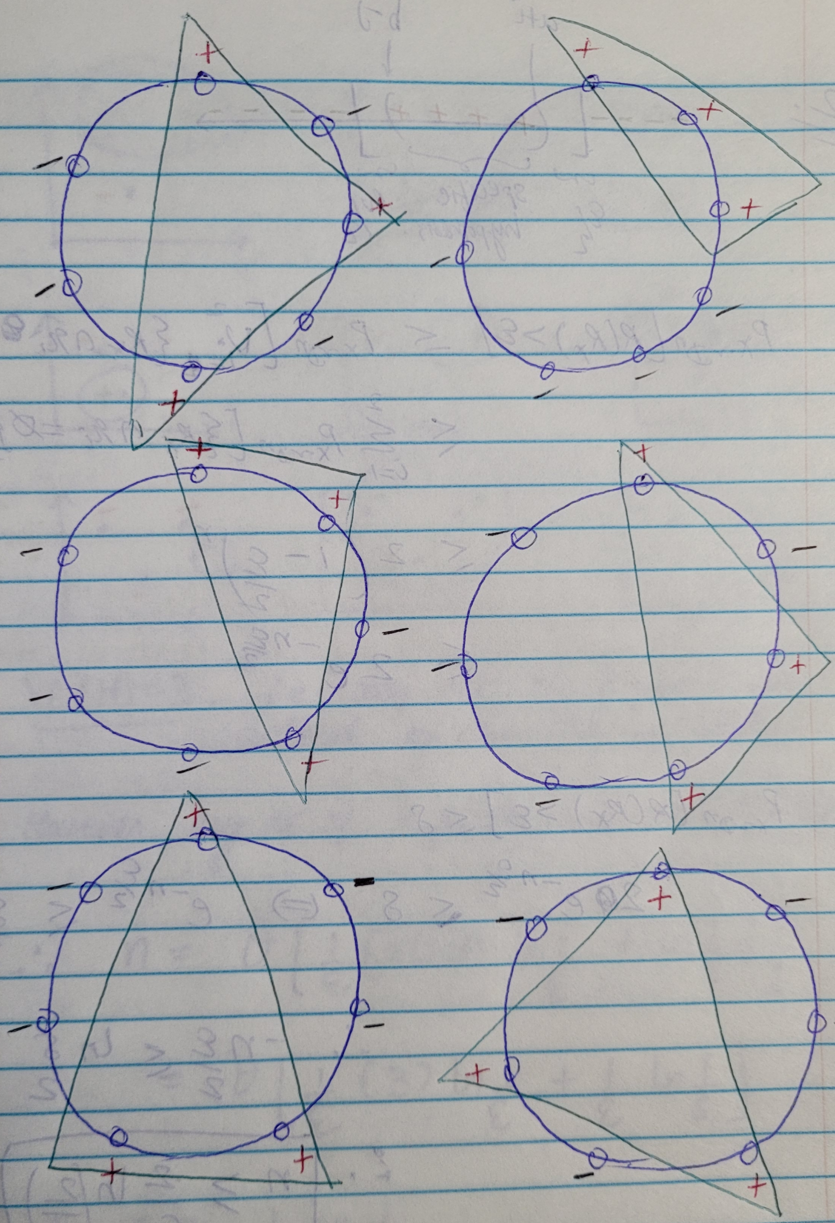
$$\therefore \boxed{n \ge \frac{2}{\varepsilon}\ln\left(\frac{2}{\delta}\right)}$$

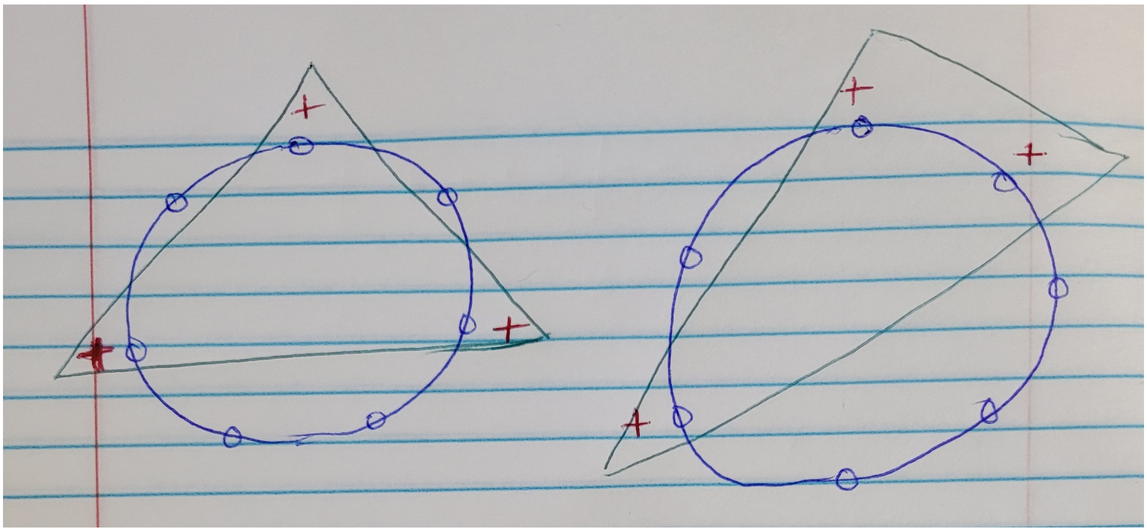
Hence it is PAC learnable.

Question 3 (5 points)

Show that the VC dimension of the triangle hypothesis class is 7 in two dimensions. (Hint: For best separation, it is best to place the seven points equidistant on a circle.)

Q3:





Question 4 (10 points)

Consider the Neyman-Pearson criteria for two univariate Cauchy distributions:

$$p(x|C_i) = \left(\frac{1}{\pi\gamma}\left[1 + \left(\frac{x-\mu_i}{\gamma}\right)^2\right]\right)^{-1} \qquad \forall i = 1,2$$

where the mode is $\mu \in \mathbb{R}$ and width $\gamma \in \mathbb{R}_0^+$ ($\gamma > 0$).

Assume a zero-one error loss, and for simplicity $\mu_2 > \mu_1$, the same width γ, and equal prior.

Answer the following questions:

1. (2 points) Suppose the maximum acceptable error rate for classifying a pattern that is actually in C_1 as if it were in C_2 is ϵ_1. Determine the decision boundary in terms of the variables given.
2. (2 points) For this boundary, what is the error rate for classifying C_2 as C_1?
3. (2 points) What is the overall error rate under zero-one loss?
4. (2 points) Apply your results to the specific case $\gamma = 1$ and $\mu_1 = -1$, $\mu_2 = 1$ and $\epsilon_1 = 0.1$.
5. (2 points) Compare your result to the Bayes error rate (i.e., without the Neyman-Pearson conditions).

Q4.

$$P(x|C_i) = \frac{1}{\pi\gamma}\left[1+\left(\frac{x-\mu_i}{\gamma}\right)^2\right]^{-1} \quad \forall i=1,2$$

error rate of misclassifying $C_1 = \varepsilon_1$

$\therefore$ error for misclassifying $C_2 = 1-\varepsilon_1$

$P(C_1) = P(C_2) = \frac{1}{2}$

zero-one loss function $\Rightarrow \lambda_{11} = \lambda_{22} = 0$

$\lambda_{21} = \lambda_{12} = 1$

$\mu_2 > \mu_1$

(1) $\eta = \frac{\lambda_{12}}{\lambda_{21}} \frac{P(C_2)}{P(C_1)} = \boxed{1}$

(2) Error rate for misclassifying C_2 and calling it $C_1 = 1-\varepsilon_1$

(3) Over all error rate = TPR + FPR

Assuming C_1 is positive class.

$$TPR = \int_{-\infty}^{\gamma} p(x|C_1)\,dx = 1 - \int_{\gamma}^{\infty} p(x|C_1)\,dx$$

$$FPR = \int_{-\infty}^{\gamma} p(x|C_2)\,dx = 1 - \int_{\gamma}^{\infty} p(x|C_2)\,dx$$

$$\therefore \text{ overall error} = 2 - \int_{\gamma}^{\infty} p(x|C_1)\,dx - \int_{\gamma}^{\infty} p(x|C_2)\,dx$$

(4) $\gamma = 1$, $\mu_1 = -1$, $\mu_2 = 1$, $\varepsilon_1 = 0.1$, $\varepsilon_2 = 0.9$

$$\text{Overall error} = 2 - \int_{1}^{\infty} \pi \left[1 + (x+1)^2\right]^{-1} dx$$

$$- \int_{1}^{\infty} \pi \left[1 + (x-1)^2\right]^{-1} dx$$

$$= 2 - \int_{1}^{\infty} \pi \cdot \left[1 + x^2 + 2x + 1\right]^{-1} dx - \int_{1}^{\infty} \pi \left[1 + x^2 - 2x + 1\right]^{-1} dx$$

$$= 2 - \int_{1}^{\infty} \pi \left[x^2 + 2x + 2\right]^{-1} dx - \int_{1}^{\infty} \frac{\pi}{x^2 - 2x + 2}\,dx$$

$$= 2 - \pi \int_{1}^{\infty} \frac{1}{(x+1)^2 + 1^2}\,dx - \pi \int_{1}^{\infty} \frac{1}{(x-1)^2 + 1^2}\,dx$$

$$= 2 - \pi \tan^{-1}(x+1)\Big|_1^{\infty} - \lambda \tan^{-1}(x-1)\Big|_1^{\infty}$$

$$= 2 - \lambda\left(\frac{\pi}{2} - \tan^{-1}(2)\right) - \lambda\left(\frac{\pi}{2} - \tan^{-1}(0)\right)$$

$$= 2 - \frac{\lambda^2}{2} + \lambda \tan^{-1}(2) - \frac{\lambda^2}{2} + 0$$

$$= \underline{\underline{2 - \lambda^2 + \lambda \tan^{-1}(2)}}$$

(5) Bayes error $= \frac{1}{2} = \max(P(C_i|x)$

Question 5 (8 points)

Consider two univariate Gaussian distributions:

$$p(x|C_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \qquad \forall i = 1, 2$$

where the mean is $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}$.

In addition, consider the following decision rule for this binary classification problem: **Decide C_1 if $x > \theta$; otherwise decide C_2.**

Answer the following questions:

1. (2 points) Show that the probability of error for this rule is

$$P(\text{error}) = P(C_1) \int_{-\infty}^{\theta} p(x|C_1)dx + P(C_2) \int_{\theta}^{\infty} p(x|C_2)dx.$$

1. (2 points) By differentiating, show that a necessary condition to minimize $P(\text{error})$ is that θ satisfy

$$P(\theta|C_1)P(C_1) = P(\theta|C_2)P(C_2).$$

1. (2 points) Does this equation define θ uniquely?

2. (2 points) Give an example where a value of θ satisfying the equation actually maximizes the probability of error (i.e. specify values for μ_1, μ_2, σ and the priors).

Q5. $p(x|C_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad \forall i = 1, 2$

Decide C_1 if $x > \theta$ otherwise C_2.

(1) $P(\text{error}) = P(\alpha_1 | C_2) + P(\alpha_2 | C_1)$

$$= \int_{R_1} P(x|C_2) P(C_2) + \int_{R_2} p(x|C_1) P(C_1)$$

$$= P(C_1) \int_{-\infty}^{\theta} p(x|C_1) dx + P(C_2) \int_{\theta}^{\infty} p(x|C_2) dx$$

(2) $\frac{dP(error)}{d\theta} = P(C_1)P(\theta|C_1) - P(C_2)P(\theta|C_2) = 0$

$\therefore P(C_1)P(\theta|C_1) = P(C_2)P(\theta|C_2)$

(3) No, it only gives the likelihood ratio, but the ratio could have different global minima.

(4) $\mu_1 = 0, \quad \mu_2 = 0.5$

$\sigma = 2$ $\quad P(C_1) = P(C_2) = \frac{1}{2}$

$\therefore P(\theta|C_1)P(C_1) = P(\theta|C_2)P(C_2)$

$$\cancel{\frac{1}{\sqrt{2\pi(4)}}} e^{-\frac{(\theta-0)^2}{2(4)}} \cancel{\left(\frac{1}{2}\right)} = \cancel{\frac{1}{\sqrt{2\pi(4)}}} e^{-\frac{(\theta-0.5)^2}{8}} \cancel{\frac{1}{2}}$$

$$\therefore -\frac{\theta^2}{\cancel{8}} = \frac{-(\theta-0.5)^2}{\cancel{8}}$$

$$\therefore -\cancel{\theta^2} = -\cancel{\theta^2} + 2\theta - 0.25$$

$$\therefore 2\theta = 0.25 \iff \theta = 0.125$$

Question 6 (6 points)

Define action α_i as the decision to assign the input to class C_1 and λ_{ik} as the *loss* incurred for taking action α_i when the input actually belongs to C_k.

In a two-class, two-action problem, if the loss is $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 10$, and $\lambda_{21} = 5$, write the optimal decision rule. How does the rule change if we add a third action of reject with $\lambda_{31} = \lambda_{32} = 1$?

Q6. Without rejection.

Decide C_1 if $R(\alpha_1|x) < R(\alpha_2|x)$
otherwise decide C_2.

where $R(\alpha_1|x) = \lambda_{11} P(C_1|x) + \lambda_{12} P(C_2|x)$
$= 0 + \underline{10 P(C_2|x)}$

& $R(\alpha_2|x) = \lambda_{21}(P(C_1|x)) + \lambda_{22} P(C_2|x)$
$= \underline{5 P(C_1|x)} + 0$

With Rejection.

Decide C_1 if $R(\alpha_1|x) < \min_i R(\alpha_i|x)$
where $i \in \{1, 2\}$

Decide C_2 if $R(\alpha_2|x) < \min R(\alpha_i|x)$, $i \in \{1, 2\}$

Reject if $R(\alpha_3|x) < R(\alpha_i|x)$, $i \in \{1, 2\}$

$R(\alpha_3|x) = \lambda_{31} P(C_1|x) + \lambda_{32} P(C_2|x)$

Question 7 (6 points)

An *association rule* is an implication of the form $X \to Y$ where X is the *antecedent* and Y is the *consequent* of the rule. One example of association rule is in *basket analysis* where we want to find the dependency between two items X and Y. In learning association rules, we can measure:

- *Support* of the association rule $X \to Y$:

$$\text{Support}(X, Y) \equiv P(X, Y) = \frac{\#\text{ customers who bought X and Y}}{\#\text{ customers}}$$

- *Confidence* of the association rule $X \to Y$:

$$\text{Confidence}(X, Y) \equiv P(Y|X) = \frac{P(X, Y)}{P(X)} = \frac{\#\text{ customers who bought X and Y}}{\#\text{ customers who bought X}}$$

Answer the following questions:

1. (3 points) Given the following data of transactions at a shop, calculate the support and confidence values of $\text{milk} \to \text{bananas}$, $\text{bananas} \to \text{milk}$, $\text{milk} \to \text{chocolate}$, and $\text{chocolate} \to \text{milk}$.

Transaction	**Items in Basket**
1	milk, bananas, chocolate
2	milk, chocolate
3	milk, bananas
4	chocolate
5	chocolate
6	milk, chocolate

1. (3 points) Generalize the confidence and support formulas for basket analysis to calculate k-dependencies, namely $P(Y|X_1, \dots, X_k)$.

Q7.

	milk	banana	chocolate
1	✓	✓	✓
2	✓		✓
3	✓	✓	
4			✓
5			✓
6	✓		✓

(1) Support (milk $\longrightarrow$ bananas) $= \frac{2}{6} = \frac{1}{3}$

confidence $= P($banana$|$milk$) = \frac{2}{4} = \frac{1}{2}$

(2) banana $\longrightarrow$ milk

support $= \frac{1}{2}$

confidence $= \frac{2}{2} = 1$

(3) milk $\longrightarrow$ chocolate

support $= \frac{3}{6} = \frac{1}{2}$

Confidence $= \frac{3}{4}$

(4) chocolate $\rightarrow$ milk

support $= \frac{1}{2}$

confidence $= \frac{3}{5}$

On-Time (5 points)

Submit your assignment before the deadline.

Submit Your Solution

Confirm that you've successfully completed the assignment.

Along with the Notebook, include a PDF of the notebook with your solutions.

`add` and `commit` the final version of your work, and `push` your code to your GitHub repository.

Submit the URL of your GitHub Repository as your assignment submission on Canvas.
