מגיש: לוין אבר

ת.ז - 322357548

Theoretical Assignment 1

Intro to Machine Learning, TAU

November 18, 2024

Question 1

Imagine you are classifying flowers as either *Iris-setosa* or *Iris-versicolor* based on two features: petal length and petal width. You have the following information:

Class	Mean Petal Length	Std Dev Petal Length	Mean Petal Width	Std Dev Petal Width
Iris-setosa	1.5	0.2	0.2	0.1
Iris-versicolor	4.7	0.3	1.4	0.2

Table 1: Summary statistics for petal length and width by flower class.

Assume that the features are independent and normally distributed.

1. Calculate the likelihoods $P(\text{petal length} = 1.6 \mid \text{Iris-setosa})$ and $P(\text{petal width} = 0.25 \mid \text{Iris-setosa})$ using the probability density function of a normal distribution, and plot them.

2. Based on these likelihoods, design a classifier that receives the petal width and petal length and predicts the the class.

3. Using the classifier from the previous section, determine which class (Iris-setosa or Iris-versicolor) is more likely for a flower with a petal length of 1.6 and a petal width of 0.25. Assume equal priors for both classes.

Question 2

A medical test is used to classify patients into two categories: *Disease* or *No Disease*. However, due to the high costs associated with misclassification, there is also an option to reject a classification if confidence is too low. You are given the following information:

- $P(\text{Disease}) = 0.2$

- $P(\text{No Disease}) = 0.8$
- The test has a sensitivity (true positive rate) of 90%, meaning $P(\text{Positive Test} \mid \text{Disease}) = 0.9$.
- The test has a specificity (true negative rate) of 85%, meaning $P(\text{Negative Test} \mid \text{No Disease}) = 0.85$.

A patient receives a positive test result.
Cost Structure:

- The cost of a false positive (FP) error (i.e., diagnosing a healthy patient as having the disease) is \$1,000.
- The cost of a false negative (FN) error (i.e., failing to diagnose a patient with the disease) is \$5,000.
- The cost of rejecting the classification for further testing is \$2,000.

Tasks:

1. Calculate the probability that this patient actually has the disease given a positive test result, $P(\text{Disease} \mid \text{Positive Test})$.
2. Based on this probability and the cost structure, decide whether to classify the patient as *Disease*, *No Disease*, or *Reject*. Choose the option with the lowest expected cost. Use a confidence threshold of 70% for deciding whether to classify the patient as having the disease. If the probability of disease is below 70% but above 30%, classify as *Reject*; otherwise, classify as *No Disease*.

Question 3

Given a real number $R \geq 0$, define the hypothesis $h_R : \mathbb{R}^d \to \{0, 1\}$ as follows:

$$h_R(x) = \begin{cases} 1 & \text{if } \|x\|_2 \leq R, \\ 0 & \text{otherwise.} \end{cases}$$

Consider the hypothesis class $H_{\text{ball}} = \{h_R \mid R \geq 0\}$. Prove directly (without using the Fundamental Theorem of PAC Learning) that H_{ball} is PAC learnable in the realizable case. Assume for simplicity that the marginal distribution of X is continuous. How does the sample complexity depend on the dimension d? Explain.

Question 4

Given a polynomial $P : \mathbb{R} \to \mathbb{R}$, define the hypothesis $h_P : \mathbb{R}^2 \to \{0, 1\}$ as follows:

$$h_P(x_1, x_2) = \begin{cases} 1 & \text{if } P(x_1) \geq x_2, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the VC-dimension of $H_{\text{poly}} = \{h_P \mid P \text{ is a polynomial}\}$. You can use the fact that given n distinct values $x_1, \ldots, x_n \in \mathbb{R}$ and $z_1, \ldots, z_n \in \mathbb{R}$, there exists a polynomial P of degree $n-1$ such that $P(x_i) = z_i$ for every $1 \leq i \leq n$.

Question 1

Imagine you are classifying flowers as either *Iris-setosa* or *Iris-versicolor* based on two features: petal length and petal width. You have the following information:

Class	Mean Petal Length	Std Dev Petal Length	Mean Petal Width	Std Dev Petal Width
Iris-setosa	1.5	0.2	0.2	0.1
Iris-versicolor	4.7	0.3	1.4	0.2

Table 1: Summary statistics for petal length and width by flower class.

Assume that the features are independent and normally distributed.

1. Calculate the likelihoods $P(\text{petal length} = 1.6 \mid \text{Iris-setosa})$ and $P(\text{petal width} = 0.25 \mid \text{Iris-setosa})$ using the probability density function of a normal distribution, and plot them.

2. Based on these likelihoods, design a classifier that receives the petal width and petal length and predicts the the class.

3. Using the classifier from the previous section, determine which class (Iris-setosa or Iris-versicolor) is more likely for a flower with a petal length of 1.6 and a petal width of 0.25. Assume equal priors for both classes.

1)

likelihood - ה [illegible]

$$\left(\begin{array}{l} p(\text{petal lenth} = 1.6 \mid \text{Iris-Setosa}) \\ p(\text{petal width} = 2.5 \mid \text{Iris-Setosa}) \end{array} \right)$$

[illegible]

[illegible]

$$\boxed{f(x|\text{class}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$$

Iris-Setosa length $\sim N(1.5, 0.04)$

Iris-Setosa width $\sim N(0.2, 0.01)$

Iris-versicolor length $\sim N(4.7, 0.09)$

Iris-versicolor width $\sim N(1.4, 0.04)$

$$\underline{x = 1.6}$$

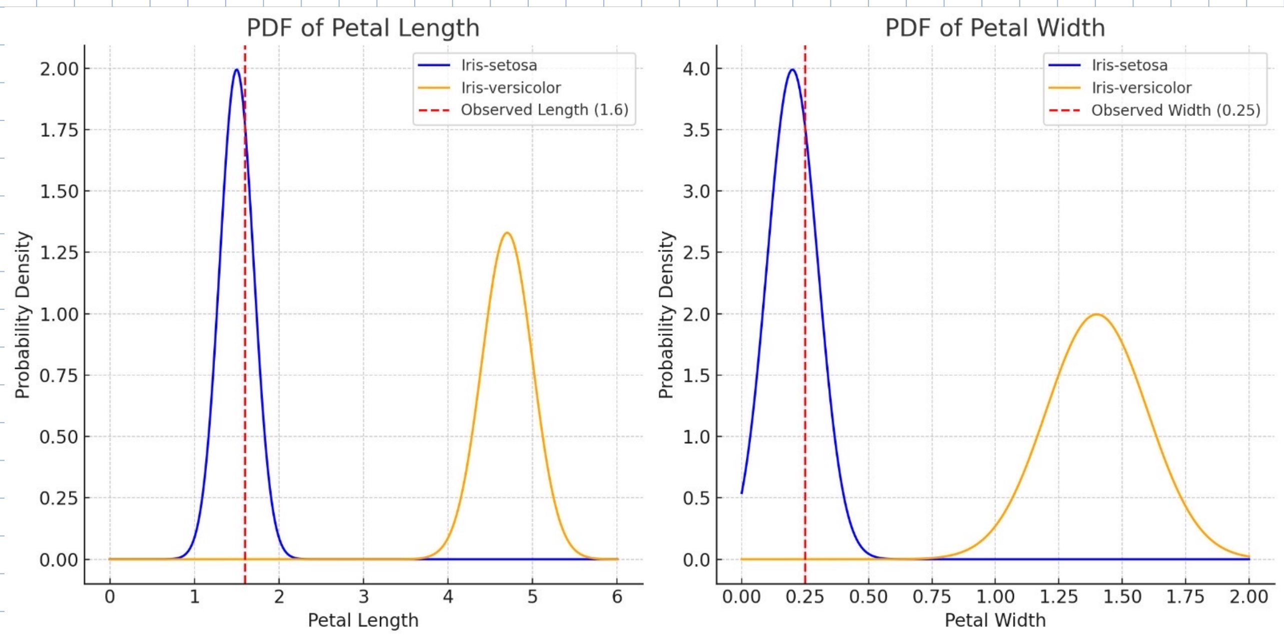
$$f(\text{petal lenth} = 1.6 \mid \text{Iris-setosa},\ 0.2^2) = \frac{1}{\sqrt{2\pi\, 0.2^2}} e^{-\frac{(1.6-1.5)^2}{2 \cdot 0.2^2}}$$

$$= 1.76$$

$$f(\text{petal width} = .25 \mid \text{Iris-setosa},\ 0.1^2) = \frac{1}{\sqrt{2\pi\, 0.1^2}} e^{-\frac{(0.25-0.2)^2}{2 \cdot 0.1^2}}$$

$$= 3.52$$

(2)



$p \sim N(\mu, \sigma^2)$

(3)

$$\bar{\mu}_1 = \begin{pmatrix} 1.5 \\ 0.2 \end{pmatrix} \longrightarrow \Sigma_1 = \begin{pmatrix} 0.04 & 0 \\ 0 & 0.01 \end{pmatrix}$$

$$\bar{\mu}_2 = \begin{pmatrix} 4.7 \\ 1.4 \end{pmatrix} \longrightarrow \Sigma_2 = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.04 \end{pmatrix}$$

$$\sigma_1^2 = \begin{pmatrix} 0.04 \\ 0.01 \end{pmatrix}$$

$$\sigma_2^2 = \begin{pmatrix} 0.09 \\ 0.04 \end{pmatrix}$$

$$p(\bar{x} \mid a_i) = \frac{1}{\sqrt{2\pi}^{2} \sqrt{|\Sigma_i|}} e^{\left(\frac{-(\underline{x}-\mu_i)^T \Sigma_i^{-1} \cdot (\underline{x}-\mu_i)}{2}\right)}$$

נבנה את ה-Classifier לפי בייס

$$C = \underset{i=\{1,2\}}{\arg\max} \left(P(\bar{x}|a_i)\cdot P(a_i)\right)$$

8

length=1.6 width=0.25

נבדוק את הסיכוי המקסימלי

Iris-setosa length $\sim N(1.5, 0.04)$

Iris-setosa width $\sim N(0.2, 0.01)$

Iris-versicolor length $\sim N(4.7, 0.09)$

Iris-versicolor width $\sim N(1.4, 0.04)$

Iris-setosa - a_1

Iris-versicolor - a_2

$$\underline{x} = \begin{pmatrix} 1.6 \\ 0.25 \end{pmatrix}$$

עבור a_1

$$P(\bar{x}|a_i) = \frac{1}{\sqrt{2\pi}^2\sqrt{|\Sigma_i|}} e^{\left(-\frac{(\underline{x}-\mu_i)^T \Sigma_i^{-1} (\underline{x}-\mu_i)}{2}\right)}$$

$$\sqrt{|\Sigma_1|} = \sqrt{4\cdot 10^{-4}} = 0.02$$

$$\Sigma_1^{-1} = \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix}$$

$$(\underline{x} - \underline{\mu}_1) = \begin{pmatrix} 1.6 \\ 0.25 \end{pmatrix} - \begin{pmatrix} 1.5 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix}$$

$$(\underline{x} - \underline{\mu}_1)^T = \begin{pmatrix} 0.1 & 0.05 \end{pmatrix}$$

$$(\underline{x} - \underline{\mu}_1)^T \cdot \Sigma_1^{-1} \cdot (x - \mu_1) = \begin{pmatrix} 0.1 & 0.05 \end{pmatrix} \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix} \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix} = 0.5$$

$$p(x|a_1) = \frac{1}{2\pi \cdot 0.02} \cdot e^{-\frac{0.5}{2}} \approx 8.74$$

עבור a_2 -

$$\sqrt{|\Sigma_2|} = \sqrt{36\cdot 10^{-4}} = 0.06$$

$$\Sigma_2^{-1} = \begin{pmatrix} 4.1 & 0 \\ 0 & 25 \end{pmatrix}$$

$$(\underline{x} - \underline{\mu_2}) = \begin{pmatrix} 1.6 \\ 0.25 \end{pmatrix} - \begin{pmatrix} 4.7 \\ 1.4 \end{pmatrix} = -\begin{pmatrix} 3.1 \\ 1.15 \end{pmatrix}$$

$$(\underline{x} - \underline{\mu_2})^T = -\begin{pmatrix} 3.1 & 1.15 \end{pmatrix}$$

$$(\underline{x} - \underline{\mu_2})^T \cdot \Sigma_2^{-1} \cdot (x - \mu_2) = -\begin{pmatrix} 3.1 & 1.15 \end{pmatrix} \begin{pmatrix} 4.1 & 0 \\ 0 & 25 \end{pmatrix} \cdot \left(-\begin{pmatrix} 3.1 \\ 1.15 \end{pmatrix} \right) = 139.733$$

$$P(\bar{x} \mid a_2) = \frac{1}{\sqrt{2}\,\pi \cdot 0.06} \cdot e^{-\frac{139.733}{2}} \rightsquigarrow 0$$

$$P(\bar{x} \mid a_1) >> P(\bar{x} \mid a_2)$$

$$P(a_2) = P(a_1) = \frac{1}{2}$$

Iris-setora ✓

Question 2

A medical test is used to classify patients into two categories: *Disease* or *No Disease*. However, due to the high costs associated with misclassification, there is also an option to reject a classification if confidence is too low. You are given the following information:

- $P(\text{Disease}) = 0.2$

- $P(\text{No Disease}) = 0.8$
- The test has a sensitivity (true positive rate) of 90%, meaning $P(\text{Positive Test} \mid \text{Disease}) = 0.9$.
- The test has a specificity (true negative rate) of 85%, meaning $P(\text{Negative Test} \mid \text{No Disease}) = 0.85$.

A patient receives a positive test result.
Cost Structure:

- The cost of a false positive (FP) error (i.e., diagnosing a healthy patient as having the disease) is \$1,000.
- The cost of a false negative (FN) error (i.e., failing to diagnose a patient with the disease) is \$5,000.
- The cost of rejecting the classification for further testing is \$2,000.

Tasks:

1. Calculate the probability that this patient actually has the disease given a positive test result, $P(\text{Disease} \mid \text{Positive Test})$.
2. Based on this probability and the cost structure, decide whether to classify the patient as *Disease*, *No Disease*, or *Reject*. Choose the option with the lowest expected cost. Use a confidence threshold of 70% for deciding whether to classify the patient as having the disease. If the probability of disease is below 70% but above 30%, classify as *Reject*; otherwise, classify as *No Disease*.

1. $P(\text{Disease}|\text{Positive Test}) = \dfrac{P(\text{Positive Test}|\text{Disease}) \cdot P(\text{Disease})}{P(\text{Positive Test})}$ $=0.2$

מכס'

נשתמש ב הסתברות השלמה:

$P(\text{Positive Test}) = P(\text{Positive Test}|\text{Disease})\,P(\text{Disease}) +$
$P(\text{Positive Test}|\text{No Disease})\,P(\text{No Disease})$

?

$P(\text{positive Test}|\text{No Disease}) + P(\text{Negative Test}|\text{No Disease}) = 1$

$\Downarrow$

$P(\text{Negative Test}|\text{No Disease}) = 1 - P(\text{positive Test}|\text{No Disease})$

$1 - 0.85 = 0.15$

ספק

$P(\text{Positive Test}) = P(\text{Positive Test}|\text{Disease})\,P(\text{Disease}) +$
$P(\text{Positive Test}|\text{No Disease})\,P(\text{No Disease})$

$= 0.9 \cdot 0.2 + 0.15 \cdot 0.8 = 0.3$

סוף.

$$P(\text{Disease}|\text{Positive Test}) = \frac{\overset{0.9}{P(\text{Positive Test}|\text{Disease})} \cdot P(\text{Disease}) = 0.2}{\underset{0.3}{P(\text{Positive Test})}}$$

$$= \frac{0.9 \cdot 0.2}{0.3} = 0.6$$

2. threshold 70%

$p < 0.3$ – No disease

$0.3 < p < 0.7$ – Disease

$p > 0.7$ – reject

ראינו שהמודל טוען חולה בסיכוי של 0.6

אז זה בדיקה חיובית – $0.3 < p < 0.7$

לכן הבחירה היא reject

וקבע הפסד מינימלי.

$l(a,c)$ cost – עלות

	y_1 disease	y_2 no disease
a_0 reject	2000	2000
a_1 positive	0	1000
a_2 negative	5000	0

בעבור חיזוי positive-

$$cost(a_1, y_i) = 0 + l(a_1, y_2)\, P(y=y_2|a_1) =$$

$$6000(1-P(y=y_1|a_1)) = 6000 \cdot 0.07 = 420$$

בעבור חיזוי reject-

$$cost(a_0, y_i) = 2000$$

$$cost(a_2, y_i) = 0 + 5000 \cdot P(y=y_1|a_2) =$$

$$0 + 5000 \cdot P(Disease|Negative) = 5000 \frac{P(Disease, negative)}{P(Negative)} =$$

$$5000 \cdot \frac{0.02}{0.7} = 142.857$$

העלות של ה reject הכי גבוהה

ולכן לא כדאי להשתמש במקורי ה- threshold

Question 4

Given a polynomial $P : \mathbb{R} \to \mathbb{R}$, define the hypothesis $h_P : \mathbb{R}^2 \to \{0, 1\}$ as follows:

$$h_P(x_1, x_2) = \begin{cases} 1 & \text{if } P(x_1) \geq x_2, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the VC-dimension of $H_{\text{poly}} = \{h_P \mid P \text{ is a polynomial}\}$. You can use the fact that given n distinct values $x_1, \dots, x_n \in \mathbb{R}$ and $z_1, \dots, z_n \in \mathbb{R}$, there exists a polynomial P of degree $n - 1$ such that $P(x_i) = z_i$ for every $1 \leq i \leq n$.

$$VC\text{-}dim(H_{poly}) = \infty$$

שים לב שלכל נקודה (x_1,x_2) כך ש- $P(x_1) \geq x_2$

בהינתן P תמיד קיים (x_1,y_1) כך ש- $y_1 > x_2$ $(y_1 = P(x_1))$

(mapping)
אם כן לכל $n \in N$ נראה שלכל σ -ו הצביעה

על $n+1$ נקודות $\{x_1, \dots, x_n, x_{n+1}\}$

לכל נקודה x_i אז נראה שקיים תיוג 1 ש

נראה ושקיים $y_i > x_{i_2}$

ואחרת- $y_i < x_{i_2}$

לכן לכל n, H_{poly} מכיל את 2^n

האפשרויות של $\{x_1, \dots, x_n\}$

נגדיר (x_1^i, x_2^i) $1 \leq i \leq n$

ועבור n נקודות (x_1^i, x_2^i), $1 \leq i \leq n$

ולכל סימון $\{y_i\}_{1 \leq i \leq n}$ נגדיר:

$$W_i = \begin{cases} x_2^i + \varepsilon & \text{if } y_i = 1 \\ x_2^i - \varepsilon & \text{if } y_i = 0 \end{cases}$$

כ(ג)אשר פס Σ מסי קון β שלהו

משפט מסקנה הלמה (עקרון, G m נק' שונות

אז נמצא קיים פולינום בדרגה $m-1$ שעובר דרכן)

ונדרוש שהפולינום יקיים $p(m_i)=w_i$ לכל i שהפולינום

מקיים את ההפרדה שדרשנו הרצויה, נשים לב כי נוכל

להחליף את הקבוע הגדול לכל G n עוד',

פולינומי G תחת G עבור ה"ל אפשרות - שיוכר

לכל n נק' פולינום בדרגה $n-1$ שמספיק רב.

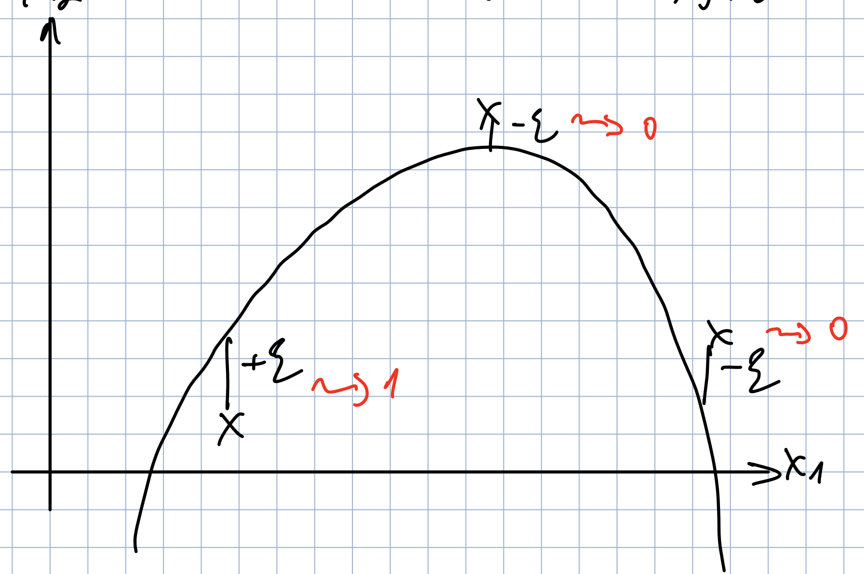
לנפץ n נקודות.

VC-dim מוגדר כ-n הגדול ביותר שניתן לנפץ בה ההפרדה.

במקרה שלנו מגיעים ל-∞. (עקרון G n קיים פולינום

שמנפץ את הנקודות)

פולינום מדרגה 2



Question 3

Given a real number $R \geq 0$, define the hypothesis $h_R : \mathbb{R}^d \to \{0, 1\}$ as follows:

$$h_R(x) = \begin{cases} 1 & \text{if } \|x\|_2 \leq R, \\ 0 & \text{otherwise.} \end{cases}$$

2

Consider the hypothesis class $H_{\text{ball}} = \{h_R \mid R \geq 0\}$. Prove directly (without using the Fundamental Theorem of PAC Learning) that H_{ball} is PAC learnable in the realizable case. Assume for simplicity that the marginal distribution of X is continuous. How does the sample complexity depend on the dimension d? Explain.

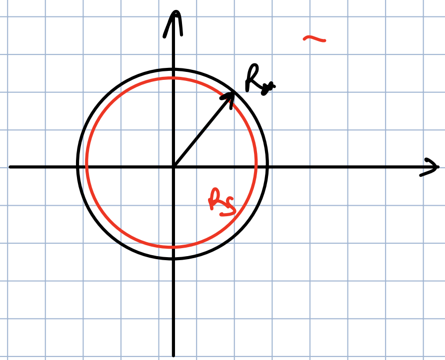
realizable $\Rightarrow$ $h^* \in H$ קיים

$y = h^*(x)$ [illegible]

$h^*(x) = \begin{cases} 1 & \|x\| \leq R^* \\ 0 & \text{o.w} \end{cases}$ [illegible] R^* נגדיר

$R_S = \max_{1 \leq i \leq n} \|x_i\|_2$, $S = \{x_i\}_{i=1}^n$ קבוצת [illegible] נגדיר

(realizable כי) $R_S \leq R^*$



עבור כל x_i כך ש- $\|x_i\| \le R_S$ מתקיים $\hat{h}(x_i) = h^*(x_i)$

ולכן אין טעות .

השגיאה היחידה שיכולה — היא עבור x_j

כך ש- $R_S < \|x_j\| \le R_*$

$$P[\text{error}(\hat{h})] = P[R_S < \|x_j\| \le R_*] = P[\|x_j\| \le R_*] - P[\|x_j\| \le R_S]$$

0
//

עבור $P[\|x_j\| \le R_*] < \varepsilon$, ולכן $P[\text{error}(\hat{h})] < \varepsilon$

$$P[\text{error}(\hat{h}) > \varepsilon] \le \prod_{i=1}^{n} (1-\varepsilon) = (1-\varepsilon)^n \le e^{-\varepsilon n} < \delta$$

$$-\varepsilon n < \ln(\delta)$$

$$\varepsilon n > -\ln(\delta)$$

$$\varepsilon n > \ln\left(\frac{1}{\delta}\right)$$

$$n > \frac{1}{\varepsilon} \ln\left(\frac{1}{\delta}\right)$$

$$n > \ln\left(\frac{1}{\delta}\right) \cdot \frac{1}{\varepsilon}$$

$$\Downarrow$$

לכן H_{ball} למידה