HOMEWORK 6: PROBABILISTIC LEARNING, CNNS, LEARNING THEORY

10-301/10-601 Introduction to Machine Learning (Summer 2022)
https://www.cs.cmu.edu/~hchai2/courses/10601/
OUT: Wednesday, July 6
DUE: Wednesday, July 13 at 1:00 PM
TAs: Sana, Brendon, Ayush, Boyang (Jack), Chu

Homework 6 covers topics on MLE/MAP, Naive Bayes, logistic regression, CNNs, and learning theory. The homework includes multiple choice, True/False, and short answer questions.

START HERE: Instructions

- **Collaboration Policy**: Please read the collaboration policy here: https://www.cs.cmu.edu/~hchai2/courses/10601/#Syllabus
- **Late Submission Policy:** See the late submission policy here: https://www.cs.cmu.edu/~hchai2/courses/10601/#Syllabus
- **Submitting your work:** You will use Gradescope to submit answers to all questions. Please follow instructions at the end of this PDF to correctly submit all your code to Gradescope.
  - **Written:** For written problems such as short answer, multiple choice, derivations, proofs, or plots, please use the provided template. Submissions can be handwritten onto the template, but should be labeled and clearly legible. If your writing is not legible, you will not be awarded marks. Alternatively, submissions can be written in LaTeX. Each derivation/proof should be completed in the boxes provided. You are responsible for ensuring that your submission contains exactly the same number of pages and the same alignment as our PDF template. If you do not follow the template, your assignment may not be graded correctly by our AI assisted grader.

Instructions for Specific Problem Types

For "Select One" questions, please fill in the appropriate bubble completely:

Select One: Who taught this course?

- ● Henry Chai
- ○ Marie Curie
- ○ Noam Chomsky

If you need to change your answer, you may cross out the previous answer and bubble in the new answer:

Select One: Who taught this course?

- ● Henry Chai
- ○ Marie Curie
- ⊗ Noam Chomsky

For "Select all that apply" questions, please fill in all appropriate squares completely:

Select all that apply: Which are scientists?

- ■ Stephen Hawking
- ■ Albert Einstein
- ■ Isaac Newton
- □ I don't know

Again, if you need to change your answer, you may cross out the previous answer(s) and bubble in the new answer(s):

Select all that apply: Which are scientists?

- ■ Stephen Hawking
- ■ Albert Einstein
- ■ Isaac Newton
- ⊠ I don't know

For questions where you must fill in a blank, please make sure your final answer is fully included in the given space. You may cross out answers or parts of answers, but the final answer must still be within the given space.

Fill in the blank: What is the course number?

10-601

10-6301

Written Questions (80 points)

1 Latex Bonus Point (1 points)

1. (1 point) **Select one:** Did you use Latex for the entire written portion of this homework?

   ◯ Yes

   ◯ No

2 MLE/MAP (18 points)

1. (2 points) **True or False:** Suppose you place a Beta prior over a Bernoulli distribution, and attempt to learn the parameter θ of the Bernoulli distribution from data. Further suppose an adversary chooses "bad", but finite hyperparameters for your Beta prior in order to confuse your learning algorithm. As the number of training examples grows to infinity, the MAP estimate of θ can still converge to the MLE estimate of θ.

   ◯ True

   ◯ False

2. (2 points) **Select one:** Let Γ be a random variable with the following probability density function (pdf):

$$f(\gamma) = \begin{cases} 2\gamma & \text{if } 0 \leq \gamma \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   Suppose another random variable Y, which is conditional on Γ, follows an exponential distribution with $\lambda = 3\gamma$. Recall that the exponential distribution with parameter λ has the following pdf:

$$f_{exp}(y) = \begin{cases} \lambda e^{-\lambda y} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   What is the MAP estimate of γ given $Y = \frac{2}{3}$ is observed?

   ◯ 0

   ◯ 1/3

   ◯ 1

   ◯ 2

3. In HW3, you derived the closed form solution for linear regression. Now, we are coming back to linear regression, viewing it as a statistical model, and deriving the MLE and MAP estimate of the parameters in the following questions.

   As a reminder, in MLE, we have

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, p(\mathcal{D}|\theta)$$

   For MAP, we have

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\, p(\theta|\mathcal{D})$$

(a) (2 points) **Select one:** Assume we have data $\mathcal{D} = \{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^{N}$, where $\mathbf{x}^{(i)} = (x_1^{(i)}, \cdots, x_M^{(i)})$. Each $y^{(i)}$ is generated given $\mathbf{x}^{(i)}$ with additive noise $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$ – that is, $y^{(i)} = \mathbf{w}^T\mathbf{x}^{(i)} + \epsilon^{(i)}$, where $\mathbf{w}$ is the parameter vector. Given this assumption, what is the distribution of $y^{(i)}$?

◯ $y^{(i)} \sim N(\mathbf{w}^T\mathbf{x}^{(i)}, \sigma^2)$

◯ $y^{(i)} \sim N(0, \sigma^2)$

◯ $y^{(i)} \sim \text{Uniform}(\mathbf{w}^T\mathbf{x}^{(i)} - \sigma, \mathbf{w}^T\mathbf{x}^{(i)} + \sigma)$

◯ None of the above

(b) (2 points) **Select one:** We'll first derive the maximum likelihood estimate of the parameters of the linear regression model. Which expression below is the correct conditional log likelihood $\ell(\mathbf{w})$ using the given data?

◯ $\sum_{i=1}^{N}[-\log(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2]$

◯ $\sum_{i=1}^{N}[\log(\sqrt{2\pi\sigma^2}) + \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2]$

◯ $\sum_{i=1}^{N}[-\log(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})]$

◯ $-\log(\sqrt{2\pi\sigma^2}) + \sum_{i=1}^{N}[-\frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2]$

(c) (3 points) **Select all that apply:** Then, the MLE of the parameters is just $\text{argmax}_{\mathbf{w}}\, \ell(\mathbf{w})$. Among the following expressions, select **ALL** that can yield the correct MLE.

□ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N}[-\log(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})]$

□ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N}[-\log(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2]$

□ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N}[-\frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2]$

□ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N}[-\frac{1}{2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})]$

□ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N}[-\frac{1}{2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2]$

(d) (3 points) **Select all that apply:** Now, we'll move on to deriving the MAP estimate of the parameters of the linear regression model. Which of the expressions below are correct expressions for the MAP estimate $\mathbf{w}_{MAP}$? (recall that $\mathcal{D}$ refers to the data, and $\mathbf{w}$ to the regression parameters (weights)).

□ $\mathbf{w}_{MAP} = \arg\max_{\mathbf{w}} p(\mathcal{D}, \mathbf{w})$

□ $\mathbf{w}_{MAP} = \arg\max_{\mathbf{w}} \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}$

□ $\mathbf{w}_{MAP} = \arg\max_{\mathbf{w}} \frac{p(\mathcal{D},\mathbf{w})}{p(\mathbf{w})}$

□ $\mathbf{w}_{MAP} = \arg\max_{\mathbf{w}} p(\mathcal{D}|\mathbf{w})p(\mathbf{w})$

□ $\mathbf{w}_{MAP} = \arg\max_{\mathbf{w}} p(\mathbf{w}|\mathcal{D})$

4. (2 points) **Select one:** A MAP estimator with a Gaussian prior $\mathcal{N}(0, \sigma^2)$ you trained gives significantly higher test error than train error. What could be a possible approach to fixing this?

   ◯ Increase variance σ^2

   ◯ Decrease variance σ^2

   ◯ Try finding the MLE instead

   ◯ None of the above

5. (2 points) **Select one:** MAP estimation with which prior is equivalent to L1 regularization? For this questions, you must also provide mathematical justification below.

   For reference:

   - The pdf of a Uniform distribution over $[a, b]$ is $f(x) = \frac{1}{b-a}$ if $x \in [a, b]$ and 0 otherwise.
   - The pdf of an Exponential distribution with rate parameter a is $f(x) = a\exp(-ax)$ for $x > 0$.
   - The pdf of a Laplace distribution with location parameter a and scale parameter b is $f(x) = \frac{1}{2b}\exp\left(\frac{-|x-a|}{b}\right)$ for all $x \in \mathbb{R}$.

   ◯ Uniform distribution over $[-1, 1]$

   ◯ Uniform distribution over $[-\mathbf{w}^T\mathbf{x}^{(i)}, \mathbf{w}^T\mathbf{x}^{(i)}]$

   ◯ Exponential distribution with rate parameter $a = \frac{1}{2}$

   ◯ Exponential distribution with rate parameter $a = \mathbf{w}^T\mathbf{x}^{(i)}$

   ◯ Laplace prior with location parameter $a = 0$

   ◯ Laplace prior with location parameter $a = \mathbf{w}^T\mathbf{x}^{(i)}$

Justification

3 Naïve Bayes (18 points)

1. (2 points) Suppose that 0.3% of all people have cancer. If someone is tested for cancer, the outcome of the test can either be positive (cancer) or negative (no cancer). The test is not perfect - among people who have cancer, the test comes back positive 97% of the time. Among people who don't have cancer, the test comes back positive 4% of the time.

   Assuming that test results are independent of each other, given the true state (cancer or no cancer), what is the probability of a test subject having cancer, given that the subject's test result is positive?

   Round your answer to 4 decimal places after the decimal point.

   Answer

2. (3 points) **Select all that apply:** In a Naïve Bayes problem, suppose we are trying to compute $P(Y \mid X_1, X_2, X_3, X_4)$. Furthermore, assume X_2 and X_3 are identical (i.e., X_3 is just a copy of X_2), and that X_2 is not independent of Y. Which of the following are true in this case?

   - □ Naïve Bayes will learn identical parameter values for $P(X_2|Y)$ and $P(X_3|Y)$.
   - □ Naïve Bayes will output probabilities $P(Y|X_1, X_2, X_3, X_4)$ that are closer to 0 and 1 than they would be if we removed the feature corresponding to X_3.
   - □ There is not enough information to determine the change in the output $P(Y|X_1, X_2, X_3, X_4)$.
   - □ None of the above

3. The following dataset describes several features of a person and then whether or not they are a fan of the K-pop boy band, BTS.

Hair Color	Height	BTS Fan?
Blonde	Short	N
Brown	Short	N
Brown	Short	Y
Brown	Medium	Y
Blonde	Medium	N
Black	Medium	Y
Black	Tall	Y
Brown	Tall	Y

   Zack, a new friend that you've met, has brown hair and medium height. We would like to determine whether they are a BTS fan, using the Naïve Bayes assumption to estimate the following probabilities.

(a) (2 points) What is the probability that someone has brown hair, medium height, and is a BTS fan?

Answer

(b) (2 points) What is the probability that someone has brown hair, medium height, and is *not* a BTS fan?

Answer

(c) (1 point) Is Zack a BTS fan?

Select one:

- ◯ Yes
- ◯ No

4. We've seen in class that a Gaussian Naïve Bayes classifier can learn more than just a *linear* decision boundary.

   (a) (4 points) Show that the decision boundary for a 2-D Gaussian Naïve Bayes classifier, $P(Y = 1 \mid x_1, x_2) = P(Y = 0 \mid x_1, x_2)$, is quadratic. That is, show that $P(Y = 1 \mid x_1, x_2) = P(Y = 0 \mid x_1, x_2)$ can be written as a polynomial function of x_1 and x_2 where the degree of each variable is at most 2. (In your proof, you may fold constants into terms C', C'', C''' so long as you are clearly showing each step.)

Answer

(b) (2 points) **Select all that apply:** Select all possible decision boundaries that can be produced by a Gaussian Naïve Bayes classifier. The shaded region is assigned class 1 and the unshaded region is assigned class 0.

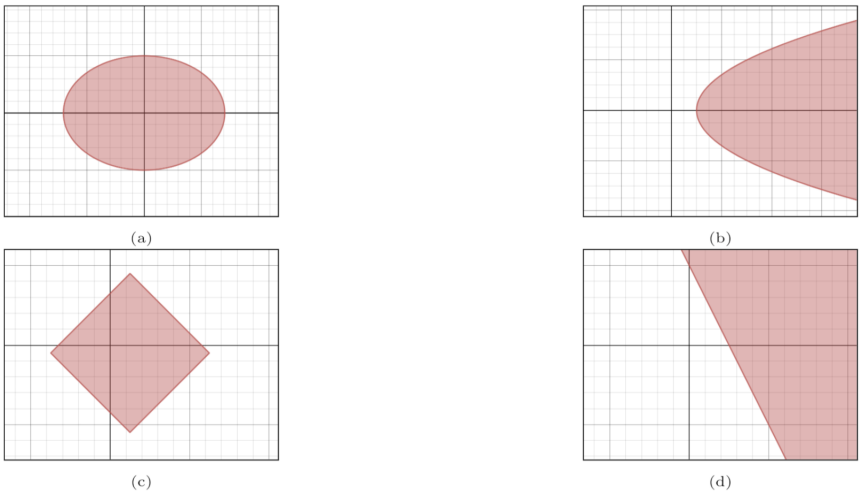


Figure 1: Decision Boundaries

- ☐ (a)
- ☐ (b)
- ☐ (c)
- ☐ (d)
- ☐ None of the above

5. (2 points) **Select all that apply:** Consider a setting where we have just one real-valued feature $X_1 \in \mathbb{R}$, from which we wish to infer the label $Y \in \{0, 1\}$.The corresponding generative story would be:

$$Y \sim \text{Bernoulli}(\phi)$$
$$X_1 \sim \text{Gaussian}(\mu_y, \sigma_y^2),$$

where the parameters are the Bernoulli parameter ϕ and the class-conditional Gaussian parameters μ_0, σ_0^2 and μ_1, σ_1^2 corresponding to $Y = 0$ and $Y = 1$, respectively.

A linear decision boundary in one dimension, of course, can be described by a rule of the form "if $X_1 > c$ then $Y = k$, else $Y = 1 - k$", where c is a real-valued threshold and $k \in \{0, 1\}$. In this one-dimensional case, is it possible to construct a Gaussian Naïve Bayes classifier with a decision boundary that cannot be expressed by a rule in the above form?

- ☐ Yes, this can occur if the Gaussians are of equal means and equal variances.
- ☐ Yes, this can occur if the Gaussians are of equal means and unequal variances.
- ☐ Yes, this can occur if the Gaussians are of unequal means and equal variances.
- ☐ Yes, this can occur if the Gaussians are of unequal means and unequal variances.
- ☐ None of the above.

4 Logistic Regression (14 points)

1. Neural trains a binary logistic regression with three features $x_1, x_2, x_3 \in \mathbb{R}$ and a bias term, and gets a final model with weight vector $\mathbf{w} = [2, 5, -3]^T$ and $b = -2$. Unfortunately, he seems to have forgotten how to work with the learned parameters, and so has asked for your help.

   (a) (2 points) Provide an equation for the decision boundary of the learned model. Simplify as much as possible.

   Answer

   (b) (1 point) Neural has held out a test point, which has features $x_1 = 2$, $x_2 = -1$ and $x_3 = -2$. What would be the predicted label of the model on this held-out test point?

   Answer

   (c) (1 point) Neural informs you the true label of the test point is $y = 1$. What is the negative log likelihood given this point?

   Round your answer to 4 decimal places after the decimal point.

   Answer

2. (2 points) **Select all that apply:** Which of the following are true regarding Gaussian Naive Bayes classifiers and logistic regression?

   - ☐ If the Naive Bayes assumption holds, Naive Bayes classifiers and logistic regression converge towards the same classifier.
   - ☐ Logistic regression generally outperforms Gaussian Naive Bayes when training data is scarce.
   - ☐ Both classifiers enable us to generate new random examples $(\mathbf{x}, y)$.
   - ☐ MAP estimation can be used to learn parameters for both Naive Bayes classifiers and logistic regression.
   - ☐ None of the above

3. (2 points) Generally, the probability threshold for a logistic regression model is defined to be 0.5. Give a reason why we may choose to use a threshold other than 0.5 when making predictions.

Answer

4. In this problem we'll explore an alternative interpretation of logistic regression, in which we model the log-odds as a linear function of the data. We will show that under this interpretation, the likelihood matches the likelihood we derived in HW4.

   Recall from HW4 that for a dataset with N training examples and with the intercept term folded into $\boldsymbol{\theta}$, the average negative log-likelihood for logistic regression is

$$\frac{1}{N}\sum_{i=1}^{N}\left[-y^{(i)}\left(\boldsymbol{\theta}^T\mathbf{x}^{(i)}\right)+\log\left(1+e^{\boldsymbol{\theta}^T\mathbf{x}^{(i)}}\right)\right] \tag{1}$$

   Our new interpretation is that the log-odds are a linear function of the features, i.e.

$$\log\frac{P(y^{(i)}=1|x^{(i)},w)}{P(y^{(i)}=0|x^{(i)},w)}=\boldsymbol{\theta}^T\mathbf{x}^{(i)} \tag{2}$$

   (a) (2 points) Write the average negative log-likelihood $\mathcal{L}(\boldsymbol{\theta})=-\frac{1}{N}\log P(\boldsymbol{y}\,|\,\mathbf{X},\boldsymbol{\theta})$ as a summation involving the variables $\boldsymbol{y}$, $P(y^{(i)}=1\,|\,x^{(i)},w)$, and $P(y^{(i)}=0\,|\,x^{(i)},w)$.

   Answer

   (b) (1 point) Using Equation 2, write $\log P(y^{(i)}=1\,|\,x^{(i)},w)$ as a function of $\boldsymbol{\theta}^T\mathbf{x}^{(i)}$ and $\log P(y^{(i)}=0\,|\,x^{(i)},w)$.

   Answer

(c) (1 point) Using Equation 2, write $P(y^{(i)} = 0 \mid x^{(i)}, w)$ as a function of $\boldsymbol{\theta}^T \mathbf{x}^{(i)}$.

Answer

(d) (2 points) Substituting your results from (b) and (c) into your result for (a), show that the average negative log-likelihood is indeed equivalent to the expression in Equation 1.

Answer

5 Convolutional Neural Networks (12 points)

1. In this problem, consider only the convolutional layer of a standard implementation of a CNN as described in Lecture 14. We are given image X and filter F below.

$X =$

1	0	-2	3	4	1
2	9	5	6	0	-1
0	-3	1	3	4	4
6	5	2	0	6	8
-5	4	-3	1	3	-2
4	1	2	8	9	7

$F =$

-1	-1	-1
-1	8	-1
-1	-1	-1

$Y =$

a	b	c	d
e	f	g	h
i	j	k	l
m	n	o	p

   (a) (1 point) Let X be convolved with F using no padding and a stride of 1 to produce an output Y. What is value of j in the output Y?

   Answer

   (b) (1 point) Suppose you had an input feature map of size (height $\times$ width) 6x4 and filter size 2x2. Using no padding and a stride of 2, what would be the resulting output size? Write your answer in the format height $\times$ width.

   Answer

2. Parameter sharing is a very important concept for CNNs because it drastically reduces the complexity of the learning problem. For the following questions, assume that there is no bias term in our convolutional layer.

   (a) (2 points) **Select all that apply:** Which of the following are parameters of a convolutional layer?

   - ☐ stride size
   - ☐ padding size
   - ☐ image size
   - ☐ filter size
   - ☐ weights in the filter
   - ☐ None of above.

(b) (2 points) **Select all that apply:** Which of the following are hyperparameters of a convolutional layer?

- □ stride size
- □ padding size
- □ image size
- □ filter size
- □ weights in the filter
- □ None of above.

(c) (2 points) Suppose for the convolutional layer, we are given grayscale images of size 22×22. Using one single 4×4 filter with a stride of 2 and no padding, what is the number of parameters you are learning in this layer?

Answer

(d) (2 points) Now suppose we do no parameter sharing. That is, each output pixel of this layer is computed by a separate 4×4 filter. Again we use a stride of 2 and no padding. What is the number of parameters you are learning in this layer?

Answer

(e) (2 points) In a sentence, describe a reason why parameter sharing is a good idea for a convolutional layer applied to image data, besides the reduction in number of learned parameters.

Answer

6 Learning Theory (17 points)

1. Alex is given a classification task to solve. He has no idea where to start, so he decides to try out a decision tree learner with 2 binary features X_1 and X_2. On the other hand, his friend Sally thinks this is a bad idea and instead decides to use logistic regression with 16 real-valued features in addition to a bias term. Sally overhears Alex talking about PAC learning and decides she would like to use it to analyze her method. She first trains her logistic regression model on N examples to obtain a training error $\hat{R}$.

   (a) (2 points) What is the the upper bound on the true error R in terms of $\hat{R}$, δ, and N? You may use big-O notation.

   Answer

   (b) (2 points) **Select one:** Sally wants to argue her method has a lower bound on the true error. Assuming Sally has obtained enough data points to satisfy PAC criterion with $\epsilon = 0.1$ and $\delta = 0.01$, which of the following is true?

   ◯ Sally is wrong. Alex's method will always classify unseen data more accurately since it is simpler as it only needs 2 binary features.

   ◯ She must first regularize her model by removing 14 features to make any comparison at all.

   ◯ It is sufficient to show that the VC Dimension of her classifier is higher than Alex's, therefore having lower bound for the true error.

   ◯ It is necessary to show that the training error she achieves is lower than the training error Alex achieves.

2. In lecture, we saw that we can use our sample complexity bounds to derive bounds on the true error for a particular algorithm. In the following parts, we will use the sample complexity bound for the infinite, agnostic case:

$$N = O\left(\frac{1}{\epsilon^2}\left[\text{VC}(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right]\right)$$

   to prove that with probability at least $(1 - \delta)$:

$$R(h) \leq \hat{R}(h) + O\left(\sqrt{\frac{1}{N}\left[\text{VC}(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right]}\right)$$

(a) (3 points) Rewrite the big-O bound in terms of N and δ using the definition of big-O notation (i.e. if $N = O(M)$ (for some value M), then there exists $c \in \mathbb{R}$ such that $N \leq cM$).

Answer

(b) (3 points) Now, using the definition of ϵ (i.e. $|R(h) - \hat{R}(h)| \leq \epsilon$), show that with probability at least $(1 - \delta)$:

$$R(h) \leq \hat{R}(h) + O\left(\sqrt{\frac{1}{N}\left[\text{VC}(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right]}\right)$$

Answer

3. (2 points) Consider the hypothesis space of functions that map M binary attributes to a binary label. A function f in this space can be characterized as $f : \{0,1\}^M \rightarrow \{0,1\}$. Your friend Paul claims that no matter the value of M, there always exists a function from this space which can shatter 2^M points.

   Is Paul correct or incorrect? If Paul is correct, briefly explain why in 1-2 *concise* sentences. If Paul is incorrect, provide a counterexample.

Answer

4. Consider instance space $\mathcal{X}$, which is the set of real numbers.

   (a) (2 points) **Select one:** What is the VC dimension of hypothesis class $\mathcal{H}$, where each hypothesis $h \in \mathcal{H}$ is of the form "if $a < x < b$ or $c < x < d$, then $y = 1$; otherwise $y = 0$"? (i.e., $\mathcal{H}$ is an infinite hypothesis class where a, b, c, and d are arbitrary real numbers).

   ◯ 2

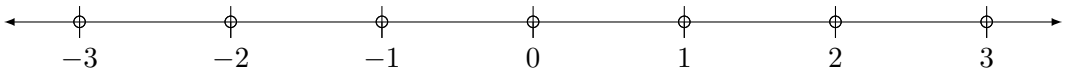
   ◯ 3

   ◯ 4

   ◯ 5

   ◯ 6

   (b) (3 points) Given the set of points in $\mathcal{X}$ below, construct a labeling of some subset of the points to show that any dimension larger than your choice of VC dimension in part (a) by *exactly* 1 is incorrect (e.g. if the VC dimension of $\mathcal{H}$ is 3, only fill in labels for 4 of the points). Fill in the boxes such that for each point in your example, the corresponding label is either 1 or 0 (for points you are not using in your example, leave the boxes blank).



-3: ☐

-2: ☐

-1: ☐

0: ☐

1: ☐

2: ☐

3: ☐

7 Collaboration Questions

After you have completed all other components of this assignment, report your answers to these questions regarding the collaboration policy. Details of the policy can be found here.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details.
2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details.
3. Did you find or come across code that implements any part of this assignment? If so, include full details.

Your Answer