HOMEWORK 6: LEARNING THEORY, MLE/MAP, AND SOCIETAL IMPACT *

10-301 / 10-601 INTRODUCTION TO MACHINE LEARNING (FALL 2026)
https://www.mlcourse.org/10601

OUT: Sunday, October 25th
DUE: Thursday, October 29th at 11:59 PM
TAs: Alan, Opally, Manaf, Saarthak, Neural the Narwhal

Summary Homework 6 covers topics on Learning Theory, MLE/MAP, Probabilistic Learning, Strategic Release Decisions, and Safety Audits.

START HERE: Instructions

- **Collaboration Policy**: Please read the collaboration policy here: https://www.mlcourse.org/10601/syllabus
- **Late Submission Policy:** See the late submission policy here: https://www.mlcourse.org/10601/syllabus
- **Submitting your work:** You will use PhiGrade to submit answers to all questions.
  - **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 and there will be a **2% penalty** (e.g., if the homework is out of 100 points, 2 points will be deducted from your final score).

*Compiled on 2026-10-25 at 08:02:28

Instructions for Specific Problem Types

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

Select One: Who taught this course?

- ● Matt Gormley
- ○ Henry Chai
- ○ 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?

- ● Matt Gormley
- ○ Henry Chai
- ⊗ Noam Chomsky

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

Select all that apply: Which are instructors for this course?

- ■ Matt Gormley
- ■ Bryan Wilder
- □ Henry Chai
- □ 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 the instructors for this course?

- ■ Matt Gormley
- ■ Bryan Wilder
- ⊠ Henry Chai
- ⊠ 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-~~6~~301

Written Questions (68 points)

1 LaTeX Point and Template Alignment (1 points)

1.1. (1 point) **Select one:** Did you use LaTeX for the entire written portion of this homework?

- ◯ Yes
- ◯ No

1.2. (0 points) **Select one:** I have ensured that my final submission is aligned with the original template given to me in the handout file and that I haven't deleted or resized any items or made any other modifications which will result in a misaligned template. I understand that incorrectly responding yes to this question will result in a penalty equivalent to 2% of the points on this assignment.
Note: Failing to answer this question will not exempt you from the 2% misalignment penalty.

- ◯ Yes

1.3. (0 points) **Select one:** Did you fill out the Exit Poll for the previous HW? Completing the exit poll will count towards your participation grade.

- ◯ Yes

2 Learning Theory (15 points)

2.1. Neural the Narwhal is given a classification task to solve, and he decides to use a decision tree learner with 2 binary features X_1 and X_2. On the other hand, you think that Neural should not have used a decision tree. Instead, you think he would get a better bound on generalization error by using logistic regression with 16 real-valued features plus a bias term. You want to use the framework of PAC learning to reason about which model is better.

You first train your logistic regression model on N examples to obtain a training error $\hat{R}$.

2.1.a. (1 point) Which of the following cases of PAC learning should you use for your logistic regression model?

- ◯ Finite and realizable
- ◯ Finite and agnostic
- ◯ Infinite and realizable
- ◯ Infinite and agnostic

2.1.b. (2 points) What is the upper bound on the true error R in terms of $\hat{R}$, δ, VC($\mathcal{H}$) and N? You may use big-$\mathcal{O}$ notation if necessary. Write only the final answer. Your work will *not* be graded.
Note: Your answer must be in terms of only the quantities listed above. Use the formulas given in the recitation handout to solve this question.

Your Answer

2.1.c. (1 point) What is the value of the VC dimension in part (b)? Provide a single numerical value.

Your Answer

2.1.d. (2 points) **Select one:** You want to argue your method has a better (smaller) bound on the true error as compared to the Neural's true error bound. Assume that you have obtained enough data points to satisfy the PAC criterion with the same ϵ and δ as Neural. Which of the following is true?

- ◯ Neural's model will always classify unseen data more accurately because it only needs 2 binary features and therefore is simpler.
- ◯ You must first regularize your model by removing 14 features to make any comparison at all.
- ◯ It is sufficient to show that the VC dimension of your classifier is higher than that of Neural's, therefore having a better bound for the true error.
- ◯ You must show that your model's total true error bound (combining training error and VC complexity) is smaller than Neural's.

2.2. (3 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 $f : \{0, 1\}^M \rightarrow \{0, 1\}$. Neural the Narwhal says that regardless of the value of M, a hypothesis class containing all possible functions in this space can always shatter 2^M points. Is Neural wrong? If so, provide a counterexample. If Neural is right, briefly explain why in 1–2 *concise* sentences.
Hint: Decision tree-like structures are one way to help understand the possible hypothesis space.

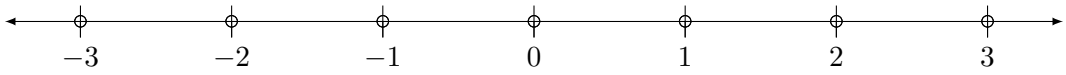
Your Answer

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

2.3.a. (3 points) **Select one:** What is the VC dimension of hypothesis class H, where each hypothesis h in H is of the form "if $a < x < b$ or $c < x < d$ then $y = 1$; otherwise $y = 0$"? (i.e., H is an infinite hypothesis class where a, b, c, and d are arbitrary real numbers).

- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6

2.3.b. (3 points) Choose VC$(H) + 1$ of the points in $\mathcal{X}$ below and give each a label of 0 or 1 so that no hypothesis in H can produce this labeling. Write N/A for every point you do not use (do *not* leave the answer box blank).



Answer for -3	Answer for -2	Answer for -1	

Answer for 0	Answer for 1	Answer for 2	Answer for 3

3 MLE/MAP (8 points)

3.1. (1 point) **True or False:** Recall that, given a dataset $\mathcal{D}$, the MLE and the MAP estimate of a parameter θ are

$$\hat{\theta}_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}}\ p(\mathcal{D} \mid \theta) \qquad \text{and} \qquad \hat{\theta}_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}}\ p(\mathcal{D} \mid \theta)\, p(\theta).$$

Suppose the parameter space is bounded. Then for every dataset, there exists a prior $p(\theta)$ under which the MAP estimate equals the MLE.

◯ True

◯ False

3.2. (1 point) **True or False:** You place a Beta(α, β) prior on the parameter θ of a Bernoulli distribution and learn θ from i.i.d. samples. An adversary chooses the hyperparameters $\alpha, \beta > 0$ to mislead your learning algorithm, but they must be finite and stay fixed. As the number of training examples grows to infinity, the difference between the MAP estimate and the MLE of θ goes to zero.

◯ True

◯ False

3.3. (2 points) Let Γ be a random variable with probability density function (pdf)

$$f_\Gamma(\gamma) = \begin{cases} 2\gamma & \text{if } 0 \le \gamma \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Given $\Gamma = \gamma$, a second random variable Y follows an exponential distribution with rate $\lambda = 3\gamma$. Recall that the exponential distribution with rate λ has pdf

$$f(y \mid \lambda) = \begin{cases} \lambda e^{-\lambda y} & \text{if } y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

You observe $Y = \frac{1}{3}$. What is the MAP estimate of γ? Give a single number.

Your Answer

3.4. (4 points) Neural the Narwhal found a mystery coin and wants to know its probability of landing heads. Neural models each toss as a draw from Bernoulli(θ), where θ is the probability of heads, then flips the coin three times and observes heads, tails, heads. An oracle reveals that $\theta \in \{0, 0.25, 0.5, 0.75, 1\}$, and *no other values of θ should be considered*.

Find the MLE and the MAP estimate of θ. For the MAP estimate, use the prior

θ	0	0.25	0.5	0.75	1
$p(\theta)$	0.90	0.04	0.03	0.02	0.01

Each of your answers must be one of the five allowed values of θ.

MLE of θ

MAP of θ

4 Probabilistic Learning (17 points)

4.1. In a previous homework, you derived the closed-form solution for linear regression. Here we revisit linear regression as a probabilistic model and derive the MLE and MAP estimates of its parameters. Recall that

$$\hat{\theta}_{\text{MLE}} = \underset{\theta}{\text{argmax}}\ p(\mathcal{D} \mid \theta) \qquad \text{and} \qquad \hat{\theta}_{\text{MAP}} = \underset{\theta}{\text{argmax}}\ p(\theta \mid \mathcal{D}).$$

We have data $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$: N examples, each with M features $\mathbf{x}^{(i)} = (x_1^{(i)}, \ldots, x_M^{(i)})^T$. Each label is generated as

$$y^{(i)} = \mathbf{w}^T \mathbf{x}^{(i)} + \epsilon^{(i)}, \qquad \epsilon^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2),$$

where $\mathbf{w} \in \mathbb{R}^M$ is the parameter vector and the noise variance σ^2 is a known constant. Throughout, treat the inputs $\mathbf{x}^{(i)}$ as fixed, so $p(\mathcal{D} \mid \mathbf{w})$ means $\prod_{i=1}^{N} p(y^{(i)} \mid \mathbf{x}^{(i)}, \mathbf{w})$.

4.1.a. (2 points) **Select one:** Under this model, what is the distribution of $y^{(i)}$ given $\mathbf{x}^{(i)}$?

- ◯ $y^{(i)} \mid \mathbf{x}^{(i)} \sim \mathcal{N}(\mathbf{w}^T \mathbf{x}^{(i)}, \sigma^2)$
- ◯ $y^{(i)} \mid \mathbf{x}^{(i)} \sim \mathcal{N}(0, \sigma^2)$
- ◯ $y^{(i)} \mid \mathbf{x}^{(i)} \sim \text{Uniform}(\mathbf{w}^T \mathbf{x}^{(i)} - \sigma,\ \mathbf{w}^T \mathbf{x}^{(i)} + \sigma)$
- ◯ None of the above

4.1.b. (2 points) **Select one:** Next, we find the MLE of $\mathbf{w}$. Which expression is the conditional log-likelihood $\ell(\mathbf{w}) = \log p(\mathcal{D} \mid \mathbf{w})$?

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

4.1.c. (2 points) **Select all that apply:** The MLE is $\hat{\mathbf{w}}_{\text{MLE}} = \text{argmax}_{\mathbf{w}}\, \ell(\mathbf{w})$. Which of the following are equal to $\hat{\mathbf{w}}_{\text{MLE}}$ for *every* dataset?

- ☐ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N} \left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)}) \right]$
- ☐ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N} \left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2 \right]$
- ☐ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N} \left[-\frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2 \right]$
- ☐ $\text{argmin}_{\mathbf{w}} \sum_{i=1}^{N} \left| y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)} \right|$
- ☐ $\text{argmin}_{\mathbf{w}} \sum_{i=1}^{N} \frac{1}{2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2$
- ☐ None of the above

4.1.d. (1 point) **Select one:** Suppose we derive $\hat{\mathbf{w}}_{\text{MLE}}$ with $\sigma = 3.7$. How would $\hat{\mathbf{w}}_{\text{MLE}}$ change if we instead used $\sigma = 5.1$?

- ◯ It would change, because the noise variance changes the scale of the objective.
- ◯ It would not change, because σ^2 only scales the squared-error term by a positive constant and adds a constant, and neither moves the maximizer.
- ◯ It would change, but only if $\sigma < 1$.
- ◯ None of the above

4.2. Now we find the MAP estimate of $\mathbf{w}$ for the same model and data $\mathcal{D}$.

4.2.a. (2 points) **Select all that apply:** Which of the following are equal to $\hat{\mathbf{w}}_{\text{MAP}}$? Recall that $\mathcal{D}$ is the data and $\mathbf{w}$ is the vector of regression weights.

- ☐ $\text{argmax}_{\mathbf{w}}\ p(\mathcal{D}, \mathbf{w})$
- ☐ $\text{argmax}_{\mathbf{w}}\ \dfrac{p(\mathcal{D} \mid \mathbf{w})\, p(\mathbf{w})}{p(\mathcal{D})}$
- ☐ $\text{argmax}_{\mathbf{w}}\ \dfrac{p(\mathcal{D}, \mathbf{w})}{p(\mathbf{w})}$
- ☐ $\text{argmax}_{\mathbf{w}}\ p(\mathcal{D} \mid \mathbf{w})\, p(\mathbf{w})$
- ☐ $\text{argmax}_{\mathbf{w}}\ p(\mathbf{w} \mid \mathcal{D})$
- ☐ None of the above

4.2.b. (2 points) **Select one:** Place an independent Gaussian prior on each weight: $w_m \sim \mathcal{N}(0, \frac{1}{\lambda})$ for $m = 1, \ldots, M$, where $\lambda > 0$. Which expression is the log joint probability $\log p(\mathcal{D}, \mathbf{w})$?

- ◯ $\sum_{i=1}^{N} \left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T \mathbf{x}^{(i)})^2 \right] + \sum_{m=1}^{M} \left[-\log \sqrt{2\pi\lambda} - \frac{1}{2\lambda} w_m^2 \right]$
- ◯ $\sum_{i=1}^{N} \left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T \mathbf{x}^{(i)}) \right] + \sum_{m=1}^{M} \left[-\log \sqrt{2\pi/\lambda} - \frac{\lambda}{2} w_m^2 \right]$
- ◯ $\sum_{i=1}^{N} \left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T \mathbf{x}^{(i)})^2 \right] - \log \sqrt{2\pi/\lambda} - \frac{\lambda}{2} \|\mathbf{w}\|_2^2$
- ◯ $\sum_{i=1}^{N} \left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T \mathbf{x}^{(i)})^2 \right] + \sum_{m=1}^{M} \left[-\log \sqrt{2\pi/\lambda} - \frac{\lambda}{2} w_m^2 \right]$

4.2.c. (2 points) **Select one:** With the same Gaussian prior, the MAP estimate maximizes the log posterior $\ell_{\text{MAP}}(\mathbf{w}) = \log p(\mathbf{w} \mid \mathcal{D})$. Which of the following equals $\text{argmax}_{\mathbf{w}}\, \ell_{\text{MAP}}(\mathbf{w})$?

- ◯ $\text{argmax}_{\mathbf{w}} \sum_{i=1}^{N} \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2 + \frac{\lambda}{2}\|\mathbf{w}\|_2^2$
- ◯ $\text{argmin}_{\mathbf{w}} \sum_{i=1}^{N} \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2 + \frac{\lambda}{2}\|\mathbf{w}\|_2^2$
- ◯ $\text{argmin}_{\mathbf{w}} \sum_{i=1}^{N} \frac{1}{2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2 + \frac{\lambda}{2}\|\mathbf{w}\|_2^2$
- ◯ $\text{argmin}_{\mathbf{w}} -\sum_{i=1}^{N} \frac{1}{2\sigma^2}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2 - \frac{\lambda}{2}\|\mathbf{w}\|_2^2$

4.2.d. (2 points) **Select one:** You compute $\hat{\mathbf{w}}_{\text{MAP}}$ with a zero-mean Gaussian prior and find that its test error is much higher than its training error. What is the most likely issue, and how could you fix it?

- ◯ Overfitting; increase the variance of the prior
- ◯ Overfitting; decrease the variance of the prior
- ◯ Underfitting; increase the variance of the prior
- ◯ Underfitting; decrease the variance of the prior

4.3. (2 points) **Select one:** MAP estimation of $\mathbf{w}$ with which prior is equivalent to ℓ_1-regularized linear regression? In every option, the weights $w_1, \ldots, w_M$ are independent and each follows the stated distribution.

For reference:

- The uniform distribution over $[a, b]$ has pdf $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$, and 0 otherwise.
- The exponential distribution with rate $a > 0$ has pdf $f(x) = ae^{-ax}$ for $x \geq 0$, and 0 otherwise.
- The Laplace distribution with location a and scale $b > 0$ has pdf $f(x) = \frac{1}{2b} \exp\left(-\frac{|x-a|}{b}\right)$ for all $x \in \mathbb{R}$.

◯ Uniform distribution over $[-1, 1]$

◯ Gaussian distribution $\mathcal{N}(0, \frac{1}{\lambda})$ for some $\lambda > 0$

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

◯ Laplace distribution with location $a = 1$

◯ Laplace distribution with location $a = 0$

◯ None of the above

5 Learning to Race (21 points)

In November 2022, OpenAI released ChatGPT and set off a race. Google hurried to announce its own chatbot, Bard, about two months later. Since then, news reports say some labs' safety testing has shrunk from months to days. This question models that race.

Two AI firms have each built a powerful new system and must decide whether to **release** it now or **hold** it back for more safety work. All amounts are in **billions of US dollars** over the next year.

- **Market.** The market is worth M. If both firms hold, or both release, they split it ($M/2$ each). If only one releases, it takes the whole market.
- **Disaster.** A release risks a public disaster. The backlash (emergency regulation, lawsuits, and collapsing public trust) costs **each firm** D, whether or not it released. For example, after the 2011 accident at one company's Fukushima nuclear plant, all 50 of Japan's reactors were taken offline within about a year. The probability of disaster is r if one firm releases and q if both do. There is no disaster if both hold.
- **Payoff.** A firm's payoff is its market share minus its expected disaster cost.

Policies and notation. A firm may also randomize: Firm i releases with probability p_i.

$p_i \in [0, 1]$	Firm i's release probability: the chance that it releases before its safety work is done ($p_i = 0$: always hold; $p_i = 1$: always release).
j	The other firm.
$u_i(a, b)$	Firm i's payoff in the payoff table when it takes action a and the other firm takes action b, where R means release and H means hold.

Scenario A: an automated AI researcher ($M = 100$, $D = 1000$, $r = 8\%$, $q = 12\%$). Each firm has built an agent that designs experiments, writes code, and trains better models, including better versions of itself. The market for such agents is worth \$100 billion a year. The disaster is that AI progress outpaces human oversight and a model behaves in ways nobody can understand or control. Governments respond by halting frontier AI development everywhere, which costs each firm \$1 trillion, about its entire value. Note that $q = 12\%$ is less than $2r = 16\%$.

5.1. (4 points) **Reading the game.**

5.1.a. Complete the payoff table for Scenario A. Each cell lists (Firm 1, Firm 2) in billions of dollars. The both-hold cell is given.

	Firm 2 holds	Firm 2 releases
Firm 1 holds	$(50, 50)$	Your Answer
Firm 1 releases	Your Answer	Your Answer

5.1.b. **Select all that apply:** Which statements about Scenario A are true?

- ☐ Holding is a dominant strategy, because both hold gives the highest total payoff.
- ☐ Scenario A is a prisoner's dilemma: both firms would be better off holding, yet they can end up releasing.
- ☐ If both firms release, a firm that switches to hold lowers its own payoff, even though holding is the safer action.
- ☐ Because $q < 2r$, the second firm's release lowers the probability of disaster.
- ☐ A firm that holds while the other firm releases avoids the disaster cost.
- ☐ The first release already lets AI progress outpace human oversight, so a second agent adds comparatively little new risk. This is why $q < 2r$.
- ☐ Neither firm has a dominant strategy, yet both hold and both release are each equilibria.

5.2. (2 points) **The tipping point.** Suppose the other firm releases with probability p. Your expected payoffs from releasing and from holding are $U(\mathrm{R}) = (1-p)\,u(\mathrm{R},\mathrm{H}) + p\,u(\mathrm{R},\mathrm{R})$ and $U(\mathrm{H}) = (1-p)\,u(\mathrm{H},\mathrm{H}) + p\,u(\mathrm{H},\mathrm{R})$, and your *advantage of releasing* is $A = U(\mathrm{R}) - U(\mathrm{H})$.

5.2.a. **Select one:** In Scenario A, your advantage of releasing is: ◯ $30 - 40p$ ◯ $50 - 130p$ ◯ $20 - 90p$ ◯ $40p - 30$

5.2.b. Releasing is better than holding exactly when p is greater than what value?

Your Answer

How the firms learn. Earlier you found the equilibria directly. Now we look at how the firms could get to one. Neither firm sees the other's decision beforehand, so imagine each firm running this check: if both firms kept their current release probabilities, would it gain by changing its own?

Each round, firm i looks at the other firm's current p_j and takes one step of gradient ascent on its own expected payoff:

$$
\begin{aligned}
U_i(\mathrm{R}) &= (1-p_j)\,u_i(\mathrm{R},\mathrm{H}) + p_j\,u_i(\mathrm{R},\mathrm{R}) && \text{average payoff if firm } i \text{ releases}\\
U_i(\mathrm{H}) &= (1-p_j)\,u_i(\mathrm{H},\mathrm{H}) + p_j\,u_i(\mathrm{H},\mathrm{R}) && \text{average payoff if firm } i \text{ holds}\\
V_i &= p_i\,U_i(\mathrm{R}) + (1-p_i)\,U_i(\mathrm{H}) && \text{expected payoff}\\
\frac{\partial V_i}{\partial p_i} &= U_i(\mathrm{R}) - U_i(\mathrm{H}) = A_i && \text{advantage of releasing } (p_j \text{ held fixed})\\
p_i &\leftarrow \mathrm{clip}_{[0,1]}\big(p_i + \eta\,A_i\big), \quad \eta = 0.005 && \text{gradient ascent step}
\end{aligned}
$$

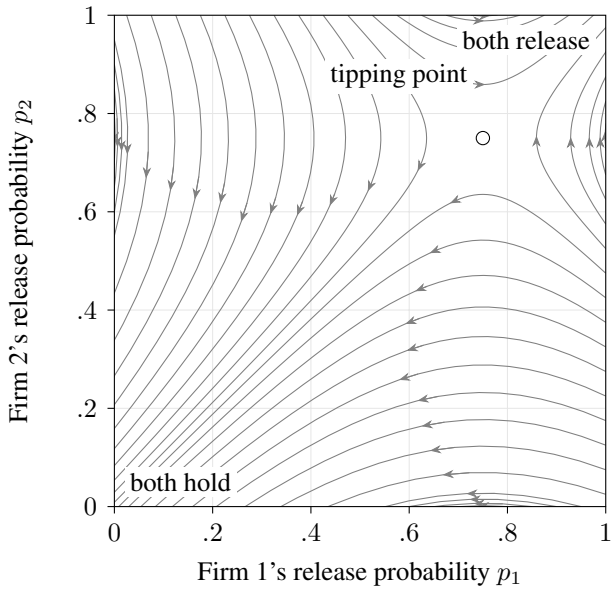
where $\mathrm{clip}_{[0,1]}$ keeps p_i between 0 and 1. So if $A_i = 4$ (payoffs in \$ billions), p_i goes up by 0.02, and a negative A_i pushes it down. Both firms update simultaneously.

The updates stop when neither p_i changes, either because $A_i = 0$ or because p_i is already at 0 or 1 and A_i pushes it further out. These stopping points are the equilibria: "both hold" means $p_1 = p_2 = 0$, and "both release" means $p_1 = p_2 = 1$. In some two-player games these updates cycle forever. That doesn't happen in any game here, so you can assume the updates converge for the following questions.

5.3. (2 points) **Same starting point. Select all that apply:** Both firms start at the same probability p_0 and follow the learning rule. Because they are identical, they move together. Which statements are true?

- □ From $p_0 = 0.8$, each firm's expected payoff rises every round, because each firm takes gradient-ascent steps on its own expected payoff.
- □ If $p_0 = 0.74$, both firms end at hold.
- □ A large enough learning rate would make firms starting at $p_0 = 0.8$ overshoot the tipping point and end at hold.
- □ Like gradient-based training on a non-convex objective, the same learning rule can reach different outcomes from different initializations.
- □ Until a probability is clipped, the gap between p and $3/4$ grows by a factor of 1.2 every round.
- □ The point $p_1 = p_2 = 0.75$ is stable: if one firm is nudged slightly away from it, learning brings both firms back to 0.75.

5.4. (3 points) **Different starting points.** The figure shows the direction of learning in Scenario A. In each case below, **Firm 1's initial release probability is** $p_1 = 0.9$.



Direction of learning in Scenario A. Arrows show how (p_1, p_2) moves.

Select all that apply: Which statements are true?

- ☐ If Firm 2 starts at $p_2 = 0.5$, then at first Firm 1 becomes more cautious and Firm 2 becomes more aggressive.
- ☐ If Firm 2 starts at $p_2 = 0.7$, it ends up releasing because its own starting probability is above the tipping point.
- ☐ If Firm 2 starts at $p_2 = 0.7$, both firms end at release.
- ☐ If Firm 2 starts at $p_2 = 0.5$, both firms end at hold.
- ☐ Whichever firm starts more aggressive ends up releasing alone.
- ☐ Each firm's direction depends on the *other* firm's current release probability: the $3/4$ threshold applies to the other firm's probability, not its own.

5.5. (4 points) **Scenario B: robotaxis ($M = 20$, $D = 100$, $r = 2\%$, $q = 5\%$).** Each firm must decide whether to launch a fully driverless taxi service nationwide now or keep testing with human safety drivers. The national robotaxi market is worth $20 billion a year. Suppose a fatal driverless crash leads regulators to ground every company's fleet nationwide, which costs each firm $100 billion in lost value. This is close to what happened in 2018: after an Uber self-driving test car killed a pedestrian, Toyota and Nvidia also paused their own self-driving tests. With two fleets, risk more than doubles: there are more driverless cars, and each system must predict robot drivers that behave differently from humans. The learning rule is the same.

5.5.a. Complete the payoff table for Scenario B. Each cell lists (Firm 1, Firm 2) in billions of dollars. The both-hold cell is given.

	Firm 2 holds	Firm 2 releases
Firm 1 holds	$(10, 10)$	Your Answer
Firm 1 releases	Your Answer	Your Answer

5.5.b. **Select all that apply:** Which statements about Scenario B are true?

- ☐ From any starting point, learning ends at both release.
- ☐ Releasing pays more only when the other firm holds.
- ☐ Both hold is an equilibrium of Scenario B.
- ☐ A single planner who trains both probabilities to maximize the firms' combined payoff pushes both firms to hold.
- ☐ Releasing pays more no matter what the other firm does.

5.5.c. **Select all that apply:** Why does learning race? Which statements are true?

- ☐ The firms race only because their learning rate is too large.
- ☐ If Firm 2 holds and Firm 1 switches from hold to release, Firm 1 gains $8 billion and Firm 2 loses $12 billion.
- ☐ Harm to crash victims is not separately counted in either firm's payoff, so neither update directly optimizes their welfare.
- ☐ If each firm learned from the combined payoff instead, both would still end up releasing.
- ☐ Firm 1's update sees only its own payoff change, so the cost its release imposes on Firm 2 is an externality that it ignores.

5.6. (2 points) **Which game is it?** Scenarios A and B differ only in M, D, r, and q. Work out, in general, which game the payoff table describes. Each row and column below gives a condition; in each cell, write the game that holds when both conditions are true. Choose from: prisoner's dilemma, stag hunt, *chicken* (each firm wants to do the opposite of the other: release only if the other firm holds), and *harmony* (holding pays more no matter what the other firm does).

	$\frac{M}{2} > (q-r)D$	$\frac{M}{2} < (q-r)D$
$\frac{M}{2} > rD$	Your Answer	Your Answer
$\frac{M}{2} < rD$	Your Answer	Your Answer

5.7. (4 points) **Changing the game with a pre-release audit.** Continue with Scenario B. Regulators fine any firm $F \geq 0$ (in billions of dollars) for launching before an independent safety audit is complete: subtract F from every payoff that firm earns when it releases. This adds a penalty term to its objective, like a regularizer. The learning rule stays the same.

5.7.a. **Select one:** What is the smallest fine F for which holding is **at least as good** as releasing, no matter what the other firm does?

- ◯ 7
- ◯ 7.5
- ◯ 8
- ◯ 10

5.7.b. **Select all that apply:** Now set $F = 7.5$. Which statements are true?

- □ If Firm 1 starts at 0.51 and Firm 2 at 0.49, learning ends with Firm 1 releasing and Firm 2 holding.
- □ If Firm 1 starts at 0.51 and Firm 2 at 0.49, the small difference shrinks and both settle at 0.5.
- □ Firms that start at exactly the same p get identical updates and stay identical.
- □ Holding pays more no matter what the other firm does.
- □ Hidden units in the same layer with identical incoming weights, biases, and outgoing weights stay identical under deterministic gradient updates that treat them symmetrically, and small random differences at initialization can break this symmetry.
- □ Releasing pays more only when the other firm holds.

6 Estimating Risk: MLE, MAP, and Safety Audits (6 points)

Return to Scenario A of the previous question, but assume Firm 1 does not know the true value of r. Firm 1 observes **zero incidents in 10 independent tests**, where an incident is a test in which the disaster would have happened. Assume each test has the same disaster probability r as a real release.

It compares MLE with MAP using a $\mathrm{Beta}(2, 8)$ prior:

$$p(r) \propto r(1-r)^7.$$

This prior counts like one incident in eight earlier trials, for example from tests of the firm's previous model. Assume Firm 2 holds. Firm 1 uses its estimate $\hat{r}$ to compare (in billions of dollars):

$$\text{Release payoff} = 100 - 1000\hat{r}, \qquad \text{Hold payoff} = 50.$$

6.1. (3 points) **Estimation and decisions. Select all that apply:** Which of the following statements are true?

- □ With the same prior, 12 incident-free tests are enough for MAP to make releasing strictly better than holding.
- □ The posterior density is proportional to $r(1-r)^{17}$, so $\hat{r}_{\mathrm{MAP}} = 1/18$.
- □ After these 10 tests, MLE favors releasing, while MAP favors holding.
- □ With the same prior, 13 is the smallest total number of incident-free tests for which MAP makes releasing strictly better than holding.
- □ If the firm observes 100 incident-free tests instead of 10, its MAP estimate remains $1/18$, because the prior has not changed.
- □ The likelihood is $(1-r)^{10}$, so $\hat{r}_{\mathrm{MLE}} = 0$.
- □ The positive MAP estimate proves that the true disaster probability is greater than zero.

6.2. (3 points) **Estimating risk versus auditing safety.**

The firm now has **13 incident-free tests**. Separately, an audit accepts a system only if all n tests are incident-free. The audit requires that any system with $r \geq 0.05$ have at most a 5% chance of passing.

Use:

$$P(\text{all tests pass} \mid r) = (1-r)^n \leq e^{-rn}.$$

Select all that apply: Which of the following statements are true?

- ☐ Once MLE and MAP both favor releasing, additional test results cannot make either method favor holding.
- ☐ The smallest n satisfying $e^{-0.05n} \leq 0.05$ is 60.
- ☐ The 13 clean tests prove that the true risk is below 5%, because MAP now favors releasing.
- ☐ Under the stated assumptions, a system whose true risk is at least 5% has at most a 5% chance of passing the 60-test audit.
- ☐ Passing the 60-test audit means there is exactly a 95% posterior probability that the system's risk is below 5%.
- ☐ If the true risk is 8%, the probability of 13 incident-free tests is approximately 0.34.

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