HW0 – Math Background Diagnostic

Foundations of Privacy (Fall 2026)

Instructions

This is a self-assessment, not a graded assignment. Its purpose is to help you decide whether you have the math background to take *Foundations of Privacy* comfortably. Work through it on your own, without external help, in roughly 90 minutes to 2 hours. After you finish, consult the self-scoring rubric at the end.

The problems exercise the following skills, all of which we will use repeatedly during the semester:

(1) Summation, product, and max/arg min notation.

(2) Basic probability: expectation, variance, independence.

(3) Conditional probability and Bayes' rule.

(4) Common continuous distributions (Laplace, Gaussian).

(5) Manipulating e^x and $\ln(\cdot)$ as inverse operations.

(6) Vector norms (ℓ_1, ℓ_2) and basic linear algebra.

(7) Partial derivatives and gradients.

(8) Reading and writing statements with quantifiers ("for all...", "there exists...").

(9) Basic proof techniques: case analysis and "without loss of generality" (WLOG).

Problem 1.

This problem checks fluency with the summation, product, and "max / argmin"-style notation used throughout the course.

(a) Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers, and let $c \in \mathbb{R}$ be a constant. Simplify (or rewrite using $\sum$ notation): (i) $\sum_{i=1}^{n} c \cdot a_i$, (ii) $\sum_{i=1}^{n} (a_i + b_i)$, (iii) $\sum_{i=1}^{n} c$ (the constant c summed n times).

(b) Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ be real numbers. Show that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j = \Big(\sum_{i=1}^{n} a_i \Big) \Big(\sum_{j=1}^{m} b_j \Big).$$

Hint: pull out factors that don't depend on the inner summation index.

(c) Simplify $\prod_{i=1}^{n} e^{x_i}$ in terms of a single $e^{(\cdot)}$ expression.

(d) Let $z_1, \dots, z_n \in \mathbb{R}$. Write the expression

$$\prod_{i=1}^{n} \tfrac{1}{2}\, e^{-|z_i|}$$

as a product of a constant and a single exponential of a sum.

(e) max, arg max, arg min. Recall: $\max_i a_i$ is the largest *value* among the a_i's, while $\arg\max_i a_i$ is the *index* (or set of indices) at which the max is attained. The same distinction holds for min and arg min. State the value of each: (i) $\max_{i\in\{1,2,3,4\}} a_i$ for $(a_1, a_2, a_3, a_4) = (5, -1, 5, 2)$; (ii) $\arg\max_{i\in\{1,2,3,4\}} a_i$ for the same a_i's; (iii) $\arg\min_{w\in\mathbb{R}}(w-3)^2$.

Problem 2.

Let $X_1, \dots, X_n$ be i.i.d. (*independent and identically distributed*) random variables with

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = 0] = 1 - p, \qquad p \in [0, 1].$$

"Identically distributed" means each X_i has the same distribution (in this case, Bernoulli(p)). "Independent" means knowing the value of one X_i tells you nothing about the others; equivalently, $\mathbb{P}[X_i = a, X_j = b] = \mathbb{P}[X_i = a]\,\mathbb{P}[X_j = b]$ for all $i \neq j$ and all values a, b. Define $S = \sum_{i=1}^{n} X_i$ and $\bar{X} = S/n$.

(a) Compute $\mathbb{E}[X_i]$ and $\mathrm{Var}(X_i)$.

(b) Compute $\mathbb{E}[S]$ and $\mathrm{Var}(S)$. State where you use independence and where you only use linearity of expectation.

(c) Compute $\mathbb{E}[\bar{X}]$ and $\mathrm{Var}(\bar{X})$. What happens to $\mathrm{Var}(\bar{X})$ as $n \to \infty$?

Problem 3.

Quick recap. For an event A, $\neg A$ ("not A") denotes the *complement* of A: the event that A does *not* occur. So $\mathbb{P}[\neg A] = 1 - \mathbb{P}[A]$. For events A, B with $\mathbb{P}[B] > 0$, the *conditional probability* of A given B is

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Bayes' rule flips the order of conditioning:

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\,\mathbb{P}[A]}{\mathbb{P}[B]}.$$

The denominator is usually computed by the *law of total probability*: $\mathbb{P}[B] = \mathbb{P}[B \mid A]\,\mathbb{P}[A] + \mathbb{P}[B \mid \neg A]\,\mathbb{P}[\neg A]$.

A diagnostic test for a rare condition has the following behavior:

- If a person *has* the condition, the test reports "positive" with probability 0.95.
- If a person *does not* have the condition, the test reports "positive" with probability 0.05.

In the population, 1% of people have the condition. A random person is tested.

(a) Let C be the event "the person has the condition" and T be the event "the test reports positive." Write $\mathbb{P}[T \mid C]$, $\mathbb{P}[T \mid \neg C]$, and $\mathbb{P}[C]$ from the problem statement.

(b) Use Bayes' rule to compute $\mathbb{P}[C \mid T]$. Show your steps.

(c) Suppose instead the prevalence were 50% ($\mathbb{P}[C] = 0.5$). Recompute $\mathbb{P}[C \mid T]$. Briefly explain (one or two sentences) why the answer changes so much even though the test's accuracy is unchanged.

Problem 4.

This problem checks your conceptual understanding of continuous probability distributions over $\mathbb{R}$. You will not need to compute any integrals.

Recap. A continuous random variable Z on $\mathbb{R}$ is described by its *probability density function* (pdf) $f : \mathbb{R} \to [0, \infty)$. The pdf is **not** itself a probability. Instead:

- The probability that Z lies in an interval $[a, b]$ is the *area under the curve* of f between a and b:
$$\mathbb{P}[a \le Z \le b] = \int_a^b f(y)\, dy.$$
- Consequently, $f(y) \ge 0$ for all y, and the total area equals 1: $\int_{-\infty}^{\infty} f(y)\, dy = 1$.

(a) *True or false, with one-sentence explanation.* If Z has continuous pdf f and $f(5) = 0.4$, then $\mathbb{P}[Z = 5] = 0.4$.

(b) *True or false, with one-sentence explanation.* A pdf f must satisfy $f(y) \le 1$ for all y.

(c) Suppose Z has a pdf f that is symmetric around 0 (i.e., $f(-y) = f(y)$ for all y). Without computing any integrals, explain why $\mathbb{P}[Z > 0] = \mathbb{P}[Z < 0] = \frac{1}{2}$.

(d) The **Laplace** distribution $\mathrm{Lap}(\lambda)$ has pdf
$$h_\lambda(y) = \tfrac{1}{2\lambda} \exp(-|y|/\lambda), \quad y \in \mathbb{R},\ \lambda > 0.$$
State (no derivation needed; you may look these up) the mean and variance of $Z \sim \mathrm{Lap}(\lambda)$.

(e) The **Gaussian** (Normal) distribution $\mathcal{N}(\mu, \sigma^2)$ has pdf
$$\phi(y) = \tfrac{1}{\sqrt{2\pi}\,\sigma} \exp\!\left(-\tfrac{(y-\mu)^2}{2\sigma^2}\right), \quad y \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma > 0.$$
State the mean and variance of $W \sim \mathcal{N}(\mu, \sigma^2)$.

(f) *Intuition.* For the Laplace distribution $\mathrm{Lap}(\lambda)$, suppose we double the scale parameter from λ to 2λ. Without computing anything, describe in one sentence what happens to the typical magnitude of Z (does it get larger, smaller, or stay the same?).

Problem 5.

This problem checks comfort with e^x and ln as a pair of inverse operations. In the course, the DP definition is stated using e^ε, but many proofs and arguments work in the equivalent ln form (the so-called "privacy loss"). You should be able to move between the two fluently.

Recap. For all real a, b and all $x, y > 0$:
$$e^{a+b} = e^a e^b, \qquad \ln(xy) = \ln x + \ln y, \qquad \ln(e^a) = a, \qquad e^{\ln x} = x.$$
ln is monotone increasing on $(0, \infty)$, and $e^x > 0$ for every real x.

(a) Simplify each expression so the answer involves e^a and e^b only (no nested exponentials): (i) e^{a+b}, (ii) e^{a-b}, (iii) $(e^a)^k$ for $k \in \mathbb{R}$.

(b) Simplify each expression for $x, y > 0$ and $k \in \mathbb{R}$: (i) $\ln(xy)$, (ii) $\ln(x/y)$, (iii) $\ln(x^k)$.

(c) Solve for x: (i) $e^{2x} = 5$, (ii) $\ln(3x) = 1$.

(d) Suppose $p, q > 0$ and $\varepsilon > 0$. Show that the inequality $p \le e^{\varepsilon} \cdot q$ is equivalent to $\ln p - \ln q \le \varepsilon$. (One direction of equivalence in each direction; cite the property of ln you are using.)

(e) *Quick computation.* If $\ln(p/q) = 0.7$, write p/q in the form $e^{?}$. What is the exponent?

Problem 6.

For a vector $v = (v_1, \ldots, v_d) \in \mathbb{R}^d$, recall the ℓ_1 *norm* (also called the *Manhattan* or *taxicab* norm) and the ℓ_2 *norm* (also called the *Euclidean norm*):

$$\underbrace{\|v\|_1 = \sum_{i=1}^{d} |v_i|}_{\ell_1 \text{ norm}}, \qquad \underbrace{\|v\|_2 = \sqrt{\sum_{i=1}^{d} v_i^2}}_{\ell_2 \text{ norm (Euclidean)}}.$$

For two vectors $u, w \in \mathbb{R}^d$, the *inner product* (or *dot product*) is

$$\langle u, w \rangle = \sum_{i=1}^{d} u_i w_i.$$

Note that $\|v\|_2^2 = \langle v, v \rangle$.

Geometric picture (in $\mathbb{R}^2$). Take $d = 2$, so a vector $v = (v_1, v_2)$ is a point in the plane. The figure below plots three sets of such points: the **diamond** (red) is $\{v : \|v\|_1 = 1\}$, i.e. all v whose ℓ_1 norm is exactly 1. Its vertices are at $(\pm 1, 0)$ and $(0, \pm 1)$. The **solid blue circle** is $\{v : \|v\|_2 = 1\}$, and the **dashed blue circle** is $\{v : \|v\|_2 = \frac{1}{\sqrt{2}}\}$. Two specific points on the diamond are marked: a vertex $v = (1, 0)$ (where $\|v\|_2 = 1$) and an edge midpoint $v = (\frac{1}{2}, \frac{1}{2})$ (where $\|v\|_2 = \frac{1}{\sqrt{2}}$).
So every point v on the diamond $\{\|v\|_1 = 1\}$ has $\frac{1}{\sqrt{2}} \le \|v\|_2 \le 1$. Scaling gives $\|v\|_2 \le \|v\|_1 \le \sqrt{2}\,\|v\|_2$ in $\mathbb{R}^2$, which is the case $d = 2$ of the inequality you will prove in part (b).

(a) Compute $\|v\|_1$ and $\|v\|_2$ for $v = (3, -4, 0, 1)$.

(b) Show that for every $v \in \mathbb{R}^d$, $\|v\|_2 \le \|v\|_1 \le \sqrt{d}\,\|v\|_2$. *Hint:* the right inequality follows from the *Cauchy–Schwarz inequality*, which states that for any two vectors $u, w \in \mathbb{R}^d$, $|\langle u, w \rangle| \le \|u\|_2 \cdot \|w\|_2$. Apply it with $u = |v| = (|v_1|, \ldots, |v_d|)$ and w the all-ones vector.

(c) Suppose $f : \mathbb{R}^n \to \mathbb{R}^d$ has the property that $\|f(x) - f(x')\|_1 \le \Delta$ for all "neighboring" inputs x, x'. Using (b), give an upper bound on $\|f(x) - f(x')\|_2$ in terms of Δ and d.

Problem 7.

Recap. For a single-variable function $f : \mathbb{R} \to \mathbb{R}$, the *derivative* $f'(w)$ measures the rate of change of f at w (the slope of the tangent line). At a minimum or maximum of a differentiable function, $f'(w) = 0$. The second derivative $f''(w)$ tells you about curvature: if $f''(w) > 0$, then w is a local minimum.

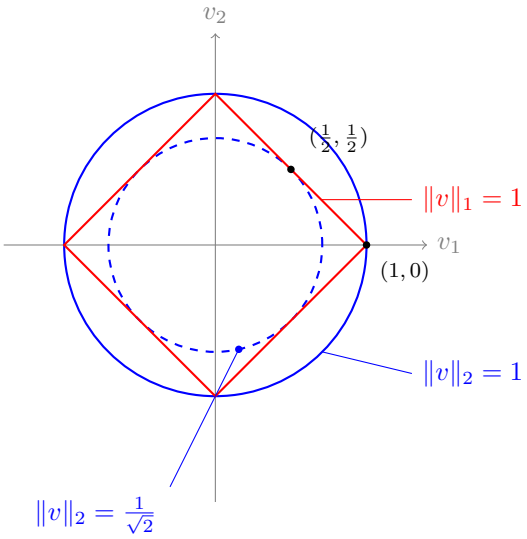

Figure 1: Unit "balls" of the ℓ_1 and ℓ_2 norms in $\mathbb{R}^2$. The ℓ_1 unit set (red diamond) is sandwiched between the ℓ_2-circle of radius $\frac{1}{\sqrt{2}}$ (inscribed, touching the edge midpoints of the diamond) and the ℓ_2-circle of radius 1 (circumscribed, touching the vertices of the diamond).

For a multivariate function $L : \mathbb{R}^d \to \mathbb{R}$, the *partial derivative* $\partial L/\partial w_j$ is the derivative with respect to w_j while treating the other coordinates as constants. The *gradient* $\nabla L(w) \in \mathbb{R}^d$ is the vector of partial derivatives:

$$\nabla L(w) = \left(\frac{\partial L}{\partial w_1}(w),\ \frac{\partial L}{\partial w_2}(w),\ \ldots,\ \frac{\partial L}{\partial w_d}(w) \right).$$

The *chain rule* for differentiation says: if $g(w) = h(r(w))$ where r is scalar-valued, then $g'(w) = h'(r(w)) \cdot r'(w)$. In particular, $\frac{d}{dw}(r(w)^2) = 2\, r(w)\, r'(w)$.

Throughout the problem, $x_1, \ldots, x_n \in \mathbb{R}$ are fixed real numbers (the data).

(a) Define $L : \mathbb{R} \to \mathbb{R}$ by $L(w) = \frac{1}{n} \sum_{i=1}^{n} (w - x_i)^2$. Compute $L'(w)$, find the unique critical point, and verify (via the second derivative) that it is a minimum. What familiar statistic of the data does the minimizer equal?

(b) Now suppose each data point is a pair (x_i, y_i) with $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, and recall the inner product $\langle w, x_i \rangle = \sum_{j=1}^{d} w_j (x_i)_j$ from Problem 6. As a function of w (with x_i fixed), $\langle w, x_i \rangle$ is linear, and its gradient is $\nabla_w \langle w, x_i \rangle = x_i$. Define $L : \mathbb{R}^d \to \mathbb{R}$ by

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} \big(\langle w, x_i \rangle - y_i \big)^2.$$

Using the chain rule, compute $\nabla L(w) \in \mathbb{R}^d$. Express your answer as a sum over i.

(c) Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(u, v) = u^2 v + e^v$. Compute $\partial f/\partial u$ and $\partial f/\partial v$, and write down $\nabla f(u, v)$.

Problem 8.

This problem checks comfort with the logical structure used throughout the course. The definition of differential privacy reads:

A randomized algorithm $\mathcal{M}$ is ε-differentially private if, ***for all*** *neighboring datasets D, D' and* ***for all*** *events $E \subseteq \text{Range}(\mathcal{M})$:*

$$\mathbb{P}[\mathcal{M}(D) \in E] \;\leq\; e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(D') \in E].$$

(a) *Proving a "for all" statement.* Prove: for every integer $n \geq 1$, $n^2 - n$ is even. Begin your proof with "Let $n \geq 1$ be an arbitrary integer" or equivalent — be explicit that you are not assuming anything specific about n.

(b) *Disproving a "for all" statement.* Consider the (false) claim: *for every integer $n \geq 1$, $n^2 - n + 1$ is prime.* Disprove it. State explicitly what kind of object suffices to disprove a "for all" statement.

(c) *Negating a nested quantifier.* Write the negation of the following statement, in symbols and in plain English:

$$\forall x \in \mathbb{R},\ \exists y \in \mathbb{R} \;:\; x + y > 0.$$

(Recall $\neg(\forall x : P(x))$ is $\exists x : \neg P(x)$, and $\neg(\exists y : Q(y))$ is $\forall y : \neg Q(y)$.)

(d) *Applying it to DP.* Suppose someone claims that an algorithm $\mathcal{M}$ is ε-DP. To **disprove** their claim, what is the smallest amount of information you need to exhibit? Pick one and justify in one sentence.

(i) A single dataset D and a single event E such that $\mathbb{P}[\mathcal{M}(D) \in E] > e^{\varepsilon}$.

(ii) A pair of neighboring datasets D, D' and a single event E such that $\mathbb{P}[\mathcal{M}(D) \in E] > e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(D') \in E]$.

(iii) All neighboring pairs D, D' and all events E where the inequality fails.

Problem 9.

This problem checks two proof techniques used throughout the course: *case analysis* and *WLOG* ("without loss of generality"). The triangle inequality is invoked directly in the privacy proof of the Laplace mechanism (Lecture 5–6).

(a) *Case analysis.* Prove the *triangle inequality* for real numbers: for all $x, y \in \mathbb{R}$,

$$|x + y| \leq |x| + |y|.$$

Hint: Recall that $|a| = a$ if $a \geq 0$ and $|a| = -a$ if $a < 0$. Split into cases based on the signs of x and y. Be explicit about which case you are in at each step.

(b) *WLOG / symmetry.* Prove the *reverse triangle inequality*: for all $x, y \in \mathbb{R}$,

$$\big|\, |x| - |y| \,\big| \leq |x - y|.$$

Hint: You may begin "WLOG, assume $|x| \geq |y|$." Then write $x = (x - y) + y$ and apply part (a).

(c) *Conceptual.* In part (b) we said "WLOG, assume $|x| \geq |y|$." In one or two sentences, explain why this assumption is valid here — i.e., what symmetry of the statement justifies it.

Self-scoring rubric

Use this guide to interpret your performance and decide whether to take the course.

What "doing it cleanly" looks like.

- You worked through each problem on your own without looking up formulas mid-problem.
- Your computations are correct and you can articulate *why* each step is valid (especially: where independence is used, what Bayes' rule is doing, what makes a "for all" proof valid).
- You finished in roughly 90 minutes to 2 hours.

Mapping to the course.

- **Problem 1 (sum/product notation).** Used in essentially every proof and computation in the course. If any of (a)–(d) was hard, practice manipulating $\sum$ and $\prod$ before the semester starts — this is foundational.
- **Problems 2, 3 (probability).** Used in every lecture. If these were hard, the entire course will be hard — consider reviewing a basic probability text (e.g. Blitzstein & Hwang, *Introduction to Probability*, Chs. 1–4) before the semester.
- **Problem 4 (continuous distributions).** The Laplace mechanism (Lecture 5–6) and Gaussian mechanism (Lecture 11–12) are central. You won't need to integrate these by hand in the course, but you should be comfortable with what a pdf means and how scale parameters affect the spread of a distribution. If parts (a)–(c) were hard, review the basics of continuous probability before the semester.
- **Problem 5 (exp/log).** Fluency with e^x and ln as inverse operations is used throughout the course (DP is defined using e^ε, but the equivalent log form — privacy loss — shows up in many proofs). If part (d) was hard, brush up on the algebra of logs and exponentials before the semester starts.
- **Problem 6 (norms, linear algebra).** Sensitivity (Lec. 5–6) is defined via ℓ_1; Gaussian mechanism uses ℓ_2. If part (b) or (c) was hard, brush up on basic norm inequalities.
- **Problem 7 (multivariable calculus).** Used in Lecture 14+ (DP-SGD, ERM). Lighter usage than the others; comfort with gradients is sufficient.
- **Problem 8 (quantifiers).** The DP definition is a triple-quantifier statement, and you will read and write proofs in this form constantly. If 8(a) or 8(d) were hard, that is a strong signal to spend time with a "transition to proof" text (e.g. Velleman, *How To Prove It*).
- **Problem 9 (case analysis, WLOG).** The triangle inequality is invoked verbatim in the Laplace mechanism privacy proof, and WLOG / symmetry arguments appear in nearly every two-sided ratio proof. If part (a) was hard, practice writing case-analysis proofs (sign cases on $\mathbb{R}$ are typical). If part (c) was hard, review what makes a statement symmetric in its variables.

The course assumes comfort with all of the above. If any of the problems above gave you significant trouble, plan to brush up on that area before the semester begins.