Artificial Intelligence PhD Qualifying Examination

High-Fidelity Mock Examination 07

Choose three of four sections; maximum score: 100 points.

Name: _____ **Student ID:** _____ **[1 point]**

Instructions. This examination consists of four sections, each worth 33 points. Choose **three** sections and indicate your choices clearly at the beginning of your answer sheet. If you answer more than three sections, only the first three sections that you select will be graded. The additional point above is awarded for writing your name and student ID. Give complete reasoning unless a question states otherwise.

A. Machine Learning Theory [33 points]

8 points Define k-nearest-neighbor classification, including a deterministic tie rule. Explain its locality inductive bias, the usual effect of increasing k on bias and variance, and the storage/training/naive-query complexity for m points in $\mathbb{R}^d$. State why normalization and metric choice must be fitted without test data.

7 points Let $\mathcal{H}$ be finite with $|\mathcal{H}| = N$. You may use, for every h and $\lambda > 0$,

$$\mathbb{E}e^{\lambda(L(h)-\widehat{L}(h))} \le e^{\lambda^2/(8m)}, \qquad \mathbb{E}e^{\lambda(\widehat{L}(h)-L(h))} \le e^{\lambda^2/(8m)}.$$

Prove

$$\mathbb{E}\sup_{h\in\mathcal{H}} |L(h) - \widehat{L}(h)| \le \sqrt{\frac{\log(2N)}{2m}},$$

and derive an expected excess-risk bound for empirical-risk minimization.

8 points On $\{0,1\}^d$, for $w \in \{0,1\}^d$ define the parity classifier

$$h_w(x) = \left(\sum_{j=1}^{d} w_j x_j\right) \bmod 2.$$

Determine the VC dimension of the parity class.

6 points Let F be twice differentiable with $\mu I \preceq \nabla^2 F(x) \preceq LI$ for all x, $0 < \mu \le L$. For gradient descent $T(x) = x - \eta\nabla F(x)$ with $\eta = 2/(L+\mu)$, prove

$$\|T(x) - T(y)\|_2 \le \frac{L-\mu}{L+\mu}\,\|x-y\|_2\,.$$

You may use the integral mean-Hessian representation of the gradient difference.

4 points Independent losses $Z_i \in [0,1]$ may have different means. For their average $\widehat{\mu} = m^{-1}\sum_i Z_i$ and $\bar{\mu} = m^{-1}\sum_i \mathbb{E}Z_i$, use only variance and Chebyshev's inequality to give a sufficient m for $|\widehat{\mu} - \bar{\mu}| \le \varepsilon$ with probability at least $1-\delta$.

B. Deep Learning and Reinforcement Learning [33 points]

(a) **Masked multi-head attention.** [10 points]

Let $X \in \mathbb{R}^{L\times d}$ contain token rows. There are h heads with $d_k = d_v = d/h$. For head j, define

$$Q_j = XW_j^Q, \qquad K_j = XW_j^K, \qquad V_j = XW_j^V,$$
$$A_j = \mathrm{softmax}_{\mathrm{row}}\left(\frac{Q_j K_j^\top}{\sqrt{d_k}} + M\right).$$

The mask has entries 0 for allowed keys and $-\infty$ for forbidden keys.

(a) [3 points] Give the dimensions of all projections and intermediate tensors, including the concatenated heads and output projection.

(b) [3 points] Prove that every unmasked row of A_j is a probability vector and that the corresponding output row is a convex combination of allowed value rows. Explain what goes wrong if every key in a row is masked.

(c) [2 points] Under an independent unit-variance component heuristic, compute the variance scale of an unscaled dot product and explain the factor $1/\sqrt{d_k}$.

(d) [2 points] Give the dominant time and attention-memory complexity and one architectural tradeoff of using more heads at fixed d.

(b) **An affine-coupling normalizing flow.** [13 points]

Split $x = (x_A, x_B)$ and define a base-to-data transformation

$$y_A = x_A, \qquad y_B = x_B \odot \exp(s_\theta(x_A)) + t_\theta(x_A).$$

The functions s_θ, t_θ need not be invertible.

(a) [3 points] Derive the inverse transformation and state why it exists.

(b) [4 points] Show that the Jacobian is block triangular and compute its log-determinant.

(c) [3 points] For a composition of such layers mapping z to data x, write the exact log-density and dataset-average NLL, with the correct determinant sign.

(d) [2 points] Explain why using the same split in every layer limits expressivity and give a remedy.

(e) [1 point] Give one numerical-stability precaution for the scale output.

(c) **DQN overestimation and Double DQN.** [10 points]

For a transition (s, a, r, s', d), $d = 1$ means terminal. Let Q_ω be the online network and $Q_{\bar{\omega}}$ a target network.

(a) [3 points] Write a squared DQN loss and target, including terminal handling and the stop-gradient convention.

(b) [3 points] Suppose $\widehat{Q}_a$ are noisy estimates with $\mathbb{E}[\widehat{Q}_a] = Q_a$. Prove $\mathbb{E}[\max_a \widehat{Q}_a] \geq \max_a Q_a$ and explain when the inequality can be strict.

(c) [2 points] Write the Double-DQN target and identify which network selects and which evaluates the next action.

(d) [2 points] Explain the distinct roles of experience replay and the target network.

C. Optimization Methods in Artificial Intelligence [33 points]

Let $X \subseteq \mathbb{R}^d$ be nonempty, closed, and convex, with Euclidean diameter at most $D > 0$: $\|x - y\| \le D$ for all $x, y \in X$. At round $t = 1, \dots, T$, the learner chooses $x^t \in X$, observes a convex loss f_t, and receives $g^t \in \partial f_t(x^t)$. Fix $\delta > 0$ and define

$$h_t = \sqrt{\delta^2 + \sum_{\tau=1}^{t} \|g^\tau\|^2}.$$

The scalar AdaGrad update is

$$x^{t+1} = P_X\left(x^t - \frac{\eta}{h_t} g^t\right), \qquad \eta > 0.$$

(1) [3 points] Identify the effective learning rate. Explain why $\delta > 0$ is needed and why h_t is nondecreasing.

(2) [5 points] For any comparator $u \in X$, use nonexpansiveness of projection to prove

$$\langle g^t, x^t - u \rangle \le \frac{h_t}{2\eta}\left(\|x^t - u\|^2 - \|x^{t+1} - u\|^2\right) + \frac{\eta}{2h_t}\|g^t\|^2.$$

(3) [5 points] Prove the changing-weight telescope

$$\sum_{t=1}^{T} h_t\left(\|x^t - u\|^2 - \|x^{t+1} - u\|^2\right) \le D^2 h_T.$$

(4) [4 points] With $h_0 = \delta$, prove

$$\sum_{t=1}^{T} \frac{\|g^t\|^2}{h_t} \le 2(h_T - \delta) \le 2h_T.$$

(5) [6 points] Define

$$\mathrm{Regret}_T(u) = \sum_{t=1}^{T} [f_t(x^t) - f_t(u)].$$

Use parts (2)–(4) to show

$$\mathrm{Regret}_T(u) \le \left(\frac{D^2}{2\eta} + \eta\right)\sqrt{\delta^2 + \sum_{t=1}^{T} \|g^t\|^2}.$$

(6) [5 points] Optimize the coefficient over η. If $\|g^t\| \le G$, give an average-regret bound. When all losses equal one convex function F, convert it into a function-value guarantee for $\bar{x}^T = T^{-1}\sum_{t=1}^{T} x^t$.

(7) [5 points] Explain three limitations: why a regret bound for time-varying losses is not automatically iterate convergence; why an early large gradient can make later steps too small; and why being adaptive does not remove the need to choose η and δ.

D. Natural Language Processing [33 points]

1. Learning a reward model from preferences. [8 points] For a prompt x, a human prefers y^+ to y^-. A reward model uses the Bradley–Terry probability

$$p_\psi(y^+ \succ y^- \mid x) = \sigma\big(r_\psi(x, y^+) - r_\psi(x, y^-)\big).$$

4 points Write the negative log-likelihood and compute its derivatives with respect to the two scalar rewards.

2 points Show that adding an arbitrary prompt-dependent constant $c(x)$ to all rewards does not change the preference likelihood. What is identifiable?

2 points Give one data-collection design that reduces position or annotator bias.

2. KL-regularized policy optimization. [9 points] For a fixed prompt x and a finite response space, consider

$$\max_{\pi} \sum_{y} \pi(y \mid x) r(x, y) - \beta\, \mathrm{KL}\big(\pi(\cdot \mid x) \| \pi_{\mathrm{ref}}(\cdot \mid x)\big), \qquad \beta > 0.$$

Assume $\pi_{\mathrm{ref}}(y \mid x) > 0$ for every candidate response.

6 points Derive the optimal normalized policy $\pi^*(y \mid x)$.

3 points Explain the limits $\beta \to 0$ and $\beta \to \infty$, and why the support assumption matters.

3. Noisy preference labels. [6 points] Let the true probability that y_1 is preferred to y_2 be p. Independently, an observed binary preference is flipped with probability $\eta < 1/2$.

3 points Derive the observed preference probability $\widetilde{p}$.

3 points Explain how increasing η affects learnability and confidence, and give one way to estimate or mitigate the noise.

4. Auditing an aligned language model. [10 points] An instruction-following model has been optimized against a learned reward model. Design an evaluation and monitoring protocol that can distinguish genuine preference improvement from reward hacking.

4 points Specify offline datasets, baselines, and at least three metric families.

2 points Design a test for distribution shift and a test for loss of response diversity.

2 points Explain how human evaluation should be randomized and blinded.

2 points State two deployment triggers for rollback or abstention, including how uncertainty enters the decision.