Homework 0: Mathematical Readiness

CMPUT 654, Fall 2026

Purpose. This assignment is not graded. It samples the proof writing, probability, linear algebra, calculus, and abstraction used in the course. Attempt it during the first week without searching for solutions. If substantial parts remain inaccessible after reviewing the relevant prerequisites, speak with the instructor promptly and make an honest decision about whether the course is a good fit.

Submission. No submission is required unless the instructor announces otherwise. Students who would like diagnostic feedback may submit a typed PDF by the end of Week 1. Use the course homework template, acknowledge every source and collaborator, and write every argument in your own words.

Total. 100 diagnostic points. The points indicate intended effort, not a course grade.

Problem 1 (Log loss identifies a conditional probability — 25 points)**.** Let $Y \sim \mathrm{Bernoulli}(\eta)$, where $\eta \in (0,1)$. For a prediction $q \in (0,1)$, define the log loss

$$\ell(q;Y) = -Y\log q - (1-Y)\log(1-q)$$

and its population risk $L_\eta(q) = \mathbb{E}[\ell(q;Y)]$.

(a) Compute $L'_\eta(q)$ and $L''_\eta(q)$, and prove that $q = \eta$ is its unique minimizer. [7]

(b) Show that

$$L_\eta(q) - L_\eta(\eta) = \eta \log\frac{\eta}{q} + (1-\eta)\log\frac{1-\eta}{1-q}.$$

Identify the expression on the right. [6]

(c) You may use the Bernoulli form of Pinsker's inequality, $D_{\mathrm{KL}}(\mathrm{Ber}(\eta)\|\,\mathrm{Ber}(q)) \ge 2(\eta-q)^2$. Deduce a bound on $|q-\eta|$ from an upper bound $L_\eta(q) - L_\eta(\eta) \le \varepsilon$. [5]

(d) A classifier predicts 1 when $q \ge 1/2$ and 0 otherwise. Find the Bayes-optimal classifier under zero–one loss and explain why a small probability-estimation error need not guarantee the correct classification when η is close to $1/2$. [7]

Problem 2 (Uniform estimation and selection — 20 points)**.** For every $j \in [m]$, let $Z_{j,1}, \ldots, Z_{j,n}$ be independent random variables taking values in $[0,1]$, with common mean μ_j for fixed j. Let $\widehat{\mu}_j = n^{-1}\sum_{i=1}^n Z_{j,i}$. You may use Hoeffding's inequality

$$\mathbb{P}(|\widehat{\mu}_j - \mu_j| \ge \varepsilon) \le 2e^{-2n\varepsilon^2}.$$

(a) Prove that, with probability at least $1-\delta$, $\max_{j\in[m]} |\widehat{\mu}_j - \mu_j| \le \varepsilon$, provided

$$n \ge \frac{\log(2m/\delta)}{2\varepsilon^2}.$$

State exactly where independence is used and where it is not used. [10]

(b) Let $\widehat{j} \in \arg\max_j \widehat{\mu}_j$ and $j^\star \in \arg\max_j \mu_j$. On the event above, prove $\mu_{j^\star} - \mu_{\widehat{j}} \le 2\varepsilon$. [10]

Problem 3 (Which interpolator does gradient descent select? — 30 points)**.** Let $A \in \mathbb{R}^{n\times d}$ have full row rank, with $n < d$, and let $b \in \mathbb{R}^n$. Consider

$$f(w) = \frac{1}{2}\|Aw - b\|_2^2$$

and gradient descent

$$w_{t+1} = w_t - \gamma A^\top(Aw_t - b), \qquad w_0 = 0.$$

(a) Show that $w^\dagger = A^\top(AA^\top)^{-1}b$ interpolates the data and is the unique minimum-Euclidean-norm vector satisfying $Aw = b$. [10]

(b) Prove that every iterate w_t lies in the row space of A. [5]

(c) Let $r_t = Aw_t - b$. Show $r_{t+1} = (I - \gamma AA^\top)r_t$. Prove that $r_t \to 0$ whenever $0 < \gamma < 2/\|A\|_2^2$. [10]

(d) Conclude that $w_t \to w^\dagger$. What can change if w_0 has a nonzero component in the null space of A? [5]

Problem 4 (A small query-complexity lower bound — 25 points)**.** An unknown index $i^\star \in [k]$ is selected. On query $j \in [k]$, a learner observes 1 if $j = i^\star$ and 0 otherwise. The learner must eventually identify $i^\star$ with probability one. It may choose queries adaptively and may randomize.

(a) Give a deterministic learner making at most $k - 1$ queries on every instance, and prove it is correct. [6]

(b) Prove that every deterministic learner requires at least $k - 1$ queries on some instance. [6]

(c) Consider querying indices in a uniformly random order and stopping after finding $i^\star$, or after eliminating $k - 1$ indices. Show that for every $i^\star$, its expected number of queries is

$$\frac{k+1}{2} - \frac{1}{k}.$$

[6]

(d) Use Yao's minimax principle with a uniform random $i^\star$ to prove that every randomized learner has worst-case expected query complexity at least $(k + 1)/2 - 1/k$. State the version of Yao's principle you use. [7]