Assignment 1

00103335: Deep Learning and Reinforcement Learning
Fall 2026

Problem 1. Let $m(x) := \mathbb{E}[Y \mid X = x]$, and let $\phi_1, \ldots, \phi_Q$ form an orthonormal basis for a Q-dimensional subspace of $L_2(P_X)$. In particular, each basis function has unit norm and

$$\mathbb{E}_X[\phi_j(X)\phi_k(X)] = \delta_{jk}, \qquad 1 \le j, k \le Q.$$

Suppose

$$m(x) = \sum_{j=1}^{Q} \theta_j \phi_j(x), \qquad \mathcal{F}_q = \operatorname{span}\{\phi_1, \ldots, \phi_q\}, \qquad \mathcal{F}_0 = \{0\}, \qquad 0 \le q \le Q.$$

We assume the estimated coefficient to be

$$Z_j = \theta_j + \frac{\nu}{\sqrt{n}}\xi_j, \qquad \xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \qquad \nu > 0, \qquad \widehat{f}_q = \sum_{j=1}^{q} Z_j \phi_j.$$

Define

$$B_q^2 := \left\| \mathbb{E}_S[\widehat{f}_q] - m \right\|_{L_2(P_X)}^2, \qquad V_q := \mathbb{E}_X[\operatorname{Var}_S\{\widehat{f}_q(X)\}].$$

(a) Compute

$$f_q^\star \in \operatorname*{arg\,min}_{f \in \mathcal{F}_q} \|f - m\|_{L_2(P_X)}^2.$$

(b) Verify that $\mathbb{E}_S[\widehat{f}_q] = f_q^\star$, and show that B_q^2 is nonincreasing in q.

(c) Verify that V_q is nondecreasing in q.

Problem 2. Let $d > n$, let $\mathbf{X} \in \mathbb{R}^{n \times d}$ have full row rank, and consider

$$\widehat{L}(b) = \frac{1}{2n}\|\mathbf{y} - \mathbf{X}b\|_2^2, \qquad b_0 = 0, \qquad b_{t+1} = b_t - \eta \nabla \widehat{L}(b_t).$$

(a) Compute $\nabla \widehat{L}(b)$ and verify that

$$b_{t+1} = b_t + \frac{\eta}{n}\mathbf{X}^\top(\mathbf{y} - \mathbf{X}b_t).$$

Show also that the Hessian is singular.

(b) Define $r_t := \mathbf{y} - \mathbf{X}b_t$. Show that

$$r_{t+1} = \left(\mathbf{I}_n - \frac{\eta}{n}\mathbf{X}\mathbf{X}^\top\right) r_t, \qquad r_0 = \mathbf{y}.$$

Let $\lambda_1, \ldots, \lambda_n > 0$ be the eigenvalues of $\mathbf{X}\mathbf{X}^\top$. Use its spectral decomposition to prove $r_t \to 0$ whenever

$$0 < \eta < \frac{2n}{\lambda_{\max}(\mathbf{X}\mathbf{X}^\top)} = \frac{2n}{\|\mathbf{X}\|_{\text{op}}^2}.$$

More precisely, prove

$$\|r_t\|_2 \le \rho^t \|\mathbf{y}\|_2, \qquad \rho := \max_i \left|1 - \frac{\eta\lambda_i}{n}\right| < 1.$$

(c) Sum the updates and use a matrix geometric series to derive

$$b_t = \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\left[\mathbf{I}_n - \left(\mathbf{I}_n - \frac{\eta}{n}\mathbf{X}\mathbf{X}^\top\right)^t\right]\mathbf{y}.$$

Deduce that

$$b_t \longrightarrow b^\dagger := \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{y}.$$

(d) Verify that $\mathbf{X}b^\dagger = \mathbf{y}$. If b is any other interpolator, show that $b = b^\dagger + v$ for some $v \in \ker(\mathbf{X})$. Prove

$$\|b\|_2^2 = \left\|b^\dagger\right\|_2^2 + \|v\|_2^2.$$

Hence show that $b^\dagger$ is the unique minimum-norm interpolator.

(e) Suppose instead that b_0 is arbitrary. Show that

$$b_t \longrightarrow b^\dagger + \Pi_{\ker(\mathbf{X})} b_0.$$

Problem 3. Let $d < n$, let $\mathbf{X} \in \mathbb{R}^{n\times d}$ have full column rank, and define its hat matrix by

$$\mathbf{H} := \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top.$$

(a) Verify directly that $\mathbf{H}^\top = \mathbf{H}$ and $\mathbf{H}^2 = \mathbf{H}$.

(b) Show that $\mathbf{H}z = z$ for every $z \in \operatorname{range}(\mathbf{X})$, and that $\mathbf{H}z = 0$ if and only if $z \in \ker(\mathbf{X}^\top)$. Deduce that $\mathbf{H}$ is the orthogonal projector onto $\operatorname{range}(\mathbf{X})$, while $\mathbf{I}_n - \mathbf{H}$ projects onto $\ker(\mathbf{X}^\top)$.

(c) Determine the eigenvalues of $\mathbf{H}$ and their multiplicities. Conclude that

$$\operatorname{rank}(\mathbf{H}) = \operatorname{tr}(\mathbf{H}) = d, \qquad \operatorname{tr}(\mathbf{I}_n - \mathbf{H}) = n - d.$$

If h_{ii} are the diagonal entries of $\mathbf{H}$, also show that $0 \le h_{ii} \le 1$ and $\sum_i h_{ii} = d$.

(d) For ordinary least squares, show that the fitted response and residual are $\widehat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ and $r = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$. Prove $\mathbf{X}^\top r = 0$ and the Pythagorean identity

$$\|\mathbf{y}\|_2^2 = \|\mathbf{H}\mathbf{y}\|_2^2 + \|(\mathbf{I}_n - \mathbf{H})\mathbf{y}\|_2^2.$$

Express the minimized empirical loss $\frac{1}{2n}\|\mathbf{y} - \mathbf{X}b\|_2^2$ in terms of this residual.

(e) Now suppose $\mathbf{y} = \mathbf{X}\beta + \eta$, where, conditionally on $\mathbf{X}$, $\mathbb{E}[\eta \mid \mathbf{X}] = 0$ and $\mathbb{E}[\eta\eta^\top \mid \mathbf{X}] = \tau^2\mathbf{I}_n$. First show that $(\mathbf{I}_n - \mathbf{H})\mathbf{X}\beta = 0$. Prove

$$\mathbb{E}\Big[\|(\mathbf{I}_n - \mathbf{H})\mathbf{y}\|_2^2 \mid \mathbf{X}\Big] = \tau^2(n-d).$$

Problem 4. Let m, p be positive integers with $m > p + 1$, let $\mathbf{Z} = [z_1, \ldots, z_p] \in \mathbb{R}^{m\times p}$ have independent standard Gaussian entries, and set $\mathbf{W} := \mathbf{Z}^\top\mathbf{Z}$.

(a) Show that $\mathbf{W}$ is positive definite almost surely.

(b) Fix j, let $\mathbf{Z}_{-j}$ contain all columns except z_j, and define

$$\mathbf{P}_{-j} := \mathbf{Z}_{-j}(\mathbf{Z}_{-j}^\top\mathbf{Z}_{-j})^{-1}\mathbf{Z}_{-j}^\top.$$

For $p = 1$, use the convention $\mathbf{P}_{-j} = 0$. Apply the Schur-complement formula to prove

$$\{[\mathbf{W}^{-1}]_{jj}\}^{-1} = z_j^\top(\mathbf{I}_m - \mathbf{P}_{-j})z_j.$$

(c) Show that

$$z_j^\top(\mathbf{I}_m - \mathbf{P}_{-j})z_j \mid \mathbf{Z}_{-j} \sim \chi^2_{m-p+1}.$$

(d) If $U \sim \chi^2_k$ and its density satisfies

$$f_U(u) = \frac{u^{k/2-1}e^{-u/2}}{2^{k/2}\Gamma(k/2)}, \qquad u > 0,$$

show that

$$\mathbb{E}[U^{-1}] = \begin{cases} 1/(k-2), & k > 2, \\ +\infty, & k \le 2. \end{cases}$$

Conclude that $\mathbb{E}[[\mathbf{W}^{-1}]_{jj}] = 1/(m-p-1)$.

(e) Show that

$$\mathbb{E}[[\mathbf{W}^{-1}]_{rs}] = 0, \qquad r \neq s,$$

and conclude that

$$\mathbb{E}[\mathbf{W}^{-1}] = \frac{\mathbf{I}_p}{m - p - 1}.$$