DSC 190/291 · Assignment 8

UCSD · Spring 2026

Released: Tuesday, May 26 · Due: Tuesday, June 2, 11:59 PM

AI policy. AI assistance is allowed and encouraged in this course. You may use AI to learn the material, explore proof structure, test examples, debug code or formalizations, and improve exposition. However, you are responsible for checking correctness and for standing behind every proof step, derivation, formalization, experiment, and explanation you submit. Use AI as a collaborator, not as an oracle: do not submit anything you cannot explain and verify. The AI usage report is a required component of the assignment.

Submission. Submit a single PDF on Gradescope containing your write-up, figures, and discussion. Also place any supporting artifacts for the assignment in your course repository under the appropriate assignment directory. This may include code, Lean files, notebooks, scripts, data, or other materials needed to inspect or reproduce your work. Your submission should make it clear how the repository artifacts relate to the write-up.

Part A: Choosing Regularization by Validation (40 points)

Let $\mathcal{W}$ be a convex parameter domain, and let $\ell(w, z) \in [0, 1]$ be convex in w and G-Lipschitz with respect to a norm $\|\cdot\|$:

$$|\ell(w, z) - \ell(w', z)| \le G\|w - w'\|$$

for all w, w', z. Let Ψ be nonnegative and α-strongly convex with respect to $\|\cdot\|$.

Split an i.i.d. sample into an independent training sample T of size n_T and validation sample V of size n_V. For a finite grid

$$\Lambda = \{\lambda_1, \dots, \lambda_K\} \subset (0, \infty),$$

define, for each $\lambda \in \Lambda$,

$$h_\lambda = A_\lambda(T) \in \arg\min_{w \in \mathcal{W}} L_T(w) + \lambda\Psi(w).$$

Then choose

$$\hat{\lambda} \in \arg\min_{\lambda \in \Lambda} L_V(h_\lambda),$$

and output $h_{\hat{\lambda}}$.

1. **(10 points) Validation selection for a finite random candidate set.**

   Condition on predictors $h_1, \dots, h_K$ that are independent of a validation set V of size n_V, and let

   $$\hat{k} \in \arg\min_{k \in [K]} L_V(h_k).$$

   Prove

$$\mathbb{E}_V L_{\mathcal{D}}(h_{\hat{k}}) \le \min_{k\in[K]} L_{\mathcal{D}}(h_k) + 2\sqrt{\frac{\log(2K)}{2n_V}}.$$

You may use Hoeffding's lemma for $[0,1]$-valued losses.

2. **(10 points) Oracle inequality over a grid.**

   Apply the previous part conditionally on the training sample T, and then take expectation over T. Prove

$$\mathbb{E}L_{\mathcal{D}}(h_{\hat{\lambda}}) \le \min_{\lambda\in\Lambda} \inf_{u\in\mathcal{W}} \left\{ L_{\mathcal{D}}(u) + \lambda\Psi(u) + \frac{2G^2}{\lambda\alpha n_T} \right\} + 2\sqrt{\frac{\log(2K)}{2n_V}}.$$

   Explain why selecting among K trained RERM rules contributes a validation term of order $\sqrt{\frac{\log K}{n_V}}$.

3. **(12 points) Adapting to an unknown comparator scale.**

   For a comparator $u \in \mathcal{W}$, write $a = \Psi(u)$ and define

$$\lambda_{\text{opt}}(a) = \sqrt{\frac{2G^2}{\alpha a n_T}}$$

   whenever $a > 0$.

   Suppose $0 < B_{\min} \le B_{\max}$ and the grid Λ has the following property: for every $a \in [B_{\min}^2, B_{\max}^2]$, there is a $\lambda \in \Lambda$ satisfying

$$\frac{\lambda_{\text{opt}}(a)}{2} \le \lambda \le 2\lambda_{\text{opt}}(a).$$

   Prove that for every $u \in \mathcal{W}$ with $B_{\min}^2 \le \Psi(u) \le B_{\max}^2$,

$$\mathbb{E}L_{\mathcal{D}}(h_{\hat{\lambda}}) \le L_{\mathcal{D}}(u) + \frac{5\sqrt{2}}{2} G\sqrt{\frac{\Psi(u)}{\alpha n_T}} + 2\sqrt{\frac{\log(2K)}{2n_V}}.$$

   Your proof should explicitly optimize the two-term expression

$$\lambda\Psi(u) + \frac{2G^2}{\lambda\alpha n_T}$$

   up to the factor-2 discretization of λ.

4. **(8 points) A dyadic grid and the price of tuning.**

   Construct a dyadic grid Λ satisfying the property in the previous part for all $a \in [B_{\min}^2, B_{\max}^2]$, where $0 < B_{\min} \le B_{\max}$. State the number of grid points K as a function of $\frac{B_{\max}}{B_{\min}}$.

   Assume n is even, set $n_T = n_V = \frac{n}{2}$, and simplify the resulting guarantee. Identify the two statistical terms in the bound: the RERM learning term and the validation-selection term. Identify the additional term caused by not knowing the comparator scale $\Psi(u)$ in advance.

Part B: Approximate RERM and Optimization Error (35 points)

Let $\mathcal{W}$ be a convex parameter domain. Assume $\ell(w, z)$ is convex and G-Lipschitz with respect to $\|\cdot\|$, and Ψ is nonnegative and α-strongly convex with respect to $\|\cdot\|$. For $\lambda > 0$ and a sample S of size n, define

$$F_S(w) = L_S(w) + \lambda\Psi(w), \quad w_S \in \arg\min_{w\in\mathcal{W}} F_S(w).$$

Let the actual algorithm return $\tilde{w}_S \in \mathcal{W}$ satisfying the deterministic optimization guarantee

$$F_S(\tilde{w}_S) \leq F_S(w_S) + \eta$$

for every sample S, where $\eta \geq 0$.

You may use the exact RERM movement bound from Week 8: if S and S' differ in one example, then

$$\|w_S - w_{S'}\| \leq \frac{2G}{\lambda\alpha n}.$$

1. **(7 points) Approximate minimizers are close to exact minimizers.**

   Prove that for every sample S,

   $$\|\tilde{w}_S - w_S\| \leq \sqrt{\frac{2\eta}{\lambda\alpha}}.$$

2. **(8 points) Stability of approximate RERM.**

   Let S and S' differ in one example. Prove that for every test point z,

   $$|\ell(\tilde{w}_S, z) - \ell(\tilde{w}_{S'}, z)| \leq \frac{2G^2}{\lambda\alpha n} + 2G\sqrt{\frac{2\eta}{\lambda\alpha}}.$$

3. **(8 points) Learning guarantee with optimization error.**

   Prove the expected true-risk bound

   $$\mathbb{E}_S L_{\mathcal{D}}(\tilde{w}_S) \leq L_{\mathcal{D}}(u) + \lambda\Psi(u) + \frac{2G^2}{\lambda\alpha n} + G\sqrt{\frac{2\eta}{\lambda\alpha}}$$

   for every comparator $u \in \mathcal{W}$.

4. **(7 points) How accurate must the optimizer be?**

   Suppose $B > 0$ and we want to compete with all $u \in \mathcal{W}$ satisfying $\Psi(u) \leq B^2$. Use

   $$\lambda = \sqrt{\frac{2G^2}{\alpha B^2 n}}$$

   and simplify the bound from the previous part.

   Give a sufficient condition on η so that

   $$G\sqrt{\frac{2\eta}{\lambda\alpha}} \leq \frac{GB}{\sqrt{\alpha n}}.$$

   State the resulting excess-risk bound under this condition.

5. **(5 points) Soft-margin linear classification.**

   Let

$$\ell(w,(x,y)) = (1 - y\langle w, x\rangle)_+, \quad y \in \{-1,+1\}, \quad \|x\|_2 \le R,$$

   and use $\Psi(w) = \frac{1}{2}\|w\|_2^2$.

   Specialize the bound from the previous part to compete with all u satisfying $\|u\|_2 \le B$, and state the approximate-RERM bound in terms of R, B, n, and η. Then give a sufficient condition on η under which the optimization-error term is no larger than $\frac{RB}{\sqrt{n}}$.

Part C: A Weighted Euclidean Geometry (15 points)

Assume linear prediction with scalar loss convex and g-Lipschitz in the prediction. Let $b_1, \ldots, b_d > 0$ and $r_1, \ldots, r_d > 0$. Suppose

$$\mathcal{C}_b = \left\{ w \in \mathbb{R}^d : |w_j| \le b_j \quad \text{for all} \quad j \right\}$$

and $|\varphi_j(x)| \le r_j$ for all x, j.

For $Q = \operatorname{diag}(q_1, \ldots, q_d)$ with $q_j > 0$, define

$$\|w\|_Q = \sqrt{w^\top Q w}, \quad \Psi_Q(w) = \frac{1}{2} w^\top Q w.$$

You may use that Ψ_Q is 1-strongly convex with respect to $\|\cdot\|_Q$.

1. **(5 points) Geometry quantities.**

   Prove that

$$\|\varphi(x)\|_{Q,*}^2 \le \sum_{j=1}^{d} \frac{r_j^2}{q_j}, \quad \sup_{w \in \mathcal{C}_b} \Psi_Q(w) = \frac{1}{2} \sum_{j=1}^{d} q_j b_j^2.$$

2. **(5 points) Fixed Q.**

   Apply the Week 8 theorem and optimize over λ to get

$$\mathbb{E} L_{\mathcal{D}}(A_\lambda(S)) \le \inf_{w \in \mathcal{C}_b} L_{\mathcal{D}}(w) + \frac{2g}{\sqrt{n}} \sqrt{\left(\sum_{j=1}^{d} q_j b_j^2\right)\left(\sum_{j=1}^{d} \frac{r_j^2}{q_j}\right)}.$$

3. **(5 points) Best diagonal geometry.**

   Optimize the right-hand side over $q_j > 0$ and show that the best excess term is

$$\frac{2g}{\sqrt{n}} \sum_{j=1}^{d} b_j r_j.$$

Part D: AI Usage Report (10 points)

Write a short report describing how you used AI in this assignment. Do not just list tools; explain what role AI played in your work and how you checked the result. Address:

1. Describe the parts of the assignment for which you used AI. For example: exploring examples, proposing conjectures, checking algebra, debugging code or formalizations, or improving exposition.
2. Describe concrete AI suggestions you accepted and explain why.
3. Describe concrete AI suggestions you rejected or substantially modified, and explain what was wrong, incomplete, or unhelpful about them.
4. Describe how you verified the correctness of what you submitted. Be specific about the relevant kind of work in this assignment: proof, derivation, code, experiment, or exposition.

AI workflow. Also describe concrete updates to your AI workflow that resulted from this assignment. This may include changes to `CLAUDE.md`, `AGENTS.md`, prompts, checklists, scripts, or skills. **Explain the 5 most recent changes you made to your AI workflow and why**.

If you did not use AI for some part of the assignment, say so explicitly.