Additional Topics

Machine Learning Theory (MLT)

Edinburgh

Rik Sarkar

Today

- Comments on coursework
- Comments on exam

- Additional discussion topics
  - Rademacher complexity
  - Privacy attacks on ML

Comments on coursework

- Marking getting closer to complete
- Expect marks end of next week
- Solutions a little earlier

- Comment: Some of you have written very long answers!
  - Instructions asked for short answers
  - Long answers suggest that you do not quite know the important point…
  - Avoid in exam!

Exam

- See past years' papers. Similar structure
- Answer 2 questions out of 3.
- You are allowed 3 sheets (6 pages) of notes
  - In any form
- What is included
  - All slides except "additional topics"
    - Corresponding topics from book Understanding Machine Learning
  - Lecture notes
  - Tutorials
  - Coursework
  - Stability definitions from Bosquet et al.
  - Random Projections Section 23.2 in Understanding Machine Learning
- Review session for exam in revision week (April 22 -- 26)

Tips on preparing notes

- Prepare your own notes!
- Handwrite!
  - Ensures legibility
  - Good for memory and understanding

- Pay attention to differential privacy

Rademacher complexity

- Suppose we have a sample set S, loss function L and hypothesis class $\mathcal{H}$
- And suppose we want to estimate the worst case generalization gap:
  - $\sup_{h \in \mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)|$
- One way to do that is, split S into $S_1 \cup S_2$, and compute
  - $\sup_{h \in \mathcal{H}} \left(L_{S_1}(h) - L_{S_2}(h)\right)$
  - Taking multiple test and training splits and taking the max
  - Larger gap implies larger complexity of $\mathcal{H}$

Rademacher complexity

- Written more formally using:
  - A combination of loss and hypothesis: $\mathcal{F} = \ell \circ \mathcal{H} : \mathrm{z} \to \ell(h, z)$
  - A selector vector $\sigma = (\sigma_1, \ldots \sigma_m) \in \{+1, -1\}^m$
    - Decides S_1 vs S_2 for each sample

- Rademacher complexity $R(\mathcal{F} \circ S) \stackrel{\text{def}}{=} \frac{1}{m} \mathbb{E}_\sigma[\sup_{\mathrm{f} \in \mathcal{F}} \sum \sigma_i \, f(z_i)]$

- A complexity measure in terms of both $\mathcal{H}$ and S.
  - In contrast of VC dimension, which is only in terms of $\mathcal{H}$

- $R(\mathcal{F} \circ S)$
  - Empirical Rademacher complexity – measured from data

- General Rademacher Complexity: Expectation over distribution
  - $\mathcal{R}(\mathcal{F}) = \mathbb{E}_{S \in \mathcal{D}}[R(\mathcal{F} \circ S)]$

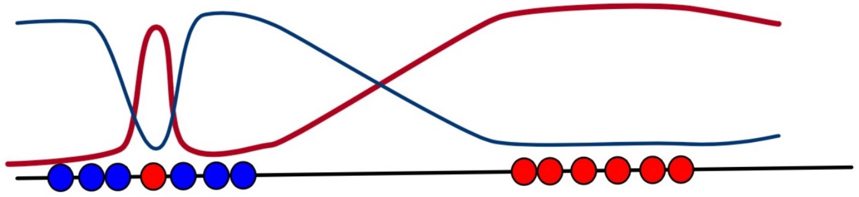
Bounds using Rademacher complexity

- Asuming $||x||_2 \leq R, ||\mathrm{w}||_2 \leq B$, Loss ρ-Lipschitz

- Then we can get bounds of the form:
  - $L_{\mathcal{D}}(w) \leq L_S(w) + \frac{2\rho BR}{\sqrt{m}} + c\sqrt{\frac{2\ln\frac{2}{\delta}}{m}}$
  - Observe: gap increases with B and R…

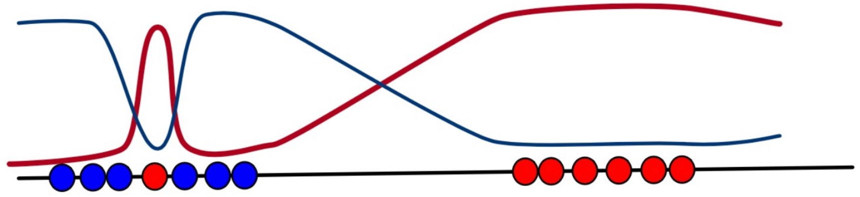
Privacy attacks

- Membership inference attacks
- Given a model M, can an attacker tell if data point x was used in training?
- Idea:
  - If datapoint x was used in training, model is more likely to get it right
- Simple strategy:
  - Set a threshold t. If $M(x)$ is correct, and with a confidence greater than t then output: “in training data”
  - Else, output ”not in training data”



More complex strategy

- Train a “shadow model”
  - E.g. using public data
- Compare prediction and confidence of $M(x)$



Why in membership inference important?

- If attacker can do membership inference correctly, they can derive values for completely unknown data points
- E.g. repeatedly make queries at successive values
  - E.g. follow gradient of loss…
- High probability outputs represent actual point values