Solution to Final Exam of DDA4210

1. Single-choice Questions (30 points in total, 2 points per questions)

1-8. CDBA BCDC

9-15. DBAC BAD

2.1 Adaptive Boosting Method (7 points)
Solution.
Lecture 2, Page 26

2.2 Rademacher Complexity (10 points)

Given a function class $\mathcal{F} = \{x \to \langle \mathbf{w}, x\rangle : \|\mathbf{w}\|_2 \leq \alpha\}$ and a set of i.i.d samples $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$, where $\max_i \|x_i\|_2 \leq \beta$. Prove that the Rademacher complexity satisfies the following inequality

$$\mathcal{R}(\mathcal{F} \circ \mathcal{S}) \leq \frac{\alpha\beta}{\sqrt{n}}.$$

Solution.

$$\begin{aligned}
\mathcal{R}(\mathcal{F} \circ \mathcal{S}) &= \mathbb{E}_\sigma \left[\sup_{f \in \mathcal{F} \circ \mathcal{S}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(\mathbf{x}_i) \right] \\
&= \mathbb{E}_\sigma \left[\sup_{\|\mathbf{w}\|_2 \leq \alpha} \frac{1}{n} \sum_{i=1}^{n} \sigma_i \langle \mathbf{w}, \mathbf{x}_i \rangle \right] \text{ (2 points)} \\
&= \mathbb{E}_\sigma \left[\frac{1}{n} \sup_{\|\mathbf{w}\|_2 \leq \alpha} \left\langle \mathbf{w}, \sum_{i=1}^{n} \sigma_i \mathbf{x}_i \right\rangle \right] \\
&\leq \mathbb{E}_\sigma \left[\frac{1}{n} \sup_{\|\mathbf{w}\|_2 \leq \alpha} \|\mathbf{w}\|_2 \left\| \sum_{i=1}^{n} \sigma_i \mathbf{x}_i \right\|_2 \right] \text{ (use Chauchy-Schwatz inequality)} \\
&\leq \mathbb{E}_\sigma \left[\frac{1}{n} \alpha \left\| \sum_{i=1}^{n} \sigma_i \mathbf{x}_i \right\|_2 \right] \\
&= \frac{\alpha}{n} \mathbb{E}_\sigma \left[\left\| \sum_{i=1}^{n} \sigma_i \mathbf{x}_i \right\|_2 \right] \text{ (3 points)} \\
&= \frac{\alpha}{n} \mathbb{E}_\sigma \left[\left(\left\| \sum_{i=1}^{n} \sigma_i \mathbf{x}_i \right\|_2^2 \right)^{1/2} \right] \\
&\leq \frac{\alpha}{n} \left(\mathbb{E}_\sigma \left[\left\| \sum_{i=1}^{n} \sigma_i \mathbf{x}_i \right\|_2^2 \right] \right)^{1/2} \text{ (use Jensen's inequality) (2 points)} \\
&= \frac{\alpha}{n} \sqrt{\mathbb{E}_\sigma \left[\sum_{i,j} \sigma_i \sigma_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right]} \\
&= \frac{\alpha}{n} \sqrt{\left(\sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \, \mathbb{E}_\sigma[\sigma_i \sigma_j] + \sum_{i=1}^{n} \langle \mathbf{x}_i, \mathbf{x}_i \rangle \, \mathbb{E}_\sigma[\sigma_i^2] \right)} \\
&= \frac{\alpha}{n} \sqrt{\sum_{i=1}^{n} \|x_i\|^2} \\
&\leq \frac{\alpha\beta}{\sqrt{n}} \text{ (3 points)}
\end{aligned}$$

2.3 Spectral Clustering (7 points)

(a) Write down the main steps of spectral clustering. (3 points)

Solution.

1) Construct a similarity matrix W
2) Compute the Laplacian matrix L
3) Perform eigenvalue decomposition on L and use the first K eigenvectors to form a matrix Z
4) normalize the columns of Z to unit ℓ_2 norm
5) Perform K-means clustering on Z

(b) Explain how to determine the number of clusters K if it is not given in advance. (2 points)

Solution.

1) K = the number of zero eigenvalues of the Laplacian matrix L
2) $K = \arg\max_j \Delta_j$, and $\Delta_j = |\lambda_{j+1} - \lambda_j|$, λ_j is the j-th eigenvalue of L

(C) What are the time and space complexities of spectral clustering? Write down them in the form of big O. (2 points)

Solution.

1) Time complexity: $O(n^3)$
2) Space complexity: $O(n^2)$

2.4 Graph Neural Networks (4 points for a, 4 points for b)
lecture06 page 23-30
some key points:

1. 2 points for GCN mathematical expression, 2 points for the objective function.
2. Should specify your notations if they are not given in the question(eg. the label matrix).
3. Readout function in the graph classification task(different from nod classification).

2.5 Generative Models ((3 points for a, 3 points for b, 1 points for c))
Solution:

(a) (1pts) Model architecture of VAE: see Fig. 1.

(1pts) Objective function:

$$\max_{\phi,\theta} \mathbb{E}_{\boldsymbol{z}\sim q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})] - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z}))$$

(1pts) Meanings of notations: please check the slides.

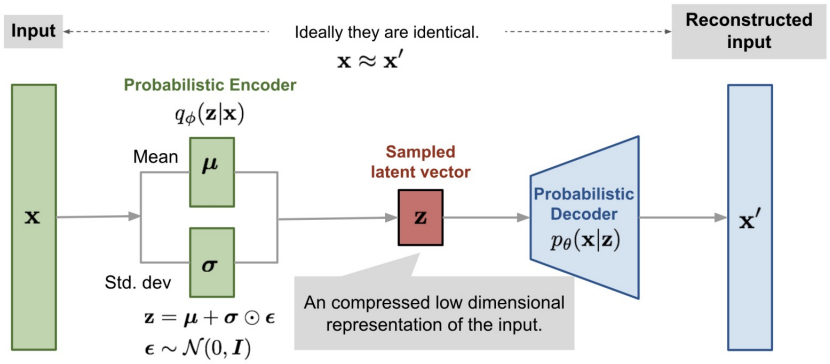

Figure 1: VAE architecture

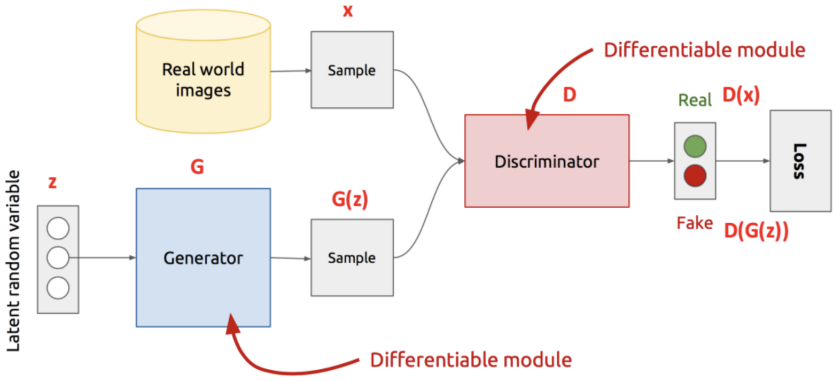

Figure 2: GAN architecture

(b) (1pts) Model architecture of VAE: see Fig. 2.

(1pts) Objective function:

$$\min_G \max_D \mathcal{L}(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_r(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

(1pts) Meanings of notations: please check the slides.

(c) (1pts) Please check the slides. Some key descriptive sentences should arise in your answer which may include

- a forward diffusion process adds small noise (e.g. Gaussian noise) to the sample in T steps slowly...
- Eventually when $T \to \infty$, x_T becomes isotropic Gaussian.

- If the diffusion process can be reversed, using $q(x_{t-1}|x_t)$, we can create a true sample from a Gaussian noise input $x_T \sim \mathcal{N}(0, I)$.

Grading Criteria:

- For (a), only 0.5 pts will be given for the model architecture if one does not specify the key parameters μ and σ in VAE;
- For (c), 0 or 0.5 pts will be given if one does not even point out two basic process.

2.6 Privacy (10 points)

Solution:

The ε-differential privacy property is **not satisfied** when a mechanism uses uniform random noise because the noise does not decay exponentially with respect to the difference in query results from neighboring databases.

To see this more concretely, consider the extreme case where the noisy output is at the edge of the range of the uniform distribution. The probability of obtaining this output from q(X) may be close to zero, whereas the probability of obtaining it from q(X') may be significantly greater than zero. The ratio of these probabilities, which is required to be at most e^{ε} in ε-differential privacy, can thus be **unbounded** in the case of uniform noise.

Therefore, the proposed mechanism does not achieve ε-differential privacy for any constant ε.

Note that the density of the uniform distribution is constant within its range, which does not provide the necessary exponential decay for ε-differential privacy. For the sake of achieving differential privacy, the Laplace mechanism or other mechanisms like the Gaussian mechanism, that use noise distributions with exponentially decreasing tails, are typically used.

Grading Criteria:

- 2 pts can be given if one gives the correct conclusion.
- 3 pts can be given if one gives the formal definition of differential privacy;
- 3 pts will be given if one specifies the unboundness of the probability ratio on neighboring sets;
- 2 pts can be give if one specifies the reasoning behind this conclusion.

2.7 Fairness (11 points) **Solution:** (a)(4 points)Group A:

$$\text{Approval rate} = \frac{300}{1000} = 0.3$$

$$\text{TPR} = \frac{TP}{TP + FN} = \frac{270}{270 + 70} = \frac{27}{34}$$

$$\text{FPR} = \frac{FP}{FP + TN} = \frac{30}{30 + 630} = \frac{1}{22}$$

Group B:

$$\text{Approval rate} = \frac{600}{2000} = 0.3$$
$$\text{TPR} = \frac{TP}{TP+FN} = \frac{540}{540+40} = \frac{27}{29}$$
$$\text{FPR} = \frac{FP}{FP+TN} = \frac{60}{60+1360} = \frac{3}{71}$$

(b)(4 points).i. Yes, ii.No, iii.No
(c)(3 points) An example: Fair in the sense of demographic parity; unfair in the sense of equal opportunity equal oods. Solution: need to be precise, eg. set different accept threshold for group A and B to improve the TPR of A to ensure equal opportunity.

a-0.5bonus 1b-1

2.8 Shapley Value (10 points)

(a) Consider a glove market with $4,000,001$ suppliers (players), denoted by N, where of these the first $2,000,000$ suppliers (denoted by N_L) can each supply one left glove and the other $2,000,001$ suppliers (denoted by N_R) can supply one right glove each. Suppose the worth of each coalition is the number of matched pairs that it can assemble. In general, given a coalition $C \subseteq N$,

$$v(C) = \min\{|C \cap N_L|, |C \cap N_R|\}.$$

What are the Shapley values for all the suppliers?

Solution: For supplier i and a coalition $C \subseteq N$,

- If $|C \cap N_L| < |C \cap N_R|$,

$$v(C \cup \{i\}) - v(C) = \begin{cases} 1 \text{ if } i \in N_L \\ 0 \text{ otherwise} \end{cases}$$

- If $|C \cap N_L| > |C \cap N_R|$,

$$v(C \cup \{i\}) - v(C) = \begin{cases} 0 \text{ if } i \in N_L \\ 1 \text{ otherwise} \end{cases}$$

- If $|C \cap N_L| = |C \cap N_R|$,

$$v(C \cup \{i\}) - v(C) = 0$$

The Shapley value for a supplier $i \in N_R$ is given by

$$\phi_i(v) = \frac{1}{|N|!} \sum_{i=1}^{|N|} \sum_{j=0}^{i/2-1} (|N| - i)!(i-1)! \binom{|N_R| - 1}{j} \binom{|N_L|}{i-j-1}$$

Based on this, we can compute the Shapley values for a supplier $i \in N_L$ by the efficiency axiom.

Grading Criteria:

- 1 pts can be given if one just gives the formula of Shapley value in slides.
- 3 pts can be given if one can give the analysis of characteristic function by cases.
- 2 pts can be given if one gives the formula of Shapley value for this problem.

(b) A machine learning model for binary classification is trained on a dataset with five features $N = \{1, 2, 3, 4, 5\}$. Feature 1 is called a dominant feature. The dominant feature with one or more other features can lead to a prediction 1. The four other features together can also lead to 1. The characteristic function is stated below.

$$v(C) = \begin{cases} 1 & \text{if } 1 \in C \text{and} |C| \geq 2 \\ 1 & \text{if } |C| \geq 4 \\ 0 & \text{otherwise} \end{cases}$$

What are the Shapley values for all features?

Solution: We first compute the Shapley value of feature 1. It follows the next two steps.

- Step 1: For $C \subseteq N \backslash \{1\}$,

$$v(C) = \begin{cases} 1 \text{ if } |C| \geq 4 \\ 0 \text{ otherwise} \end{cases} \quad \text{and} \quad v(C) = \begin{cases} 1 \text{ if } |C| \geq 1 \\ 0 \text{ otherwise} \end{cases}$$

- Step 2: Its Shapley value is

$$\phi_1(v) = \sum_{C \subseteq N \backslash \{1\}} \frac{|C|!|d - |C| - 1|!}{d!} (v(C \cup \{1\}) - v(C))$$

$$= \binom{4}{0}\frac{0!4!}{5!}(0-0) + \binom{4}{1}\frac{1!3!}{5!}(1-0) + \binom{4}{2}\frac{2!2!}{5!}(1-0) + \binom{4}{3}\frac{3!1!}{5!}(1-0) + \binom{4}{4}\frac{4!0!}{5!}(1-1)$$
$$= 4 \times \frac{1}{20} + 6 \times \frac{1}{30} + 4 \times \frac{1}{20}$$
$$= \frac{3}{5}$$

Next, Let's prove that the other four features are equivalent to each other. For any $i, j \in N\backslash\{1\}$ and $C \subseteq N\backslash\{i, j\}$,

- if $1 \in C$, $v(C \cup \{i\}) = 1 = v(C \cup \{j\})$;
- if $1 \notin C$, $|C| \leq 2$ which implies that $v(C \cup \{i\}) = 0 = v(C \cup \{j\})$.

It means that any two features in $N\backslash\{1\}$ are equivalent to each other. Hence, the other four features have equal treatment of equals.

By the axioms of symmetry and efficiency, it has

$$\phi_2(v) = \phi_3(v) = \phi_4(v) = \phi_5(v) = \frac{1}{4}(1 - \frac{3}{5}) = \frac{1}{10}.$$

Grading Criteria:

- 5 pts will be given if all the Shapley values are correct.
- 2 pts can be given if one gives the analysis of characteristic function by cases.
- 1 pts can be given if one just gives the formula of Shapley value.