Machine Learning II

Bjoern Andres, Jannik Irmai, David Stein

Machine Learning for Computer Vision
TU Dresden

Summer Term 2024

Summary. In this part of the course, we show that also the learning of partial functions can be NP-hard. Specifically, we show that separating labeled data by a pair of DNFs defining a partial Boolean function is NP-complete.

Definition. For any finite, non-empty set S, called a set of *samples*, any $X \neq \emptyset$, called an *attribute space* and any $x : S \to X$, the tuple (S, X, x) is called *unlabeled data*.

For any $y : S \to \{0, 1\}$, given in addition and called a *labeling*, the tuple (S, X, x, y) is called *labeled data*.

Definition: Let (S, X, x, y) labeled data with $X = \{0,1\}^J$ and $J \neq \emptyset$ finite. Let $f\colon \Theta \to \mathbb{R}^X$. Let $R\colon \Theta \to \mathbb{R}_0^+$ called a regularizer.

- For any $m \in \mathbb{N}_0$, the instance of the *partial separability problem* is to decide if there exist $\theta, \theta' \in \Theta$ such that

$$
\begin{aligned}
& R(\theta) + R(\theta') \leq m && (1)\\
\forall s \in y^{-1}(1)\colon \quad & f_\theta(x_s) > 0 && (2)\\
\forall s \in y^{-1}(0)\colon \quad & f_{\theta'}(x_s) > 0 && (3)\\
\forall x \in X\colon \quad & f_\theta(x) \leq 0 \ \lor \ f_{\theta'}(x) \leq 0 && (4)
\end{aligned}
$$

- The instance of the *partial separation problem* has the form

$$
\begin{aligned}
\inf_{\theta, \theta' \in \Theta} \quad & R(\theta) + R(\theta') && (5)\\
\text{subject to} \quad & \forall s \in y^{-1}(1)\colon \quad f_\theta(x_s) > 0 && (6)\\
& \forall s \in y^{-1}(0)\colon \quad f_{\theta'}(x_s) > 0 && (7)\\
& \forall x \in X\colon \qquad f_\theta(x) \leq 0 \ \lor \ f_{\theta'}(x) \leq 0 && (8)
\end{aligned}
$$

- For any $L\colon \mathbb{R} \times \{0,1\} \to \mathbb{R}_0^+$ called a *loss function* and any $\lambda \in \mathbb{R}_0^+$, the instance of the *supervised partial learning problem* has the form

$$
\inf_{\theta, \theta' \in \Theta} \lambda(R(\theta) + R(\theta')) + \frac{1}{|y^{-1}(1)|} \sum_{s \in y^{-1}(1)} L(f_\theta(x_s), 1) + \frac{1}{|y^{-1}(0)|} \sum_{s \in y^{-1}(0)} L(f_{\theta'}(x_s), 1) \tag{9}
$$

Definition: For any finite, non-empty set $X = \{0,1\}^J$ and for the sets

$$\Gamma = \left\{ \left(V, \bar{V}\right) \in 2^J \times 2^J | V \cap \bar{V} = \emptyset \right\} \tag{10}$$

$$\Theta = 2^\Gamma \quad , \tag{11}$$

the family $f : \Theta \to \{0,1\}^X$ such that for any $\theta \in \Theta$ and any $x \in X$,

$$f_\theta(x) = \sum_{(J_0, J_1) \in \theta} \prod_{j \in J_0} x_j \prod_{j \in J_1} (1 - x_j) \tag{12}$$

is called the family of J-variate *disjunctive normal forms (DNFs)*.

Moreover: For $R_l, R_d \colon \Theta \to \mathbb{N}_0$ such that for all $\theta \in \Theta$,

$$R_l(\theta) = \sum_{(J_0, J_1) \in \theta} (|J_0| + |J_1|) \tag{13}$$

$$R_d(\theta) = \max_{(J_0, J_1) \in \theta} (|J_0| + |J_1|) \tag{14}$$

$R_l(\theta)$ and $R_d(\theta)$ are called the *length* and *depth*, respectively, of the DNF defined by θ.

Definition. For any set S and any $\emptyset \notin \Sigma \subseteq 2^S$, the set Σ is called a *cover* of S iff

$$\bigcup_{U \in \Sigma} U = S \ . \tag{15}$$

Definition. Let S be any set, let $\emptyset \notin \Sigma \subseteq 2^S$ and let $m \in \mathbb{N}$. Deciding whether there exists a $\Sigma' \subseteq \Sigma$ such that Σ' is a cover of S, and $|\Sigma'| \leq m$ is called the instance of the *set cover problem* with respect to S, Σ and m.

Definition. For any instance (S', Σ, m) of the set cover problem, the *Haussler data* induced by (S', Σ, m) is the labeled data (S, X, x, y) such that

- $S = \{0\} \mathbin{\dot\cup} S'$
- $X = \{0,1\}^\Sigma$
- $x_0 = 0^\Sigma$ and

$$\forall s \in S' \ \forall \sigma \in \Sigma: \quad x_s(\sigma) = \begin{cases} 1 & \text{if } s \in \sigma \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

- $y_0 = 0$ and $\forall s \in S' \colon y_s = 1$

Lemma. For any instance (S', Σ, m) of the set cover problem, consider the instance of the partial separability problem for the family $f\colon \Theta \to \{0,1\}^\Sigma$ of DNFs, $R \in \{R_l, R_d\}$, the Haussler data (S, X, x, y) and the bound $2m$ (for R_l) and $m+1$ (for R_d).

The function $h\colon 2^\Sigma \to \Theta^2$ such that for any $\Sigma' \subseteq \Sigma$, we have $h(\Sigma') := (\theta, \theta')$ with $\theta = \{(\{\sigma\}, \emptyset) \mid \sigma \in \Sigma'\}$ and $\theta' = \{(\emptyset, \Sigma')\}$ has the following properties:

1. $h(\Sigma')$ is computable in time $O(\text{poly}(|\Sigma'||S|))$.
2. If Σ' solves the instance of the set cover problem then $h(\Sigma')$ solves the instance of the partial separability problem.

The function $g\colon \Theta^2 \to 2^\Sigma$ such that for all $\theta, \theta' \in \Theta^2$:
$g(\theta, \theta') \in \text{argmin}\ \{|\Sigma'| : \Sigma' \in \{\Sigma'_0, \Sigma'_1\}\}$ with

$$\Sigma'_0 = \bigcup_{(\Sigma_0, \Sigma_1) \in \theta} \Sigma_0 \tag{17}$$

$$\Sigma'_1 \in \begin{cases} \{\Sigma_1 \subseteq \Sigma \mid (\emptyset, \Sigma_1) \in \theta'\} & \text{if non-empty} \\ \{\emptyset\} & \text{otherwise} \end{cases} \tag{18}$$

has the following properties:

1. $g(\theta, \theta')$ is computable in time $O(\text{poly}(R_l(\theta) + R_l(\theta')))$
2. If (θ, θ') solves the instance of the partial separability problem then $g(\theta, \theta')$ solves the instance of the set cover problem.

Corollary. The partial separability problem is NP-complete.

Proof (sketch). $(\Rightarrow)$ $(\theta, \theta') = h(\Sigma')$ solves the instance of the partial separability problem by construction.

$(\Leftarrow)$ Firstly, we show that Σ'_0 is a solution to the instance of the set cover problem: On the one hand:

$$
\begin{aligned}
& f_{\theta'}(0^{\Sigma}) = 1 \\
\Rightarrow \quad & f_{\theta}(0^{\Sigma}) = 0 && (19) \\
\Rightarrow \quad & \forall (\Sigma_0, \Sigma_1) \in \theta \colon \Sigma_0 \neq \emptyset \ . && (20)
\end{aligned}
$$

On the other hand:

$$
\begin{aligned}
& \forall s \in S' \colon f_{\theta}(x_s) = 1 \\
\Rightarrow \quad & \forall s \in S' \ \exists (\Sigma_0, \Sigma_1) \in \theta \colon \ (\forall \sigma \in \Sigma_0 : x_s(\sigma) = 1) \wedge (\forall \sigma \in \Sigma_1 : x_s(\sigma) = 0) && (21) \\
\Rightarrow \quad & \forall s \in S' \ \exists (\Sigma_0, \Sigma_1) \in \theta \ \exists \sigma \in \Sigma_0 : \quad x_s(\sigma) = 1 \qquad \text{by (20)} && (22) \\
\Rightarrow \quad & \forall s \in S' \ \exists \sigma \in \Sigma'_0 : \quad x_s(\sigma) = 1 && (23) \\
\Rightarrow \quad & \forall s \in S' \ \exists \sigma \in \Sigma'_0 : \quad s \in \sigma \ . && (24)
\end{aligned}
$$

Secondly, we show that Σ_1' is a solution to the instance of the set cover problem: On the one hand:

$$\begin{aligned}
& f_{\theta'}(0^\Sigma) = 1 \\
\Rightarrow \quad & \exists (\Sigma_0, \Sigma_1) \in \theta' : \quad \Sigma_0 = \emptyset && (25) \\
\Rightarrow \quad & \{\Sigma_1 \subseteq \Sigma \mid (\emptyset, \Sigma_1) \in \theta'\} \neq \emptyset \ . && (26)
\end{aligned}$$

On the other hand:

$$\begin{aligned}
& \forall s \in S' : \quad f_\theta(x_s) = 1 \\
\Rightarrow \quad & \forall s \in S' : \quad f_{\theta'}(x_s) = 0 && (27) \\
\Rightarrow \quad & \forall s \in S' \ \exists (\Sigma_0, \Sigma_1) \in \theta' : \ (\exists \sigma \in \Sigma_0 : x_s(\sigma) = 0) \vee (\exists \sigma \in \Sigma_1 : x_s(\sigma) = 1) && (28) \\
\Rightarrow \quad & \forall s \in S' \ \exists \sigma \in \Sigma_1' : \quad x_s(\sigma) = 1 \qquad \text{by (26)} && (29) \\
\Rightarrow \quad & \forall s \in S' \ \exists \sigma \in \Sigma_1' : \quad s \in \sigma \ . && (30)
\end{aligned}$$

Thirdly,

$$
\begin{aligned}
|g(\theta, \theta')| &\leq \min\{|\Sigma'_0|, |\Sigma'_1|\} && (31)\\
&\leq \frac{|\Sigma'_0| + |\Sigma'_1|}{2} && (32)\\
&\leq \tfrac{1}{2}(R_l(\theta') + R_l(\theta)) \ . && (33)
\end{aligned}
$$

Fourthly,

$$
\begin{aligned}
|g(\theta, \theta')| &\leq \min\{|\Sigma'_0|, |\Sigma'_1|\} && (34)\\
&\leq |\Sigma'_0| && (35)\\
&\leq R_d(\theta) && (36)\\
&\leq R_d(\theta) + R_d(\theta') - 1 \ . && (37)
\end{aligned}
$$

□