Machine Learning I

Bjoern Andres (Lectures)
Shengxian Zhao and Jérome Thiessat (Exercises)

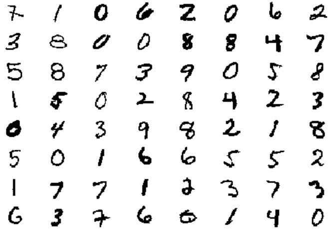
Machine Learning for Computer Vision
TU Dresden

Winter Term 2021/2022

Contents. This part of the course is about a special case of supervised learning: the supervised learning of disjunctive normal forms.

- We state the problem by defining labeled data, a family of functions, a regularizer and a loss function
- We prove that the problem is hard to solve (technically: NP-hard), by relating it to the well-known set cover problem.

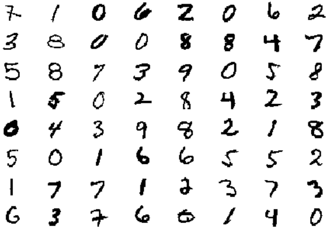
Deciding with Disjunctive Normal Forms



S

$$\rightarrow \begin{pmatrix}0\\1\\2\\3\\4\\5\\6\\7\\8\\9\end{pmatrix}$$

Deciding with Disjunctive Normal Forms



S

$$\rightarrow \quad \begin{pmatrix} 0 & \text{not the digit 7} \\ 1 & \text{the digit 7} \end{pmatrix}$$

Data

We consider binary attributes. More specifically, we consider some finite, non-empty set V, called the set of **attributes**, and labeled data $T = (S, X, x, y)$ such that $X = \{0,1\}^V$.

Hence, $x\colon S \to \{0,1\}^V$ and $y\colon S \to \{0,1\}$.

Family of functions

Let $\Gamma = \{(V_0, V_1) \in 2^V \times 2^V \mid V_0 \cap V_1 = \emptyset\}$ and $\Theta = 2^\Gamma$.

Definition. For any $\theta \in \Theta$ and the $f_\theta \colon \{0,1\}^V \to \{0,1\}$ such that

$$\forall x \in \{0,1\}^V : \quad f_\theta(x) = \bigvee_{(V_0,V_1)\in\theta} \prod_{v \in V_0} (1 - x_v) \prod_{v \in V_1} x_v \ , \tag{1}$$

the form on the r.h.s. of (1) is called the **disjunctive normal form (DNF)** defined by V and θ. The function f_θ is said to be defined by the DNF.

Example. $\{\ (\emptyset, \{v_1, v_2\}),\ (\{v_1\}, \{v_3\})\ \} = \theta \in \Theta$ defines the function

$$f_\theta(x) = x_{v_1}\, x_{v_2} \ \vee\ (1 - x_{v_1})\, x_{v_3} \ . \tag{2}$$

Regularization

In order to quantify the complexity of DNFs, we consider the following regularizers.

Definition. The functions $R_d, R_l : \Theta \to \mathbb{N}_0$ whose values are defined below for any $\theta \in \Theta$ are called the **depth** and **length**, resp., of the DNF defined by θ.

$$R_d(\theta) = \max_{(V_0,V_1)\in\theta}(|V_0| + |V_1|) \tag{3}$$

$$R_l(\theta) = \sum_{(V_0,V_1)\in\theta}(|V_0| + |V_1|) \tag{4}$$

Loss function

We consider the **0/1-loss** L, i.e.

$$\forall r \in \mathbb{R} \; \forall \hat{y} \in \{0,1\}\colon \quad L(r, \hat{y}) = \begin{cases} 0 & r = \hat{y} \\ 1 & \text{otherwise} \end{cases} \; . \tag{5}$$

Definition. For any $R \in \{R_l, R_d\}$ and any $\lambda \in \mathbb{R}_0^+$, the instance of the **supervised learning problem of DNFs** with respect to T, L, R and λ has the form

$$\min_{\theta \in \Theta} \quad \lambda R(\theta) + \frac{1}{|S|} \sum_{s \in S} L(f_\theta(x_s), y_s) \tag{6}$$

Definition. Let $m \in \mathbb{N}$. The instance of the **bounded depth DNF problem** w.r.t. T and m is to decide whether there exists a $\theta \in \Theta$ such that

$$R_d(\theta) \leq m \tag{7}$$

$$\forall s \in S\colon \quad f_\theta(x_s) = y_s \ . \tag{8}$$

The instance of the **bounded length DNF problem** w.r.t. T and m is to decide whether there exists a $\theta \in \Theta$ such that

$$R_l(\theta) \leq m \tag{9}$$

$$\forall s \in S\colon \quad f_\theta(x_s) = y_s \ . \tag{10}$$

Next, we will reduce the hard-to-solve (technically: NP-hard) set cover problem to the bounded length/depth DNF problem, thereby showing that these problems are hard to solve (NP-hard) as well. The reduction is by Haussler (1988).

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

$$\bigcup_{U \in \Sigma} U = S \ . \tag{11}$$

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

- $S = S' \mathbin{\dot\cup} \{1\}$
- $X = \{0,1\}^\Sigma$
- $x_1 = 1^\Sigma$ and

$$\forall s \in S' \; \forall \sigma \in \Sigma \colon \quad x_s(\sigma) = \begin{cases} 0 & \text{if } s \in \sigma \\ 1 & \text{otherwise} \end{cases} \tag{12}$$

- $y_1 = 1$ and $\forall s \in S' \colon y_s = 0$

Lemma 2: For any instance (S', Σ, m) of the set cover problem, the Haussler data (S, X, x, y) induced by (S', Σ, m), and any $\Sigma' \subseteq \Sigma$:

$$\bigcup_{\sigma \in \Sigma'} \sigma = S' \quad \Leftrightarrow \quad \forall s \in S' \colon \prod_{\sigma \in \Sigma'} x_s(\sigma) = 0$$

Proof.

$$\begin{aligned}
& \bigcup_{\sigma \in \Sigma'} \sigma = S' \\
\Leftrightarrow \quad & \forall s \in S' \ \exists \sigma \in \Sigma' \colon \quad s \in \sigma && (13) \\
\Leftrightarrow \quad & \forall s \in S' \ \exists \sigma \in \Sigma' \colon \quad x_s(\sigma) = 0 && (14) \\
\Leftrightarrow \quad & \forall s \in S' \colon \quad \prod_{\sigma \in \Sigma'} x_s(\sigma) = 0 && (15)
\end{aligned}$$

Theorem 1. The set cover problem is reducible to the bounded depth/length DNF problem.

Proof. The proof is for any $R \in \{R_d, R_l\}$.
Let (S', Σ, m) any instance of the set cover problem.
Let $T = (S, X, x, y)$ the Haussler data induced by (S', Σ, m).
We show: There exists a cover $\Sigma' \subseteq \Sigma$ of S' with $|\Sigma'| \leq m$ iff there exists a $\theta \in \Theta$ such that $R(\theta) \leq m$ and $\forall s \in S\colon f_\theta(x_s) = y_s$.

($\Rightarrow$) Let $\Sigma' \subseteq \Sigma$ a cover of S and $|\Sigma'| \leq m$.
Let $V_0 = \emptyset$ and $V_1 = \Sigma'$ and $\theta = \{(V_0, V_1)\}$. Thus,

$$\forall x' \in X\colon \quad f_\theta(x') = \prod_{\sigma \in \Sigma'} x'(\sigma) \tag{16}$$

On the one hand, $\forall s \in S'\colon f(x_s) = 0$, by Lemma 2, and $f(1^\Sigma) = 1$, by definition of f_θ. Thus, $\forall s \in S\colon f(x_s) = y_s$.
On the other hand, $R(\theta) = |\Sigma'| \leq m$.

($\Leftarrow$) Let $\theta \in \Theta$ such that $R(\theta) \leq m$ and $\forall s \in S \colon f_\theta(x_s) = y_s$.
There exists a $(\Sigma_0, \Sigma_1) \in \theta$ such that $\Sigma_0 = \emptyset$, because $1 = y_1 = f_\theta(x_1) = f_\theta(1^\Sigma)$. Moreover:

$$\begin{aligned}
& \forall s \in S' \colon \quad f(x_s) = 0 \\
\Rightarrow \quad & \forall s \in S' \colon \quad \bigvee_{(V_0,V_1)\in\theta} \prod_{v\in V_0} (1 - x_s(v)) \prod_{v \in V_1} x_s(v) = 0 && (17) \\
\Rightarrow \quad & \forall s \in S' \; \forall (V_0, V_1) \in \theta \colon \quad \prod_{v\in V_0} (1 - x_s(v)) \prod_{v \in V_1} x_s(v) = 0 && (18)
\end{aligned}$$

Thus, for $(\emptyset, \Sigma_1) \in \theta$ in particular:

$$\forall s \in S' \colon \quad \prod_{\sigma \in \Sigma_1} x_s(\sigma) = 0 \tag{19}$$

And by virtue of Lemma 2:

$$\bigcup_{\sigma \in \Sigma_1} \sigma = S' \tag{20}$$

Furthermore, $|\Sigma_1| \leq R(\theta) = m$. $\square$

Summary: Supervised learning of DNFs is hard. More specifically, the NP-hard set cover problem is reducible to the bounded length/depth DNF problem by construction of Haussler data.