

# Computer Vision I

Bjoern Andres

Machine Learning for Computer Vision  
TU Dresden



<https://mlcv.cs.tu-dresden.de/courses/26-winter/cv1/>

Winter Term 2026/2027

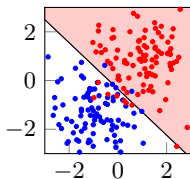
### Contents.

- ▶ This part of the course motivates and defines the  $l_2$ -regularized **logistic regression** problem and the **steepest descent algorithm**.
- ▶ Moreover, it defines the **supervised learning** problem and shows that logistic regression is a special case of supervised learning.

**Problem.** Given a finite set  $V$  of pixels, the set  $X = \mathbb{R}^V$  of images, a finite image collection  $x : S \rightarrow X$  and binary labels  $y : S \rightarrow \{0, 1\}$ , **find** a function  $g : X \rightarrow \{0, 1\}$  to assign a label  $g(x)$  to **any** image  $x \in X$ .

**Example.** Learning to identify precisely the images of the hand-written digit 7.

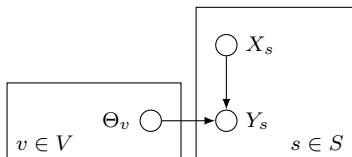
|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 7 | 1 | 0 | 6 | 2 | 0 | 6 | 2 |
| 3 | 8 | 0 | 0 | 8 | 8 | 4 | 7 |
| 5 | 8 | 7 | 3 | 9 | 0 | 5 | 8 |
| 1 | 5 | 0 | 2 | 8 | 4 | 2 | 3 |
| 0 | 4 | 3 | 9 | 8 | 2 | 1 | 8 |
| 5 | 0 | 1 | 6 | 6 | 5 | 5 | 2 |
| 1 | 7 | 7 | 1 | 2 | 3 | 7 | 3 |
| 6 | 3 | 7 | 6 | 0 | 1 | 4 | 0 |



To begin with, consider **linear functions**. More specifically, consider  $\Theta = \mathbb{R}^V$  and  $f : \Theta \rightarrow \mathbb{R}^X$  such that for all  $\theta \in \Theta$  and all  $\hat{x} \in X$ :

$$f_{\theta}(\hat{x}) = \langle \theta, \hat{x} \rangle = \sum_{v \in V} \theta_v \hat{x}_v \quad (1)$$

## Machine Learning for Computer Vision (1/3)



We introduce a probabilistic model:

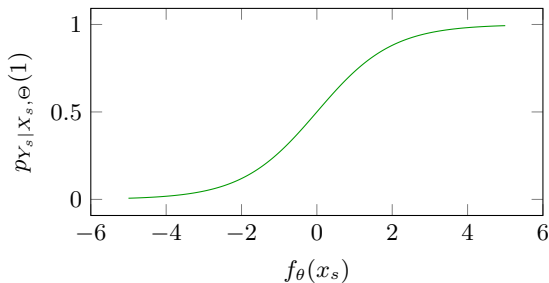
- For any  $s \in S$ , let  $X_s$  be a random variable whose value is a vector  $x_s \in \mathbb{R}^V$ , the **image**  $x_s$  indexed by  $s$  in the collection.
- For any  $s \in S$ , let  $Y_s$  be a random variable whose value is a binary number  $y_s \in \{0, 1\}$ , the **label**  $y_s$  of the image  $x_s$  indexed by  $s$  in the collection.
- For any  $v \in V$ , let  $\Theta_v$  be a random variable whose value is a real number  $\theta_v \in \mathbb{R}$ , a **parameter** of the function we seek to learn.

We assume the conditional independencies of the Bayes net depicted above.  
Thus,

$$P(X, Y, \Theta) = \prod_{s \in S} (P(Y_s \mid X_s, \Theta) P(X_s)) \prod_{v \in V} P(\Theta_v) . \quad (2)$$

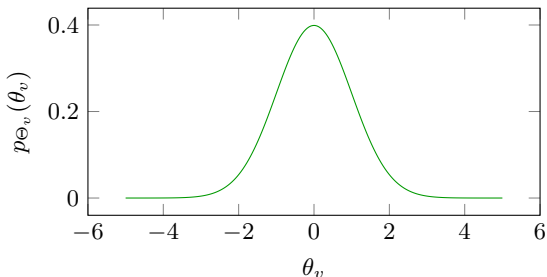
- We assume a **sigmoid distribution**

$$\forall s \in S : \quad p_{Y_s|X_s, \Theta}(1) = \frac{1}{1 + e^{-f_{\theta}(x_s)}} \quad (3)$$



- We assume a **normal distribution** with  $\sigma \in \mathbb{R}^+$ :

$$\forall v \in V : \quad p_{\Theta_v}(\theta_v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\theta_v^2/2\sigma^2} \quad (3)$$



We attempt to **learn** parameters by maximizing their likelihood

$$\begin{aligned} & \prod_{s \in S} (P(Y_s | X_s, \Theta) P(X_s)) \prod_{v \in V} P(\Theta_v) \\ & \propto \prod_{s \in S} P(Y_s | X_s, \Theta) \prod_{v \in V} P(\Theta_v) . \end{aligned} \quad (4)$$

We attempt to **infer** labels by maximizing the conditional probability

$$P(Y | X, \Theta) = \prod_{s \in S} P(Y_s | X_s, \Theta) . \quad (5)$$

**Lemma 1.** Estimating maximally probable parameters  $\theta$ , given images  $x$  and labels  $y$ , i.e.,

$$\max_{\theta \in \mathbb{R}^V} p_{\Theta|X,Y}(\theta, x, y)$$

is equivalent to the optimization problem

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

with  $L$ ,  $R$  and  $\lambda$  such that

$$\forall r \in \mathbb{R} \ \forall b \in \{0, 1\}: \quad L(r, b) = -br + \ln(1 + e^r) \quad (7)$$

$$\forall \theta \in \Theta: \quad R(\theta) = \|\theta\|_2^2 \quad (8)$$

$$\lambda = \frac{1}{2\sigma^2} \quad (9)$$

It is called the  **$l_2$ -regularized logistic regression problem**.

*Proof.* Firstly,

$$\begin{aligned}
 & \operatorname{argmax}_{\theta \in \mathbb{R}^V} \prod_{s \in S} p_{Y_s|X_s, \Theta}(y_s, x_s, \theta) \prod_{v \in V} p_{\Theta_v}(\theta_v) \\
 &= \operatorname{argmax}_{\theta \in \mathbb{R}^V} \sum_{s \in S} \ln p_{Y_s|X_s, \Theta}(y_s, x_s, \theta) + \sum_{v \in V} \ln p_{\Theta_v}(\theta_v) \quad (10)
 \end{aligned}$$

Secondly,

$$\begin{aligned}
 & \ln p_{Y_s|X_s, \Theta}(y_s, x_s, \theta) \\
 &= y_s \ln p_{Y_s|X_s, \Theta}(1, x_s, \theta) + (1 - y_s) \ln p_{Y_s|X_s, \Theta}(0, x_s, \theta) \\
 &= y_s \ln \frac{p_{Y_s|X_s, \Theta}(1, x_s, \theta)}{p_{Y_s|X_s, \Theta}(0, x_s, \theta)} + \ln p_{Y_s|X_s, \Theta}(0, x_s, \theta) \quad (11)
 \end{aligned}$$

Thus, with (3) and (4):

$$\operatorname{argmin}_{\theta \in \mathbb{R}^V} \sum_{s \in S} \left( -y_s f_{\theta}(x_s) + \ln \left( 1 + e^{f_{\theta}(x_s)} \right) \right) + \frac{1}{2\sigma^2} \|\theta\|_2^2 \quad (12)$$

□

**Lemma 2.** The objective function

$$\varphi(\theta) = \lambda R(\theta) + \sum_{s \in S} L(\langle \theta, x_s \rangle, y_s) \quad (13)$$

of the **linear**  $l_2$ -regularized logistic regression problem is convex.

*Proof.* (Not part of this course. See Machine Learning 1).

The  $l_2$ -regularized logistic regression problem can be solved, e.g., by the steepest descent algorithm:

**Algorithm.** Steepest descent with line search and tolerance parameter  $\epsilon \in \mathbb{R}_0^+$

---

```
 $\theta := 0$ 
repeat
   $d := \nabla \varphi(\theta)$ 
   $\eta := \operatorname{argmin}_{\eta' \in \mathbb{R}} \varphi(\theta - \eta' d)$  (line search)
   $\theta := \theta - \eta d$ 
  if  $\|d\| < \epsilon$ 
    return  $\theta$ 
```

---

**Lemma 3.** Maximally probable labels  $y$ , given images  $x$  and parameters  $\theta$ , i.e., solutions to

$$\operatorname{argmax}_{y \in \{0,1\}^S} p_{Y|X,\Theta}(y, x, \theta) \quad (14)$$

are such that

$$\forall s \in S: \quad y_s = \begin{cases} 1 & \text{if } f_\theta(x_s) > 0 \\ 0 & \text{if } f_\theta(x_s) < 0 \end{cases} . \quad (15)$$

*Proof.* Firstly,

$$\begin{aligned}
 & \operatorname{argmax}_{y \in \{0,1\}^S} p_{Y|X,\Theta}(y, x, \theta) \\
 &= \operatorname{argmax}_{y \in \{0,1\}^S} \prod_{s \in S} p_{Y_s|X_s,\Theta}(y_s, x_s, \theta) \\
 &= \operatorname{argmax}_{y \in \{0,1\}^S} \sum_{s \in S} \ln p_{Y_s|X_s,\Theta}(y_s, x_s, \theta) \\
 &= \operatorname{argmax}_{y \in \{0,1\}^S} \sum_{s \in S} \left( y_s \ln \frac{p_{Y_s|X_s,\Theta}(1, x_s, \theta)}{p_{Y_s|X_s,\Theta}(0, x_s, \theta)} + \ln p_{Y_s|X_s,\Theta}(0, x_s, \theta) \right) \\
 &= \operatorname{argmin}_{y \in \{0,1\}^S} \sum_{s \in S} \left( -y_s f_\theta(x_s) + \ln \left( 1 + e^{f_\theta(x_s)} \right) \right) \\
 &= \operatorname{argmin}_{y \in \{0,1\}^S} \sum_{s \in S} L(f_\theta(x_s), y_s) .
 \end{aligned}$$

Secondly,

$$\min_{y \in \{0,1\}^S} \sum_{s \in S} \left( -y_s f_\theta(x_s) + \ln \left( 1 + e^{f_\theta(x_s)} \right) \right) = \sum_{s \in S} \max_{y_s \in \{0,1\}} y_s f_\theta(x_s) .$$

## Contents (reflection).

- ▶ So far, we have derived the  $l_2$ -regularized **logistic regression** problem from a probabilistic model.
- ▶ Next, we define the **supervised learning** problem and observe that logistic regression is a special case.

**Definition 1.** For any finite, non-empty set  $S$ , called a set of **samples**, any  $X \neq \emptyset$ , called an **input space** and any  $x : S \rightarrow X$ , the tuple  $(S, X, x)$  is called **unlabeled data**. For any  $y : S \rightarrow \{0, 1\}$ , given in addition and called a **labeling**, the tuple  $(S, X, x, y)$  is called **labeled data**.

**Definition 2.** For

- ▶ any labeled data  $T = (S, X, x, y)$ ,
- ▶ any  $\Theta \neq \emptyset$  and  $f : \Theta \rightarrow \mathbb{R}^X$ ,
- ▶ any  $R : \Theta \rightarrow \mathbb{R}_0^+$ , called a **regularizer**,
- ▶ any  $L : \mathbb{R} \times \{0, 1\} \rightarrow \mathbb{R}_0^+$ , called a **loss function** and
- ▶ any  $\lambda \in \mathbb{R}_0^+$ :

the instance of the **supervised learning problem** has the form

$$\inf_{\theta \in \Theta} \lambda R(\theta) + \sum_{s \in S} L(f_{\theta}(x_s), y_s) \quad (16)$$

**Example 1. For**

- ▶ any labeled data  $T = (S, X, x, y)$  with  $X = \mathbb{R}^V$  and  $V$  finite,
- ▶  $\Theta = \mathbb{R}^V$  and  $f : \Theta \rightarrow \mathbb{R}^X$  such that for all  $\theta \in \Theta$  and all  $\hat{x} \in X$ :

$$f_{\theta}(\hat{x}) = \langle \theta, \hat{x} \rangle = \sum_{v \in V} \theta_v \hat{x}_v \quad (17)$$

- ▶  $R : \Theta \rightarrow \mathbb{R}_0^+$  such that for all  $\theta \in \Theta$ :  $R(\theta) = \|\theta\|_2^2$
- ▶  $L : \mathbb{R} \times \{0, 1\} \rightarrow \mathbb{R}_0^+$  such that for all  $r \in \mathbb{R}$  and all  $b \in \{0, 1\}$ :

$$L(r, b) = -rb + \ln(1 + e^r) \quad (18)$$

- ▶ any  $\lambda \in \mathbb{R}_0^+$ :

the **supervised learning problem**

$$\inf_{\theta \in \Theta} \lambda R(\theta) + \sum_{s \in S} L(f_{\theta}(x_s), y_s) \quad (19)$$

is called  **$l_2$ -regularized linear logistic regression**.