Machine Learning II

Jannik Irmai, Jannik Presberger, David Stein, Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

https://mlcv.cs.tu-dresden.de/courses/25-summer/ml2/

Summer Term 2025

The main idea is to define a particular family f of functions f_θ we can learn:

Definition. For

- any labeled data (S, X, x, y),
- any non-empty Θ', Θ'' and $\Theta := \Theta' \times \Theta''$,
- any $g\colon \Theta' \to \mathbb{R}^X$, i.e. $g_{\theta'}\colon X \to \mathbb{R}$,
- any $\alpha\colon \Theta'' \to (\mathbb{R}_0^+)^{X \times X}$ called **attention**, i.e. $\alpha_{\theta''}\colon X \times X \to \mathbb{R}_0^+$,

define for any $(\theta', \theta'') = \theta \in \Theta$ and any $x \in X$:

$$f_\theta(x) = \frac{1}{n_{\theta'' S}(x)} \sum_{s \in S} \alpha_{\theta''}(x, x_s)\, g_{\theta'}(x_s)$$

with

$$n_{\theta'' S}(x) := \sum_{s \in S} \alpha(x, x_s) \quad .$$

Remark. Instead of mapping a feature vector x to the real number $g_\theta(x)$, we relate x to *every labeled sample* x_s by $\alpha_{\theta''}(x, x_s)$ and average the real numbers $g_\theta(x_s)$ obtained for the labeled samples, weighted by $\frac{1}{n_{\theta'' S}(x)} \alpha_{\theta''}(x, x_s)$.

From now on, we concentrate on the following special case.

Example. $X = \mathbb{R}^J$ with $J \neq \varnothing$ finite. Moreover, $\Theta' = \mathbb{R}^J$ and for all $x \in X$ and all $\theta' \in \Theta'$:

$$g_{\theta'}(x) = \langle \theta', x \rangle \ .$$

Remark.

$$\begin{aligned} f_\theta(x) &= \frac{1}{n_{\theta'' S}(x)} \sum_{s \in S} \alpha_{\theta''}(x, x_s)\, g_{\theta'}(x_s) \\ &= \frac{1}{n_{\theta'' S}(x)} \sum_{s \in S} \alpha_{\theta''}(x, x_s) \langle \theta', x_s \rangle \\ &= \frac{1}{n_{\theta'' S}(x)} \left\langle \theta', \sum_{s \in S} \alpha_{\theta''}(x, x_s)\, x_s \right\rangle \end{aligned}$$

Example. a) *Dot product attention*: $\Theta'' = \Theta^q \times \Theta^k$ with $\Theta^q = \mathbb{R}^{K \times J}$ and $\Theta^k = \mathbb{R}^{K \times J}$ with $K \neq \varnothing$ finite. Moreover, for all $(\theta^q, \theta^k) = \theta'' \in \Theta''$ and all $x, x_s \in X$:

$$\alpha_{\theta''}(x, x_s) = e^{\langle \theta^q x, \theta^k x_s \rangle}$$

b) *Mahalanobis distance attention*: $\Theta'' = \mathbb{R}^{K \times J}$ with $K \neq \varnothing$ finite. Moreover, for all $\theta'' \in \Theta''$ and all $x, x_s \in X$:

$$\alpha_{\theta''}(x, x_s) = e^{-\|\theta''(x - x_s)\|_2^2}$$