Remarks on loss funct (info geo)

giovedì 9 febbraio 2023 15:20

$\underline{\underline{DL}}$ $\quad L(x, w) \overset{def}{=} -\lg \dfrac{e^{s_{y_x}(x)}}{\sum_j e^{s_j(x)}}$ $\quad$ cross entropy

image $x \in \mathbb{R}^{784}$, weights $w \in \mathbb{R}^{10000}$

$x = \boxed{2} \to \begin{bmatrix} \cdot \\ \cdot \\ 2 \\ \cdot \\ \cdot \end{bmatrix} \begin{matrix} 0 \\ 1 \\ = y_x \\ \\ \end{matrix}$

cat = 1
dog = 2
horse = 3

$1 = y_x$, $x \sim$ CAT $\begin{bmatrix} 0.1 \\ 0.2 \\ 0.7 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$

$s(x) = 10\left\{\begin{bmatrix} 12 \\ 144 \\ 13 \\ \vdots \\ 7 \end{bmatrix}\right. \rightsquigarrow \dfrac{e^{s_i(x)}}{\sum_j e^{s_j(x)}} = p_i(x, w) \qquad \sum p_i(x, w) = 1$

General

Amari Loss, discrete

$H(p) = \mathbb{E}_p[(\overbrace{-\lg p}^{})] \overset{\downarrow}{=} -\sum p_i \lg p_i \qquad q(x)\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right.$

Shannon entropy $\quad p(x, w)$ empirical prob $P = (0.1, 0.2, 0.7)$ $\quad 1 \cdot \lg 1 + 0 \lg 0 + 0 \lg 0 = 0$

cross entropy $q(x)$ $\quad i = y_x$, $0, 1$ $\quad q = (\overset{y_x}{1}, 0, 0)$

$H(q, p) \overset{def}{=} -\mathbb{E}_q[\lg p] = -\sum q_i(x) \lg p_i(x, w) = -\lg \dfrac{e^{s_{y_x}}}{\sum_j e^{s_j}}$

$KL(q \| p) \overset{def}{=} \sum q_i \lg \dfrac{q_i}{p_i} = \underbrace{\sum_i q_i \lg q_i}_{0 \,=\, H(q)} - \underbrace{\sum q_i \lg p_i}_{+\, H(q, p)} = L(x, w)$

Fisher matrix $\qquad F = \mathbb{E}_p[\nabla_w(\lg p) \cdot \nabla_w(\lg p)^t]$

$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot (u_1 \cdots u_n) = \begin{pmatrix} v_1 u_1 & v_1 u_2 & \\ v_2 u_1 & \cdots & \end{pmatrix}$

Prop: $\quad KL(p(x, w) \| p(x, w + \delta w)) \simeq \frac{1}{2} \delta w^t F \delta w$ $\quad$ ↑ DL

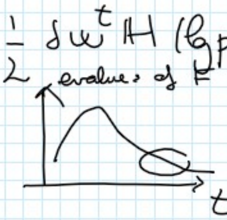
Pf $\quad$ to save time

$KL(p(w) \| p(w + \delta w)) = \sum p_i(w) \lg \dfrac{p_i(w)}{p_i(w + \delta w)} = \left[w_{t+1} = w_t - \alpha \nabla L \right]$

$= \sum p_i(w) \lg p_i(w) - \sum p_i(w) \underbrace{\lg p_i(w + \delta w)}_{\lg p_i(w) + \frac{\nabla p_i}{p_i} \cdot \delta w + \frac{1}{2} \delta w^t H(\lg p_i) \delta w} =$

$= \cancel{\sum p_i(w) \lg p_i(w)} - \sum \cancel{p_i(w)} \lg p_i(w) +$

$- \sum \cancel{p_i(w)} \dfrac{\nabla p_i(w)}{\cancel{p_i(w)}} \delta w - \frac{1}{2} \delta w^t \sum p_i(w) H(\lg p_i) \delta w$



$= \cancel{-\nabla(\sum p_i)} - \frac{1}{2} \delta w^t \underbrace{\sum p_i(w) H(\lg p_i)}_{-F} \delta w \qquad F = \mathbb{E}_p[H(\lg p)]$

F is thought as a metric (but it is not in DL)

Metric $\quad w \in \mathbb{R}^n \longrightarrow F(w) = \mathbb{E}_p[\nabla \lg p \cdot \nabla \lg p^t]$

F is thought as a metric (but it is not in DL)

Metric $\quad \underset{\text{pt par. space}}{w \in \mathbb{R}^n} \longrightarrow F(w) = \mathbb{E}_p[\nabla \lg p \cdot \nabla \lg p^t]$

metric : $p \in \mathbb{R}^n \to$ scal. prod on $T_p\mathbb{R}^n = \mathbb{R}^n$ non deg.

Obs $\quad \ker F(w) = \operatorname{span}\{\nabla_w p_i\}^\perp \subset \mathbb{R}^n \qquad p \quad$ 0.1 | 0.2 $\quad p_1 p_2 \cdots p_{10}$

big dim ! $\qquad n \simeq 10000$

(small rk $F(w)$)

In fact

$$u^t F u = u^t \mathbb{E}_p[\nabla \lg p_i (\nabla \lg p_i)^t] u =$$

$$= \sum p_i \underbrace{(u \cdot \nabla \lg p_i)}_{u_k \partial_k \lg p_i} \underbrace{(u \cdot \nabla \lg p_i)^t}_{\partial_j \lg p_i \cdot u_j} =$$

$$= \sum p_i (u \cdot \nabla \lg p_i)^2 = \sum \cancel{p_i} \frac{(u \cdot \nabla p_i)^2}{p_i^{\cancel{2}}} \geq 0$$

$\ker F \subset \operatorname{span}\{\nabla p_i\}^\perp$

If $u \in \ker F \to Fu = 0 \to u^t F u = 0$

$u \cdot \nabla p_i = 0 \quad \forall i \qquad = 0$

$\ker F \supset \operatorname{span}\{\nabla p_i\}^\perp \quad u \cdot \frac{\nabla p_i}{p_i} = 0$

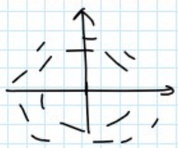
$$Fu = \mathbb{E}_p[\nabla \lg p_i \cdot \nabla \lg p_i^t u] = \mathbb{E}_p[\nabla \lg p \cdot [u \cdot \frac{\nabla p_i}{p_i}]) = 0$$

Hence : $\ker F = \operatorname{span}\{\nabla p_i\}^\perp \Longleftrightarrow \operatorname{rk} F$ is bound by C

Obs. $\quad w \longmapsto \mathcal{D}_w = \operatorname{span}\{\underbrace{\nabla p_1 \cdots \nabla p_c}_{10}\} \qquad F|_{\mathcal{D}_w}$ is non deg.

1) Frobenius Thm NO

2) Subriemann.



$[\nabla p_i, \nabla p_j] \ldots$

Obs $\quad F(x) \rightsquigarrow \mathcal{D}$ good (Th Frobenius)
(involutive)

Mnist $\quad \mathbb{R}^{784} \ni x \qquad \boxed{1} \quad \nabla p_1 \quad \nabla p_7 \qquad 3 \; 5 \quad \boxed{8} \quad 3$