MATH 462

K-class Losses

K-L divergence.

Loss design. Label smoothing

22.10.2021

Lecture 14

project:

FIND Research paper AI

identify novel loss.

Analyse Loss.

EXAMPLES

Next Class. SIMCLR

preview project

- list of DL papers
- loss.

Define (x, y)

image ↗ ↘ class

loss used. why different?

<u>SimCLR</u> Bengio U de M. / <u>data types</u>

<u>Valid</u> NOT Clear

$$\ell(h, y) = (\log(h) - \log(y))^2$$

Data types Design.

Classification K classes

CIFAR-10 dogs cats airplanes

$x \to y \in [1, \dots, K]$

Ambiguous dogs/cats

How?

Answer: KL-divergence · Label smoothing
current

Review Margin Loss

$s \in \mathbb{R}^k \quad y \in \mathcal{Y}_k = \{1, \dots, k\}$

margin loss defn:

$$m(s, y) = s_y - \max_{j \neq y} s_j$$

$$\ell_{mk}(s, y) = \max(0, 1 - m(s, y))$$

<u>EX</u>

$s = (5, 4.2, 3)$
$y = 1$

$$m(s, y) = 5 - \max(4.2, 3)$$
$$= 0.8$$
$$\ell(s, y) = (1 - 0.8)^+ = 0.2$$

$s = (0.5, 4.2, 3)$
$y = 1$

$$m(s, y) = 0.5 - 4.2 = -3.7$$
$$\ell(s, y) = (1 + 3.7)^+ = 4.7$$

Loss Design

$s \in \mathbb{R}^K$ $y \in \mathcal{Y}_K = \{1, \dots, k\}$

margin loss defn:

$$m(s, y) = s_y - \max_{j \neq y} s_j$$

$$\ell_{MK}(s, y) = \max(0, 1 - m(s, y))$$

generalize to $y \in \mathcal{Y}_K$ $k = 4$

y

dog, cat, airplane, car

$s = (5, 5, 0, 0)$	✓		$(1, 0, 0, 0)$
$s = (0, 0, 3, 4)$	✗		$(0, 1, 0, 0)$
$s = (0, 5, 0, 0)$	✓	→	$(1, 1, 0, 0)$
$s = (5, 2, 4.5, 3)$	~		$(0.5, 0.5, 0, 0)$

$K = K_+ \cup K_-$

dog/cat car airpln

$m(s, u)$

$$\max_{j \in K_+} s_j - \max_{j \in K_-} s_j$$

work for ✓ vs. ✗

EX

$m = 5$

$m = -4$

$m = 5$

$m = 0.5$

Current practise

Generalize

$y \in \mathcal{Y}_K$ → represent $y = e_y = (0, 0, 1, 0)$

one hot vector $y = 3$

loss between $p(x)$ prob of class $j = p_j$

Loss $\ell(p, y) = -\log p_y$ ⊛

Augment e_y prob vector. $(0, 0, 1, 0)$

allow q prob vect $(.1\ 0.1\ .7\ 0.1)$

label smoothing

WANT $loss(p, q)$ which reduces to ⊛

when $q = e_y$.

Ans.

write $\Delta^{k-1} = \left\{ (p_1, \dots, p_k) \in \mathbb{R}^k \;\middle|\; \begin{array}{l} \sum_{i=1}^{k} p_i = 1 \\ p_i \geq 0 \quad i=1,\dots,k \end{array} \right\}$

① $KL(q \| p) = - \sum_{i=1}^{k} q_k \log\left(\frac{p_k}{q_k}\right)$ defn "$0 \log 0 = 0$"

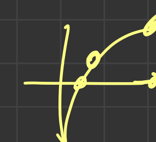
$l_{KL}(p,q) = KL(q \| p)$

check $q = e_y \Rightarrow -\log p_k/1$ ✓

EX $q = (0.5, 0.5, 0, 0)$

$$l_{KL}(p,q) = -\frac{1}{2}\left(\log \frac{p_1}{0.5} + \log \frac{p_2}{0.5}\right)$$

$$= \frac{1}{2} \log\left(\frac{0.5^2}{p_1 p_2}\right)$$



NOTE 0 $p_1 = p_2 = 1/2$

Q. 0.6 0.4

EX $p = (0.6, 0.4, 0, 0)$ $l_{KL}(p,q) = \frac{1}{2} \log\left(\frac{0.25}{0.24}\right)$

$p = (0.2, 0.1, 0.7, 0)$ $l_{KL}(p,q) = \frac{1}{2} \log\left(\frac{0.25}{0.02}\right)$

① $KL(q\|p) = -\sum_{i=1}^{k} q_k \log\left(\frac{p_k}{q_k}\right)$

HW Thm $KL(q\|p) \geq 0$ with equality iff $p = q$.

ANS Jensen's Ineq.

MATH involved in K-class.

Important functions.

① LSE Log Sum exp $s \in \mathbb{R}^K$

$$LSE(s) = \log\left(e^{s_1} + e^{s_2} + \dots + e^{s_K}\right)$$

$$\bar{s} = s_{max} = \max_{i=1}^{K}(s_i)$$

LSE(s) soft max.

Claim

$$s_{max} \le LSE(s) \le s_{max} + \log K$$

proof

$$e^{s_{max}} \le e^{s_1} + \dots + e^{s_K} \le K e^{s_{max}}$$

$$\log(\quad) \le \log(\quad) \le \log(\quad)$$

Ex $s = (0, 3, 0.2)$ $LSE(s) = \log(e^0 + e^3 + e^{0.2})$

$= 3.105$

<u>Note</u>

$$l_{\sigma k}(s, y) = -\log(\sigma_y(s))$$

$$\sigma_j(s) = e^{s_j} / e^{s_1} + \dots + e^{s_k}$$

$$\Rightarrow l_{\sigma k}(s,y) = \log(e^{s_1} + \dots + e^{s_k}) - \log(e^{s_y})$$

$$= \underbrace{LSE(s)}_{\text{"sofmax(s)"}} - s_y$$

Looks like margin!

<u>Ex</u> $s = (0, 3, 0.2)$ $\quad l_{\sigma k}(s,y) = 3.2 - 3 \doteq 0.2$

$y = 2$ $\quad l_{\sigma k}(0, 30, 0.2) = \dots$ smaller

Important Fn

$$\sigma_j(s) = \frac{e^{s_j}}{e^{s_1 + \dots + s^K}}$$

Claim $\quad \sigma(s) = \nabla_s \, LSE(s)$

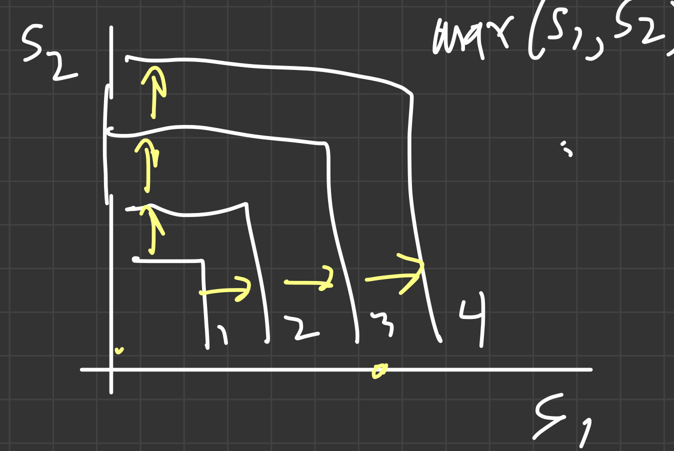
First $\quad K = 2$

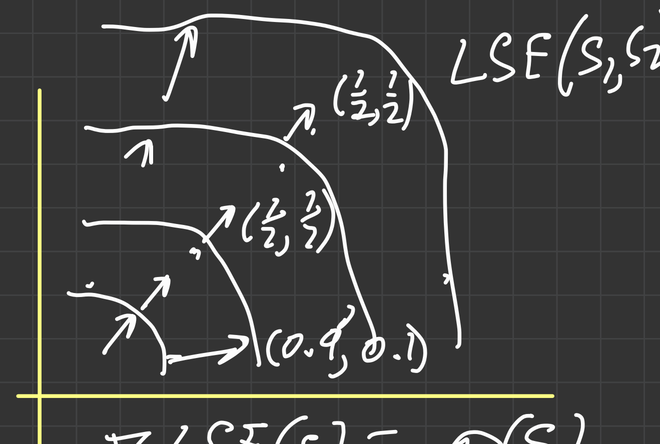
$$f(s) = LSE((0, s)) = \log(e^0 + e^s)$$

$$f'(s) = \frac{e^s}{1 + e^s} = \frac{1}{1 + e^{-s}} = \sigma(s)$$

Next $\quad K \quad f(s) = \log(e^{s_1} + \dots + e^{s_K})$

$$\frac{\partial f}{\partial s_j} = \frac{e^{s_j}}{e^{s_1} + \dots + e^{s_K}}$$





$$\nabla LSE(s) = \sigma(s)$$

Ex $\quad \sigma(0, 10) = \frac{1}{e^0 + e^{10}} (e^0, e^{10})$

$$= (0.00005, 0.99995)$$

$$\cong (0, 1)$$

LSE identity / trick

loss $s_1 = (1, 5, 2)$

$s_2 = (11, 15, 12)$

$s_2 = s_1 + 10$

def $c \in \mathbb{R}$
$s \in \mathbb{R}^k$
$(s+c)_i = (s_i + c)$

margin loss $\ell_{margin}(s_1, y) = \ell_{margin}(s_2, y)$

$\ell_{CE}(s_1, y) \neq \ell_{CE}(s_2, y)$

Note

$$LSE(s+m) = m + LSE(s)$$