# Math 462

## Admin

HW 3 due
Midterm grades
HW 2 grades

Topics
Multiclass.
Generalization
project

20.10.2021
Lecture 13

# From binary to k-classification (multi-class)

Review binary.

$x \in \mathbb{R}^d$

$S_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$ $\qquad \mathcal{Y}_\pm = \{-1, 1\}$

$\ell_{0\text{-}1}(s, y)$ measure errors NOT for training

Want $c(x) \in \mathcal{Y}_\pm$

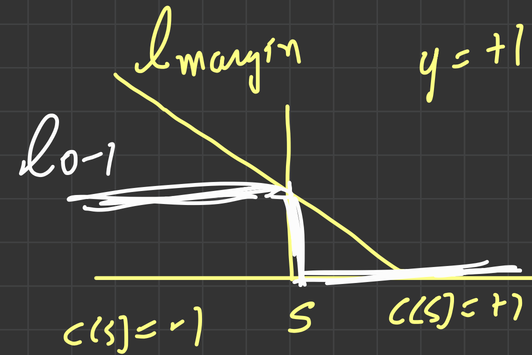
Use $s(x) \in \mathbb{R}^1$

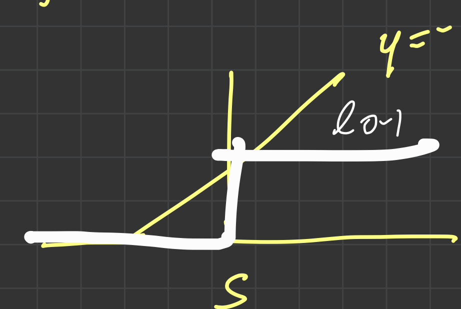
class map $c : \mathbb{R} \to \mathcal{Y}_\pm$

$c(s) = \text{sign}(s)$.

surrogate loss

$$\ell(s, y), \quad s \in \mathbb{R}, \quad y \in \mathcal{Y}_\pm$$




__Ex__ margin loss
log-logistic

__Thm__ $\ell_{class}(s, y) \geq c\, \ell_{0\text{-}1}(c(s), y)$

Binary Class

log-logistic loss (as score-based)

$$\ell_{\log\text{-}\sigma}(x,y) = \begin{cases} -\log\left(\underbrace{\frac{1}{1+e^{-x}}}_{\sigma(x)}\right) & y=+1 \\ -\log\left(1-\sigma(x)\right) & y=-1 \end{cases}$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

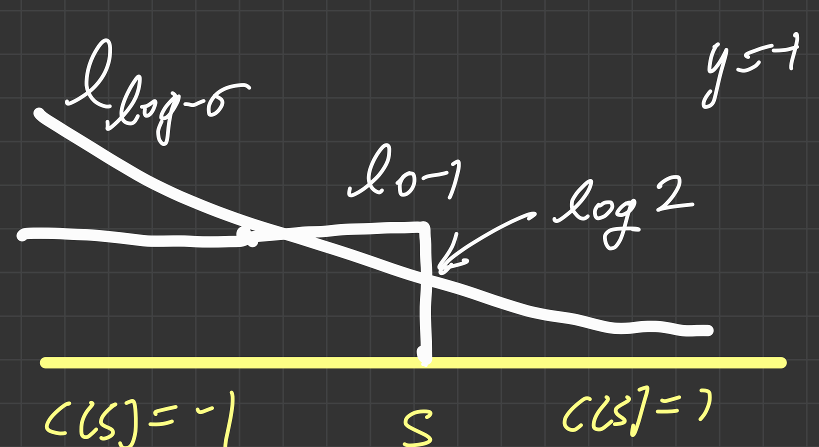
$$= \begin{cases} \log\left(1+e^{-x}\right) & y=+1 \\ \log\left(1+e^{+x}\right) & y=-1 \end{cases}$$

$$l_{\log\text{-}\sigma}(x,y) \doteq \log(1+e^{-yx}) \qquad y=-1,+1$$

Loss

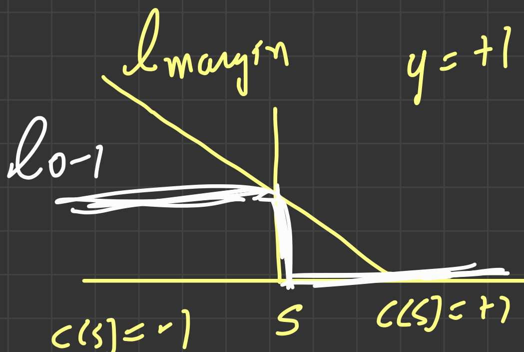
encourage correct

monotone (bigger when worse score)




$$\underbrace{l_{\log\text{-}\sigma}(s,y) \geq \frac{1}{\log 2}\, l_{0\text{-}1}(c(s),y)}_{\text{bound on errors from loss}}$$

# Compare 2 classification losses





<u>note</u> 0 when $s \geq 1$

p.w. linear

$\ell' = \begin{cases} 0 \\ \pm 1 \end{cases}$

signal same

$s_1$ worse $s_2$

$\ell'(s_1) = \ell'(s_2)$

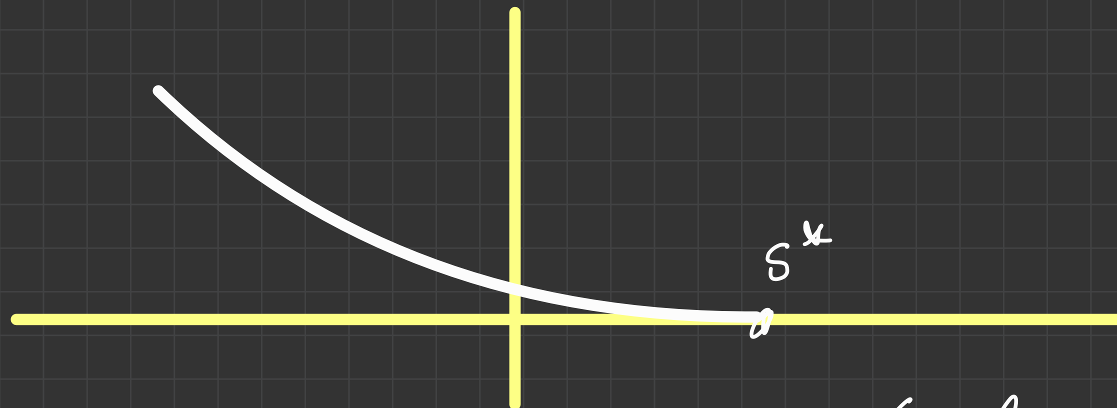
<u>note</u> alway $\geq 0$

diff'ble, convex

$s_1$ worse $s_2$

$\ell'(s_1)$ stronger $\ell'(s_2)$

can we combine losses




shift down & do $\max(0, \ell_{shift})$

properties 0 at $s^*$

$\ell'$ decreasing

$\ell$ strongly convex for $s \leq s^*$.

From binary to K-classification
(multi-class)

Review binary.

$S_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$ $\quad x \in \mathbb{R}^d$

$\mathcal{Y}_\pm = \{-1, 1\}$

$l_{0\text{-}1}(s, y)$ measure errors NOT for training

Want $c(x) \in \mathcal{Y}_\pm$

Use $s(x) \in \mathbb{R}^1$

class map $c: \mathbb{R} \to \mathcal{Y}_\pm$

$c(s) = \mathrm{sign}(s)$.

surrogate loss

$l(s, y)$, $s \in \mathbb{R}$, $y \in \mathcal{Y}_\pm$

EX margin loss
log-logistic

Thm $l_{class}(s, y) \geq c\, l_{0\text{-}1}(c(s), y)$

Linear model

$s(x) = w \cdot x$

K-classes

$x \in \mathbb{R}^d$

$y \in \{1, \dots, K\}$

$S_m$ ✓

$l_{0\text{-}1}$ ✓

$s \in \mathbb{R}^K$ K-scores
one for each class

class map

$c(s) = \mathrm{argmax}\{s_1, \dots, s_K\}$
highest score.

Thm ✓

need $l: \mathbb{R}^K \times \mathcal{Y}_K \to \mathbb{R}^+$

linear

$s(x) = (w_1 \cdot x, \dots, w_K \cdot x)$

Break Regression $y \in \mathbb{R}^K$

Linear Model

$$h(x) = (w_1 \cdot x, \dots, w_K \cdot x) = W \cdot x \qquad W \quad K \times d \text{ matrix}$$

$$w = W$$

$$\hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} \underbrace{(h_w(x_i) - y_i)^2}_{\|h_w(x_i) - y_i\|^2 = \sum_{j=1}^{K} (h^j(x) - y_i^j)^2}$$

$$\nabla_w \hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} (h_w(x_i) - y_i) \underbrace{\nabla_w h_w(x)}_{\begin{pmatrix} x, \\ x, \\ \vdots \\ x \end{pmatrix}}$$

K-class margin loss

2-class $\quad l_{margin}(s,y) = \begin{cases} \max(0, 1-s), & y=1 \\ \max(0, 1+s), & y=-1 \end{cases}$

vector version

$(s_1, s_2) \qquad m(s,y) = \dfrac{s_y - s_{not\,y}}{2}$

$$l_{m\text{-}2}((s_1,s_2),y) = \max(0, 1-m(s,y))$$

Ex $\quad (s_1,s_2) = (s, -s)$ recover binary case

Ex $\quad l(0.6,+1) \quad l(0.6,-1) \quad l_{margin}$
$\quad l(1.2,+1) \quad l(-1.5,-1)$

same for $l_{m\text{-}2}$
$s = (0.6, -0.6)$
etc

$l(0.6,+1) = \max(0, 1-0.6) = 0.4$

$l(0.6,-1) = \max(0, 1+0.6) = 1.6$

$l_{m\text{-}2}((0.6,-0.6),+1)$

$m = 0.6$

same

K-class margin loss

2-class $\ell_{margin}(s,y) = \begin{cases} \max(0, 1-s), & y=1 \\ \max(0, 1+s), & y=-1 \end{cases}$

vector version

$(s_1, s_2)$

$m_2(s,y) = \underbrace{s_y - s_{not\,y}}_{2}$

correct $m \geq 0$

wrong $m \leq 0$

$m \geq 1$ $\ell = 0$

$\Leftarrow \Big\{ \; m_k(s,y) = s_y - \max_{j \neq y} s_j$

$\ell_{m.k}((s_1,s_2), y) = \max(0, 1 - m(s,y))$

Want $\ell(\text{wrong}) \geq 1$

want $\ell(\text{correct}) \leq 1$

$\ell(\text{correct} + \text{marg}) = 0$

Dragoș.

$$\frac{(K-1)}{K} S_y - \frac{\sum_{j \neq y} S_j}{K} = m$$

$y = 1$

$S_1 = 3$ $\quad S_2 \le 3$ $\quad S_3 \le 3$

correct

Good example of loss design challenges.

$S_j \le S_y \quad j \le y$ ✓

$\Rightarrow m \geq 0$

$S = (3, 2.9, 1.3) \quad m \geq 0$

$y = 1$

$$S = (3, \underbrace{4, 0.1, 0.1}_{4/3}) \qquad 3 - 4/3 < 0$$

but incorrect.

$y = 1$

$C(S) = 2$

Loss

$$\hat{L}(w) = \frac{1}{m}\sum_{i=1}^{m}\left(\log(h_w(x_i)) - \log(y_i)\right)^2$$

paper regression problem $h, y \in \mathbb{R}^K$

instead of using

$$l_2(h,y) = (h-y)^2$$

because $y_i$ large range of values.

so $(\log(h) - \log(y))^2$ better.

prove $\hat{L}_2(w)$: $(w, y_i) \Rightarrow \text{mean}(y_i)$

using $h_w = w$ $w^* = AM(y_i)$. $h_w = w$ $l_2$-log loss

$l_2$ loss $w^* - GM(y_i)$

$GM \neq AM$

$$GM(y_i) = \left(\prod_{i=1}^{m} y_i\right)^{1/m}$$

$$AM(y_i) = \frac{1}{m}\sum y_i$$

Mot. GM:

wikipedia

K-class Logistic

2 class $\quad \ell_{\sigma-2}(s,y) = \log\left(\underbrace{1+e^{-sy}}\right)$

$y=1 \quad = -\log(\sigma(s))$

$C(s) = \underset{i}{\arg\max}\ s_i$

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

$s \in \mathbb{R}^k$ generalisation

$$\sigma(s_1,\dots,s_k) = \frac{1}{\sum_{i=1}^{k} e^{s_i}}\left(e^{s_1}, e^{s_2}, \dots, e^{s_k}\right)$$

K-class loss:

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

$$\left(\sigma_y(s) = e^{s_y} / \sum e^{s_i}\right)$$

EX Given s recover $\ell_{\sigma 2}$ from $(s_1,s_2) = (\quad)$

$(s_1,s_2) = (s,0)$

Given s Find $s_1, s_2$

$$\ell_{\log\sigma}((s_1,s_2),1) = \ell_{\sigma 2}(s)$$

EX $s = 3$, $y=1$ loss $-\log\left(\frac{e^s}{1+e^s}\right)_{s=3} \xrightarrow{s_1,s_2} -\log\left(\frac{e^{s_1}}{e^{s_1}+e^{s_2}}\right)$

what $\ell_{\log\text{-}\sigma}(s,y) \geq \frac{1}{c}\ell_{0\text{-}1}(c(s),y)$

Yes $\quad c = \log k$

loss $= \log\left(\underbrace{\frac{e^{s_1}+\cdots+e^{s_k}}{e^{s_y}}}\right)$

$$\left(1+\sum_{j\neq y} e^{s_j - s_y}\right)$$

define

$m_j = s_j - s_y$

① $c(s) = y \Rightarrow m_j \leq 0 \Rightarrow e^{m_j} \leq 1$

$\Rightarrow$ loss $\leq \log(1+k-1) = \log(k)$