MULTI-CLASS ADA BOOST CORRECTION

R. AZAÏS
(2024)

1 $$g^*(x) = \underset{k \in \mathcal{Y}}{\operatorname{argmax}} \; \pi_k(x)$$

2 $$\mathcal{L}(f,\lambda) = \mathbb{E}\left[L(Z,f) \mid X=x \right] - \lambda \sum_{k=1}^{K} f_k$$

$$= \mathbb{E}\left[\exp\left(-\frac{1}{K} \sum_{k=1}^{K} Z_k f_k \right) \middle| X=x \right] - \lambda \sum_{k=1}^{K} f_k$$

$$= \mathbb{E}\left[\prod_{k=1}^{K} \exp\left(-\frac{1}{K} Z_k f_k \right) \middle| X=x \right] - \lambda \sum_{k=1}^{K} f_k$$

$$= \sum_{j=1}^{K} \pi_j(x) \exp\left(-\frac{f_j}{K} + \sum_{k \neq j} \frac{f_k}{K(K-1)} \right) - \lambda \sum_{k=1}^{K} f_k$$

Or $$\sum_{k \neq j} \frac{f_k}{K(K-1)} = \frac{-f_j}{K(K-1)} \quad \left(\text{avec } \sum_{k=1}^{K} f_k = 0 \right)$$

Et $-\frac{\delta_j}{K} - \frac{\delta_j}{K(K-1)} = \frac{-\delta_j K + \delta_j - \delta_j}{K(K-1)} = \frac{-\delta_j}{K-1}$

Donc $\mathcal{L}(\delta, \lambda) = \sum_{j=1}^{K} \pi_j(x) \exp\left(-\frac{\delta_j}{K-1}\right)$

$$- \lambda \sum_{j=1}^{K} \delta_k$$

Alors

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \delta_k} = -\frac{\pi_k(x)}{K-1} \exp\left(-\frac{\delta_k}{K-1}\right) - \lambda \\ \frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=1}^{K} \delta_k \end{cases}$$

Par conséquent, $\frac{\partial \mathcal{L}}{\partial \delta_k} = 0$ nous mène

à $\exp\left(-\frac{\delta_k}{K-1}\right) = \frac{-\lambda(K-1)}{\pi_k(x)}$

i.e. $\delta_k = (K-1) \ln(\pi_k(x)) - (K-1) \ln(-\lambda(K-1))$

Avec $\dfrac{\partial \mathcal{L}}{\partial \lambda} = 0$, on a aussi

$$(K-1)\ln(-\lambda(K-1)) = (K-1)\sum_{j=1}^{K} \ln(\pi_j(x))$$

Finalement,

$$f_k = (K-1)\left[\ln \pi_k(x) - \sum_{j=1}^{K} \ln(\pi_j(x))\right]$$

On trouve $\boxed{f_k^*(x) = (K-1)\left[\ln \pi_k(x) - \sum_{j=1}^{K} \ln \pi_j(x)\right]}$

---

3 Immédiatement,

$$\exp\left(\frac{f_k^*}{K-1}\right) = \pi_k(x) \times \underbrace{\exp\left(-\sum_{j=1}^{K} \ln \pi_j(x)\right)}_{= \text{cste}}$$

Donc $\boxed{\pi_k(x) \propto \exp\left(\frac{f_k^*}{K-1}\right)}$

Et $\boxed{g^*(x) = \underset{k \in \mathcal{Y}}{\operatorname{argmax}}\ f_k^*(x)}$

4

$$\hat{L}(\beta, h) = \sum_{i=1}^{m} \omega^{(i)} L\left(z^{(i)}, \beta h(x^{(i)})\right)$$

$$= \sum_{i=1}^{m} \omega^{(i)} \exp\left(-\frac{\beta}{K} \sum_{k=1}^{K} z_k^{(i)} h(x^{(i)})_k\right)$$

- Si $z^{(i)} = h(x^{(i)})$, alors

$$\sum_{k=1}^{K} z_k^{(i)} h(x^{(i)})_k = 1 + \frac{K-1}{(K-1)^2}$$

$$= \frac{K-1+1}{K-1}$$

$$= \frac{K}{K-1}$$

- Si $z^{(i)} \neq h(x^{(i)})$, alors

$$\sum_{k=1}^{K} z_k^{(i)} h(x^{(i)})_k = \frac{-2}{K-1} + \frac{K-2}{(K-1)^2}$$

$$= \frac{-2K+2+K-2}{(K-1)^2}$$

$$= -\frac{K}{(K-1)^2}$$

Finalement,

$$\hat{L}(\beta, h) = \sum_{z^{(i)} = h(x^{(i)})} \omega^{(i)} \exp\left(-\frac{\beta}{K-1}\right)$$

$$+ \sum_{z^{(i)} \neq h(x^{(i)})} \omega^{(i)} \exp\left(\frac{\beta}{(K-1)^2}\right)$$

$$= \exp\left(-\frac{\beta}{K-1}\right) \sum_{i=1}^{n} \omega^{(i)}$$

$$+ \sum_{i=1}^{n} \omega^{(i)} \mathbb{1}_{\{z^{(i)} \neq h(x^{(i)})\}} \left[\exp\left(\frac{\beta}{(K-1)^2}\right) - \exp\left(-\frac{\beta}{K-1}\right)\right]$$

Le minimum en h pour $h \in \mathcal{H}$ est obtenu en

$$\boxed{h^* = \underset{h \in \mathcal{G}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}_{\{z^{(i)} \neq h(x^{(i)})\}}}$$

(tant que $\exp\left(\frac{\beta}{(K-1)^2}\right) - \exp\left(-\frac{\beta}{K-1}\right) > 0$; c'est bien le cas pour $\beta > 0$)

On pose $\mathcal{E}^* = \dfrac{\sum_{i=1}^{m} \omega^{(i)} \mathbb{1}_{\{z^{(i)} \neq h^*(x^{(i)})\}}}{\sum_{i=1}^{m} \omega^{(i)}}$

Alors

$$\frac{\mathcal{L}(\beta, h^*)}{\sum_{i=1}^{m} \omega^{(i)}} = \exp\left(-\frac{\beta}{K-1}\right) + \mathcal{E}^* \left[\exp\left(\frac{\beta}{(K-1)^2}\right) - \exp\left(-\frac{\beta}{K-1}\right)\right]$$

$$= \exp\left(-\frac{\beta}{K-1}\right)(1-\mathcal{E}^*) + \mathcal{E}^* \exp\left(\frac{\beta}{(K-1)^2}\right)$$

On dérive par rapport à β :

$$-\frac{1-\mathcal{E}^*}{K-1} \exp\left(-\frac{\beta}{K-1}\right) + \frac{\mathcal{E}^*}{(K-1)^2} \exp\left(\frac{\beta}{(K-1)^2}\right) = 0$$

$$\frac{1-\mathcal{E}^*}{\mathcal{E}^*} = \frac{1}{K-1} \exp\left(\frac{\beta}{K-1} \underbrace{\left[1 + \frac{1}{K-1}\right]}_{\frac{K}{K-1}}\right)$$

Donc $\boxed{\left[\ln\left(\frac{1-\varepsilon^*}{\varepsilon^*}\right) + \ln(k-1)\right]\frac{(k-1)^2}{k} = \beta^*}$

---

5 On a

$$\sum_{k=1}^{K} z_k^{(i)} f(x^{(i)})_k = \sum_{m=1}^{M} \beta_m \sum_{k=1}^{K} z^{(i)} h_m(x^{(i)})_k$$

Donc
$$L(z^{(i)}, f) = \prod_{m=1}^{M} L\left(z^{(i)}, \beta_m h_m(x^{(i)})\right)$$

$$= L\left(z^{(i)}, \beta_M h_M(x^{(i)})\right) \times \underbrace{\prod_{m=1}^{M-1} L\left(z^{(i)}, \beta_m h_m(x^{(i)})\right)}_{\omega_M^{(i)}}$$

Finalement

$$\boxed{\sum_{i=1}^{n} L(z^{(i)}, f) = \sum_{i=1}^{n} \omega_M^{(i)} L\left(z^{(i)}, \beta_M h_M(x^{(i)})\right)}$$

Et $\boxed{\omega_M^{(i)} = \omega_{M-1}^{(i)} \times L\left(Z^{(i)}, \beta_{M-1} h_{M-1}(X^{(i)})\right)}$

avec $\omega_0^{(i)} = 1$ et $h_0 = 0$

---

6 $\sum_{i=1}^{m} L(Z^{(i)}, f)$ est la version empirique de la fonction minimisée en $\underline{3}$ pour définir f^*!

Pour m entre 1 et M:

- $\omega_m^{(i)} = \omega_{m-1}^{(i)} L\left(Z^{(i)}, \beta_{m-1} h_{m-1}(X^{(i)})\right)$

(En particulier $\omega_1^{(i)} = 1$)

- $h_m = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{m} \omega_m^{(i)} \mathbb{1}_{\{h(X^{(i)}) \neq Z^{(i)}\}}$

- $\varepsilon_m = \dfrac{\sum_{i=1}^{m} \omega_m^{(i)} \mathbb{1}_{\{h_m(X^{(i)}) \neq Z^{(i)}\}}}{\sum_{i=1}^{m} \omega_m^{(i)}}$

- $\alpha_m = \ln \dfrac{1-\varepsilon_m}{\varepsilon_m} + \ln K-1$

- $\beta_m = \dfrac{(K-1)^2}{K} \alpha_m$

Retourner $f = \sum_{m=1}^{M} \beta_m h_m$

Rq : Vérifier que $\sum_{k=1}^{K} f_k = 0$ ...

---

7 Le classifieur induit par <u>3</u> et <u>6</u> est

$$H(x) = \underset{k \in \mathcal{Z}}{\operatorname{argmax}} \; f_k(x)$$

$$= \underset{k \in \mathcal{Z}}{\operatorname{argmax}} \; \sum_{m=1}^{M} \beta_m \, h_m(x)_k$$

Si $g_m : \mathbb{R}^d \to \mathcal{Y}$ est le classifieur associé à h_m, i.e.

$$g_m(x) = k \text{ ssi } h_m(x)_k = 1$$

Alors $\beta_m h_m(x)_k = \beta_m \mathbb{1}_{\{g_m(x)=k\}}$

$$- \frac{\beta_m}{K-1} \underbrace{\mathbb{1}_{\{g_m(x)\neq k\}}}_{1-\mathbb{1}_{\{g_m(x)=k\}}}$$

$$= \beta_m \underbrace{\left(1+\frac{1}{K-1}\right)}_{\frac{K}{K-1}} \mathbb{1}_{\{g_m(x)=k\}}$$

$$- \frac{\beta_m}{K-1}$$

Donc le classifieur $G : \mathbb{R}^d \to \mathcal{Y}$ associé à $H : \mathbb{R}^d \to \mathcal{Z}$ est donné par

$$G(x) = \underset{k \in \mathcal{Y}}{\operatorname{argmax}} \sum_{m=1}^{M} \beta_m \mathbb{1}_{\{g_m(x)=k\}}$$

$$= \underset{k \in \mathcal{Y}}{\operatorname{argmax}} \sum_{m=1}^{M} \alpha_m \mathbb{1}_{\{g_m(x)=k\}}$$

car $\alpha_m \propto \beta_m$

8 $\alpha_m = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right) + \ln(K-1)$

terme qui généralise
de manière non-triviale
AdaBoost

Rq: On aurait pu prendre en multi-classe
$$\alpha_m = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right)$$
comme pour le problème à 2 classes.

EXERCICE BASÉ SUR MULTI-CLASS ADABOOST, par Zhu et al. (Statistics and its Interface, 2009)
