Homework 3

Fondations of Machine Learning

Xiang Pan (xp2030)

April 5, 2022

Boosting

1

(a)

$$\ell(yh(x), r(x)) = \begin{cases} 1_{yh(x)\le 0}, & r(x) > 0 \\ c, & r(x) \le 0 \end{cases} \tag{1}$$

where c is a positive constant less than 1/2. For simplicity, define $b = 2\sqrt{\frac{1-c}{c}}$.

$$\Psi_1(yh(x), r(x)) = \max\left\{e^{r(x)-yh(x)}, ce^{-br(x)}\right\} \tag{2}$$

$$\Psi_2(yh(x), r(x)) = e^{r(x)-yh(x)} + ce^{-br(x)} \tag{3}$$

Show that Ψ_1 is convex in $(yh(x), r(x))$ and it upper-bounds ℓ. Show that Ψ_2 is convex in $(yh(x), r(x))$ and it upper-bounds Ψ_1.

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \tag{4}$$

Solution:
Bounding:

$$\Psi_1(yh(x), r(x)) = \max\left\{e^{r(x)-yh(x)}, ce^{-br(x)}\right\} \tag{5}$$

if $r(x) > 0$ and $yh(x) > 0$, then $\ell(x) = 0$.
We have $\ell(x) = 0 < \Phi_1(x)$
if $r(x) > 0$ and $yh(x) \le 0$, then $\ell(x) = 1$. $r(x) > yh(x)$, $e^{r(x)-yh(x)} > 1$.
We have $\ell(x) = 1 < \Phi_1(x)$
if $r(x) \le 0$, then $\ell(x) = c$.

$$P(c) = ce^{-br(x)} \tag{6}$$

$$P(c) = ce^{-br(x)} = ce^{-2\sqrt{\frac{1-c}{c}}r(x)} \geq c \tag{7}$$

We have $\ell(x) = c \leq \Phi_1(x)$
We always have the two components of Ψ_1 are positive, thus $\Psi_1 < \Psi_2$.
Thus, we have proved $\ell(x) < \Psi_1(x) < \Psi_2(x)$.
We let $\Psi_1(x) = \Psi_1(x)$
Convex:

(b)

$$\Psi_{1,\mathcal{F}} = \{(x, y) \mapsto \min\{\Psi_1(y\mathbf{h}(x), \mathbf{r}(x)), 1\}, (\mathbf{h}, \mathbf{r}) \in \mathcal{F}\}$$

We denote the set $S = \{(h_1, r_1), (h_2, r_2), \cdots, (h_N, r_N)\}$

$$Q(t) = \min\{t, 1\}, \quad t > 0 \tag{8}$$

$$\psi = Q \circ \phi \tag{9}$$

$$\mathfrak{R}_m(\Psi_{1,\mathcal{F}}) \leq \mathfrak{R}_m(\mathcal{H}) + (b+1)\mathfrak{R}_m(\mathcal{R})$$

We hide x for simplicity.
Frist we prove that $\mathcal{H}$ is convex.

$$| Q \cdot \phi_1(h_i, r_i) - Q \cdot \phi_2(h_j, r_j) | \leq | (h_i(x) - h_j(x)) + (b+1)[r_i(x) - r_j(x)] | \tag{10}$$

$$\begin{aligned}
\mathfrak{R}_m(\Psi_{1,\mathcal{F}}) &= \mathfrak{R}_m(\Psi_{1,\mathcal{S}}) && \text{(Convex Hull Theorem)} && (11)\\
&\leq \mathfrak{R}_m(\mathcal{H}) + (1+b)\mathfrak{R}_m(\mathcal{R}) && \text{(Talagrand's lemma)} && (12)
\end{aligned}$$

$$\frac{1}{m}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{h \in H} \sum_{i=1}^{m} \sigma_i h(x_i)\right] \tag{13}$$

$$\frac{1}{m}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{r \in R} \sum_{i=1}^{m} \sigma_i r(x_i)\right] \tag{14}$$

References

[1] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Multi-class classification with maximum margin multiple kernel. In *International Conference on Machine Learning*, pages 46–54. PMLR, 2013.

[2] Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of machine learning research*, 2(Dec):265–292, 2001.