A Proof of Lemma 3

Proof. From Theorem 1, we know that $\psi(\theta)$ is a convex function. It's also known that any convex function on a convex open subset of $\mathbb{R}^n$ is semi-differentiable. Thus, we can denote the right derivative of $\psi(\theta)$ at $\theta = 0$ by $\partial_+\psi(0)$. Using Maclaurin expansion, we have,

$$\psi(\theta) = \psi(0) + \partial_+\psi(0) * \theta + o(\theta) = \partial_+\psi(0) * \theta + o(\theta) \; . \tag{47}$$

Followed by Proposition 1, if $\partial_+\psi(0) \neq 0$, we have,

$$M = \partial_+\psi(0) \quad \text{and} \quad \alpha = 1 \; . \tag{48}$$

If $\partial_+\psi(0) = 0$, then $\psi(\theta) = o(\theta)$, which means that $\psi(\theta)$ is the infinitesimal of higher order than θ as $\theta \to 0+$. Then, by definition, for any given ϕ, we can compute ψ. Because $\psi(\theta)$ is the higher order infinitesimal of θ as $\theta \to 0+$, there exist unique $\alpha > 1$ and $M \in \mathbb{R}^+$ such that,

$$\lim_{\theta \to 0+} \frac{\psi(\theta)}{M\theta^\alpha} = 1 \tag{49}$$

which completes the proof. □

B Proof of Lemma 4

Proof. Let $\psi^{-1}(\mu) = \theta$, then $\mu = \psi(\theta)$. From Theorem 1, we know that ψ is monotonic increasing within $[0, 1]$, and $\psi(0) = 0$. Thus,

$$\mu \to 0+ \quad \text{implies} \quad \theta \to 0+ \; .$$

We have,

$$\lim_{\mu \to 0+} \frac{\psi^{-1}(\mu)}{S\mu^I} = \lim_{\theta \to 0+} \frac{\theta}{S(\psi(\theta))^I} \; . \tag{50}$$

By substituting the definitions of S and I, the right hand side of (50) becomes,

$$\lim_{\theta \to 0+} \frac{M^{\frac{1}{\alpha}}\theta}{(\psi(\theta))^{\frac{1}{\alpha}}} = \left[\lim_{\theta \to 0+} \frac{M\theta^\alpha}{\psi(\theta)}\right]^{\frac{1}{\alpha}} = 1 \; . \tag{51}$$

The last equality follows (21) in Lemma 3, which completes the proof. □

C Proof of Lemma 5

Proof. From Proposition 1, we have that there exists $0 < A, B < +\infty$ such that,

$$\mathcal{O}\left(\frac{1}{n^p}\right) = A\frac{1}{n^p} + o\left(\frac{1}{n^p}\right) \tag{52}$$

and

$$\mathcal{O}\left(\psi^{-1}(\frac{1}{n^p})\right) = B\psi^{-1}\left(\frac{1}{n^p}\right) + o\left(\psi^{-1}(\frac{1}{n^p})\right) \; . \tag{53}$$

Substituting (52) and (53) into (23), it's equivalent to proving that for any $0 < A < +\infty$, there exists $0 < B < +\infty$, such that,

$$\psi^{-1}\left(A\frac{1}{n^p} + o(\frac{1}{n^p})\right) = B\psi^{-1}\left(\frac{1}{n^p}\right) + o\left(\psi^{-1}(\frac{1}{n^p})\right) \; . \tag{54}$$

To prove (54), by Proposition 1, we only need to prove that, for any $0 < A < +\infty$, there exists $0 < B < +\infty$, such that,

$$\lim_{n \to +\infty} \frac{\psi^{-1}(A\frac{1}{n^p} + o(\frac{1}{n^p}))}{\psi^{-1}(\frac{1}{n^p})} = B \; . \tag{55}$$

Followed by Lemma 4 and proposition 1, we have

$$\psi^{-1}(\mu) = S\mu^I + o(\mu^I) \; . \tag{56}$$

Substituting (56) into (55), we have,

$$\begin{aligned}
&\lim_{n \to +\infty} \frac{\psi^{-1}(A\frac{1}{n^p} + o(\frac{1}{n^p}))}{\psi^{-1}(\frac{1}{n^p})} \\
&= \lim_{n \to +\infty} \frac{S(A\frac{1}{n^p} + o(\frac{1}{n^p}))^I + o((A\frac{1}{n^p} + o(\frac{1}{n^p}))^I)}{S(\frac{1}{n^{p*I}}) + o((\frac{1}{n^{p*I}}))} \\
&= \lim_{n \to +\infty} \frac{S(A^I\frac{1}{n^{p*I}} + o(\frac{1}{n^{p*I}}))}{S(\frac{1}{n^{p*I}})} \\
&= A^I \; .
\end{aligned} \tag{57}$$

This means that for any $0 < A < +\infty$, there exists $B = A^I \in (0, +\infty)$ such that (55) holds true, which completes the proof. □

D Proof of Theorem 2

Proof. From Theorem 1, we have that

$$R(f_n) - R^* \leq \psi^{-1}(R_\phi(f_n) - R_\phi^*) \tag{58}$$

where f_n is the minimizer of the empirical surrogate risk $\hat{R}_\phi$ with sample size n. Under our assumption that the Bayes optimal classifier is within the hypothesis class $\mathcal{F}$, then, with a high probability, we have,

$$R_\phi(f_n) - R_\phi^* \leq \mathcal{O}\left(\frac{1}{n^p}\right) \; . \tag{59}$$

Note that p is often equal to $1/2$ for the worst cases. With (59) and Lemma 5, we have,

$$\begin{aligned}
&\psi^{-1}(R_\phi(f_n) - R_\phi^*) \\
&\leq \psi^{-1}\left(\mathcal{O}(\frac{1}{n^p})\right) = \mathcal{O}\left(\psi^{-1}(\frac{1}{n^p})\right) \; .
\end{aligned} \tag{60}$$

Substituting (56) into (60), we have,

$$\begin{aligned}
&\mathcal{O}\left(\psi^{-1}(\frac{1}{n^p})\right) = \mathcal{O}\left(S(\frac{1}{n^p})^I + o((\frac{1}{n^p})^I)\right) \\
&= \mathcal{O}\left(\frac{S}{n^{pI}}\right) \; ,
\end{aligned} \tag{61}$$

which completes the proof. □

E Proof of Theorem 3

Proof. We have that $0 < \alpha < +\infty$, so $I > 0$ holds trivially. To prove $I = \frac{1}{\alpha} \leq 1$, it is equivalent to proving that there exists $0 \leq C < +\infty$ such that,

$$\psi(\theta) = C\theta + o(\theta) \; , \tag{62}$$

because $I < 1$ holds true if and only if $C = 0$; and $I = 1$ holds true if and only if $0 < C < +\infty$. From proposition 1, the equation (62) implies,

$$\lim_{\theta \to 0+} \frac{\psi(\theta)}{\theta} = C \; . \tag{63}$$

Since any convex function on a convex open subset in $\mathbb{R}^n$ is semi-differentiable, $\psi(\theta)$ is at least right differentiable at $\theta = 0$. We therefore have,

$$\lim_{\theta \to 0+} \frac{\psi(\theta)}{\theta} = \lim_{\theta \to 0+} \frac{\psi(\theta) - \psi(0)}{\theta - 0} = \partial_+ \psi(0) = C \tag{64}$$

where $\partial_+ \psi(0)$ denotes the right derivative of $\psi(\theta)$ at $\theta = 0$. The proof ends. □

F Proof of Theorem 4

Proof. Following Proposition 1 and Lemma 3, we have,

$$\psi_i(\theta) = M_i \theta^{\alpha_i} + o(\theta^{\alpha_i}) \quad \text{for} \quad i = 1, 2 \quad \text{and} \quad 0 < M_i < \infty \tag{65}$$

Then, we get,

$$\lambda = \lim_{\theta \to 0+} \frac{M_1 \theta^{\alpha_1} + o(\theta^{\alpha_1})}{M_2 \theta^{\alpha_2} + o(\theta^{\alpha_2})} = \frac{M_1}{M_2} \lim_{\theta \to 0+} \theta^{\alpha_1 - \alpha_2} \tag{66}$$

Thus, we can conclude:

- for $\lambda = \frac{M_1}{M_2} \in (0, +\infty)$, then we have $\alpha_1 = \alpha_2$ and $I_1 = I_2$. Therefore, the minimizers w.r.t. ϕ_1 and ϕ_2 converge equally fast to the Bayes optimal classifier;
- for $\lambda = +\infty$, we have $\alpha_1 < \alpha_2$ and $I_1 > I_2$. Thus the minimizer w.r.t. ϕ_1 converges faster to the Bayes optimal classifier;
- for $\lambda = 0$, then we have $\alpha_1 > \alpha_2$ and $I_1 < I_2$, which means that the minimizer w.r.t. ϕ_2 converges faster to the Bayes optimal classifier.

□

G Proof of Theorem 5

Proof. Notice that $\psi(R(f_n) - R^*) \leq R_\phi(f_n) - R^*_\phi$. If we scale ϕ as $\tilde{\phi}(z) = k_2 \phi(z)$, the both sides of the inequality will be multiplied by k_2, which holds trivially.
We now consider $\tilde{\phi}(z) = \phi(k_1 z)$. Observe that,

$$\begin{aligned} \tilde{C}^*(\eta) &= \inf_{z \in \mathbb{R}} \eta \phi(k_1 z) + (1 - \eta) \phi(-k_1 z) \\ &= \inf_{k_1 z \in \mathbb{R}} \eta \phi(k_1 z) + (1 - \eta) \phi(-k_1 z) \\ &= \inf_{z' \in \mathbb{R}} \eta \phi(z') + (1 - \eta) \phi(-z') = C^*(\eta) \end{aligned} \tag{67}$$

where $z' = k_1 z$. Then,

$$\tilde{\psi}(\theta) = \tilde{\phi}(0) - \tilde{C}^*(\eta) = \phi(0) - C^*(\eta) = \psi(\theta) \; , \tag{68}$$

which leads to the same I. □