1.

2.

3. (a) Theorem 7.4 in UML tells us that

$$|L_{\mathcal{P}}(h) - \hat{L}_D(h)| \leq \min_{n:h\in\mathcal{H}_n} \epsilon_n(m, w(n)\cdot\delta) \quad \forall h \in \mathcal{H}$$

with probability at least $1-\delta$ If we assume that $\mathcal{H}$ is countable then we can consider it to be a union of singletons such that

$$\epsilon(h) = \sqrt{\frac{\log(2/(w(h)\cdot\delta))}{2m}} \quad \forall h \in \mathcal{H}, \sum_{h\in\mathcal{H}} w(h) = 1.$$

$epsilon(h)$ is minimized when all of the weight is concentrated on one function $h' \in \mathcal{H}$ such that $w(h') = 1$. In this case $e(h) = \sqrt{\frac{\log(2/\delta)}{2m}}$. Thus if we replace δ with $\delta/2$, set $m = \alpha N$, and let $D = \mathcal{V}$, it follows that

$$|L_P(h) - \hat{L}_{\mathcal{V}}(h)| \leq \sqrt{\frac{\log(4/\delta)}{2\alpha N}} \quad \forall h \in \mathcal{H}$$

with probability at leat $1-\delta/2$.

(b) h^{SRM} is an element of $\mathcal{H}$ and so, by definition, $\hat{L}_{\mathcal{V}}(h^{Val}) \leq \hat{L}_{\mathcal{V}}(h^{SRM})$. The result from part (a) then tells us that

$$\hat{L}_{\mathcal{V}}(h^{Val}) \leq \hat{L}_{\mathcal{V}}(h^{SRM}) \leq L_{\mathcal{P}}(h^{SRM}) + \sqrt{\frac{\log(4/\delta)}{2\alpha N}}$$

with probability at least $1-\delta/2$.

(c) We can combine the previous result with part (a) to get

$$L_{\mathcal{P}}(h^{Val}) \leq \hat{L}_{\mathcal{V}}(h^{Val}) + \sqrt{\frac{\log(4/\delta)}{2\alpha N}} \leq L_{\mathcal{P}}(h^{SRM}) + \sqrt{\frac{\log(4/\delta)}{2\alpha N}} + \sqrt{\frac{\log(4/\delta)}{2\alpha N}}$$

with probability at least $(1-\delta/2)$. This in turn implies that

$$\left|L_{\mathcal{P}}(h^{Val}) - L_{\mathcal{P}}(h^{SRM})\right| \leq 2\sqrt{\frac{\log(4/\delta)}{2\alpha N}}$$

with probability at least $1-\delta$.