Problem 1

Solution

(a) Let $t = \sqrt{1+x}$. Since $x \geq 0$, we have $t \geq 1$. Let $h(t) = 2\log(t) - t$, then we have to prove that $h(t) \leq 0$ holds for all $t \geq 1$. Since $h'(t) = \frac{2}{t} - 1$, we have $h(t) \leq h(2) < 0$.

(b) Let $h(m) = \frac{\log(m)}{m}, m \geq 1$. We have $h'(m) = \frac{1-\log(m)}{m^2}$. Thus, $h(m)$ is decreasing when $m \geq e$. If $n = 1$, then $m < \log(m) \implies m < 0$. (Acutually, the inequality on left hand is not possible for all m, and thus it can imply any statement.) Similarly, it is trivial if $m = 1$. For the cases where $n \geq 2$ and $m \geq 2$, we proceed by contradiction. Suppose that $m \geq 2n\log(n) > e$. Since $\frac{m}{\log(m)}$ is increasing when $m \geq e$ and $n \geq 2\log(n)$ (conclusion of (a)), we have $\frac{m}{\log(m)} \geq \frac{2n\log(n)}{\log(2n\log(n))} \geq \frac{2n\log(n)}{\log(n^2)} = n$. Thus $m < n\log(m) \implies m < 2n\log(n)$.

Problem 2

Solution

The Lagrangian of this problem is

$$L(w, b, \xi, \mu, \lambda) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\mu_i(1 - \xi_i - y_i(w^\intercal x_i + b)) - \sum_{i=1}^{n}\lambda_i\xi_i,\ \mu_i \geq 0 \ and\ \lambda_i \geq 0.$$

Let the gradient of L w.r.t w, b and ξ be 0. We have

$$\nabla_w L = w - \sum_{i=1}^{n}\mu_i y_i x_i = 0,$$

$$\nabla_b L = \sum_{i=1}^{n}\mu_i y_i = 0,$$

$$\nabla_{\xi_i} L = C - \lambda_i - \mu_i = 0,\ \forall i.$$

Since $C = \lambda_i - \mu_i = 0,\ \forall i$ and $\lambda_i \geq 0$, we have $\mu_i \leq C$. By substituting $w = \sum_{i=1}^{n}\mu_i y_i x_i$, we have $D(\mu, \lambda) = \sum_{i=1}^{n}\mu_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j y_i y_j (x_i^\intercal x_j)$. Thus the dual problem is of the same form as the soft-margin version with additional constraints that $\mu_i \leq C,\ \forall i$.

Problem 3

Solution

Let $L_t(\alpha) = \sum_{i=1}^{n}\exp\left(-\sum_{s=1}^{t-1}\alpha_s y_i h_s(x_i) - \alpha y_i h_t(x_i)\right)$. Then, $\alpha_t = \arg\min L_t(\alpha)$. Suppose that $h_{t+1} = h_t$. Then, $L_{t+1}(\alpha) = L_t(\alpha + \alpha_t)$. Thus, $\alpha_{t+1} = 0 = \arg\min L_{t+1}(\alpha)$. However, since $\epsilon_t < \frac{1}{2}$, $\alpha_{t+1} = \frac{1}{2}\ln\frac{1-\epsilon_{t+1}}{\epsilon_{t+1}} > 0$. Thus, by contradiction, $h_{t+1} \neq h_t$.

Problem 4

Solution

(a) 91%.

(b) As shown in Fig. 1, test accuracy improves as number of learners increases.

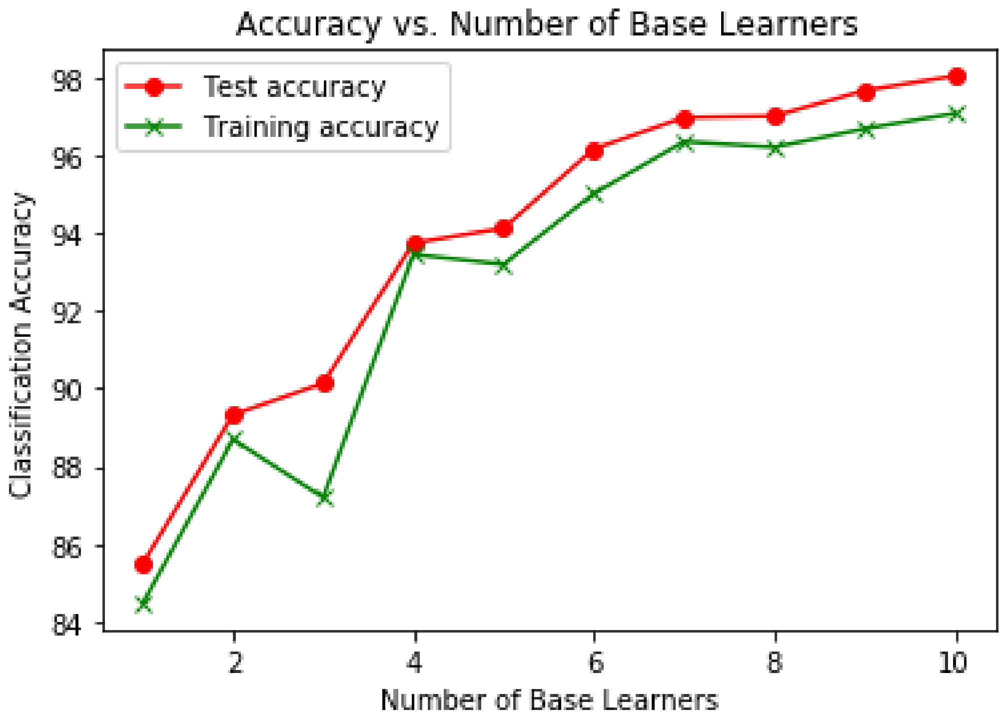

Figure 1: Results of AdaBoost.

Problem 4

Solution

(a) Accuracies are $50\%, 50\%, 52\%, 69\%, 97\%, 98\%$ respectively.

(b) Linear kernel: 98%. Poly kernel: 99%. RBF kernel: 100%. Sigmoid kernel: 81%.