EECS 553 HW3

Lingqi Huang

September 2024

1 Problem 1

(a):Notice that if X_i is misclassified, then the true label and the predicted label are distinct, so we must have

$$1 - \epsilon_i \leq y_i(w^T x_i + b) \leq 0$$

and so we conclude that $\epsilon_i \geq 1$.

(b): We first denote that margin be the plane such that all x_i satisfies

$$y_i(w^T x_i + b) = 1$$

Now given any $\tilde{x}_i$ that does not on the margin, we must have

$$y_i(w^T \tilde{x}_i + b) = 1 - \epsilon_i$$

Now notice we can rewrite the above formula to be

$$y_i(w^T \tilde{x}_i + b) = y_i(w^T x_i + b) - \epsilon_i$$

Combine terms we will then conclude that

$$\epsilon_i = -y_i w^T (\tilde{x}_i - x_i) \Rightarrow \epsilon_i^2 = w^T ||\tilde{x}_i - x_i||^2 w \Rightarrow \epsilon_i = ||w|| \cdot ||\tilde{x}_i - x_i||$$

So we conclude that $epsilon_i$ is proportional to the distance from x_i to the margin hyperplane, and the coefficient is $||w||$.

2 Problem 2

We select $\lambda = \frac{1}{C}$, and we can rewrite the constraint to be

$$\epsilon_i \geq \max\{0, 1 - y_i(w^T x_i + b)\}$$

Then, we can rewrite the optimization problem to be

$$\min_{w,b} \frac{\lambda}{2}||w||^2 + \frac{1}{n}\left[(1-\alpha)\sum_{i:y_i=1} \max\{0, 1 - y_i(w^T x_i + b)\} + \alpha \sum_{i:y_i=-1} \max\{0, 1 - y_i(w^T x_i + b)\}\right]$$

Now we can rewrite the last term to be $L(y, f(x))$, where $f(x) = w^T x + b$, and we can write

$$L(y, f(x)) = \begin{cases} (1-\alpha) \cdot \max\{0, 1 - y_i f(x_i)\} & \text{if } y_i = 1 \\ \alpha \cdot \max\{0, 1 - y_i f(x_i)\} & \text{if } y_i = -1 \end{cases}$$

Thus, the optimize problem can be written as a regularized ERM that is

$$\min_{w,b} \frac{\lambda}{2}||w||^2 + \frac{1}{n}\sum_{i=1}^{n} L(y_i, f(x_i))$$

Which yields the same classifier as the quadratic program.

3 Problem 3

(a): We assume that the occurrence of a word does not depend on previous occurrence of the same word.
(b): we notice that

$$\hat{y} = \arg\max_{k\in\{0,1\}} \log\left(\hat{\pi}_k \prod_{j=1}^{d}\left(\frac{n_{kj}+\alpha}{n_k+\alpha d}\right)^{x_j}\right) \tag{1}$$

$$= \arg\max_{k\in\{0,1\}} \log(\hat{\pi}_k) + \sum_{j=1}^{d} x_j \log\left(\frac{n_{kj}+\alpha}{n_k+\alpha d}\right) \tag{2}$$

$$= \arg\max_{k\in\{0,1\}} \log(\hat{\pi}_k) + \sum_{j=1}^{d} x_j \log(n_{kj}+\alpha) - \sum_{j=1}^{d} x_j \log(n_k+\alpha d) \tag{3}$$

$$\tag{4}$$

(c): Please see the code in my py file on Canvas. The estimated $\log(\pi_0) = -0.697$, the estimated $\log(\pi_1) = -0.689$.
(d): The test error would be 0.126.
(e): The test error would be about 0.4987.