Machine Learning, 2024 Spring
Assignment 7

Name: Zhou Shouchen

Student ID: 2021533042

Notice

Plagiarizer will get 0 points.
LaTeXis highly recommended. Otherwise you should write as legibly as possible.

Problem 1 Referring to Figure 4.10, why are both curves increasing with K? Why do they converge to each other with increasing K? (20pt)

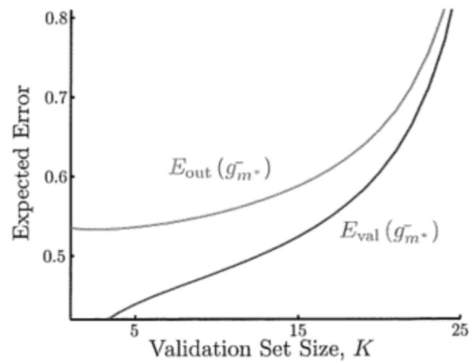

Figure 4.10: Optimistic bias of the validation error when using a validation set for the model selected.

Solution

Problem 2

1. From Figure 4.12, $\mathbb{E}[E_{out}(g^{-}_{m^*})]$ is initially decreasing. How can this be, if $\mathbb{E}[E_{out}(g^{-}_{m^*})]$ is increasing in K for each m? (10pt)
2. From Figure 4.12 we see that $\mathbb{E}[E_{out}(g^{-}_{m^*})]$ is initially decreasing, and then it starts to increase. What are the possible reasons for this? (10pt)
3. When $K = 1$, $\mathbb{E}[E_{out}(g^{-}_{m^*})] < \mathbb{E}[E_{out}(g_{m^*})]$. How can this be, if the learning curves for both models are decreasing? (10pt)

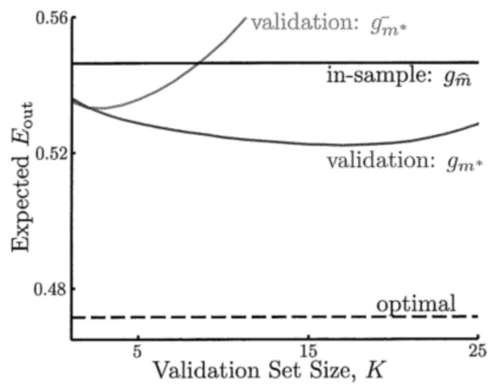

Figure 4.12: Model selection between $\mathcal{H}_2$ and $\mathcal{H}_5$ using a validation set. The solid black line uses E_{in} for model selection, which always selects $\mathcal{H}_5$. The dotted line shows the optimal model selection, if we could select the model based on the true out-of-sample error. This is unachievable, but a useful benchmark. The best performer is clearly the validation set, outputting g_{m^*}. For suitable K, even $g^{-}_{m^*}$ is better than in-sample selection.

Solution

Problem 3

Definition 1 (leave-one-out cross-validation) *Select each training example in turn as the single example to be held-out, train the classifier on the basis of all the remaining training examples, test the resulting classifier on the held-out example, and count the errors.*

Let the superscript -i denote the parameters we would obtain by finding the SVM classifier f without the ith training example. Define the *leave-one-out CV error* as

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(y_i, f(\mathbf{x}_i; \mathbf{w}^{-i}, b^{-i})), \tag{1}$$

where $\mathcal{L}$ is the zero-one loss. Prove that

$$\text{leave-one-out CV error} \leq \frac{\text{number of support vectors}}{\text{n}} \tag{2}$$

(20pt)
Solution

Problem 4

The l_1-norm SVM can be formulated as follows

$$\begin{aligned} &\min_{\mathbf{w},b} \|\mathbf{w}\|_1 \\ &\text{s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \leq 1, i = 1, ..., n. \end{aligned} \tag{3}$$

Please deprive the equivalent linear programming formulation of (3) (10pt), give its dual formulation (10pt). Also, please explain how to determine the support vector SV according to the optimal multiplier. (10pt)

Solution