

Compsci 571 HW5

Yilin Gao (yg95)

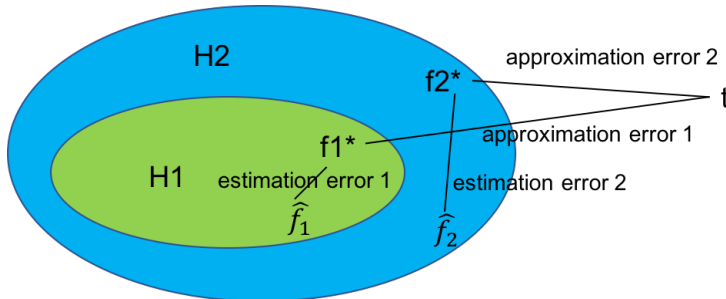
March 14, 2018

1 Hoeffding's Inequality

- (a)
- (b)

2 VC Dimension

- (a) The VC dimension of \mathcal{H}_1 is p , because its feature space is p -dimensional, and all separating hyperplanes are linear in the feature space and goes through the origin point.
The VC dimension of \mathcal{H}_2 is $\frac{(p+2)(p+1)}{2}$, because its feature space is $\frac{(p+2)(p+1)}{2}$ -dimensional, and all separating hyperplanes are linear in the feature space and goes through the origin point.
- (b) The approximation and estimation errors for \mathcal{H}_1 , \mathcal{H}_2 and \hat{f}_1, \hat{f}_2 are depicted as following:



In the picture, t stands for the global optimal function based on true risk, f_1^* and f_2^* stand for the optimal functions based on true risk in function classes \mathcal{H}_1 and \mathcal{H}_2 respectively, and \hat{f}_1 and \hat{f}_2 are optimal functions based on empirical risk in function classes \mathcal{H}_1 and \mathcal{H}_2 respectively.

As n increases, the estimation error increases because with more data we are able to estimate the model more accurately, and decrease the discrepancy between true risk and empirical risk, i.e., the estimation error. And the approximation error won't be affected significantly because the approximation error is dependent on the function class \mathcal{H} , but not the data.

- (c) One function class F such that its VC dimension is not equal to its number of parameters is the set of functions generated by the K-nearest neighbors algorithm with $K = 1$. It only has 1 parameter K , while its VC dimension is infinity.

3 Ridge Regression

- (a)
- (b)
- (c)