Problem Set Template

Claire Goeckner-Wald

November 14, 2016

Validation

1) [d] $k = 6$

For $k = 6$ and $N = 25$, the validation error is 0.0. See attached code.

2) [e] $k = 7$

For $k = 7$ and $N = 25$, the out-of-sample error is 0.072. See attached code.

3) [d] $k = 6$

For $k = 6$ and $N = 10$, the validation error is 0.08. See attached code.

4) [d] $k = 6$

For $k = 6$ and $N = 10$, the out-of-sample error is 0.192. See attached code.

5) [b] 0.1, 0.2

For problems one and three, the chosen model's out-of-sample errors for $k = 7$, where $N = 25$ and $N = 10$, respectively, are 0.084 and 0.192. This is closet in Euclidean distance to 0.1 and 0.2.

Validation Bias

6) [d] 0.5, 0.5, 0.4

The expected values of e_1 and e_2 are 0.5, because their expected value is the average of $[0, 1]$, which is $\frac{0+1}{2}$. The expected value of $\min e_1, e_2 = \frac{1}{3}$. See attached code for an experimental proof.

$$\frac{1}{1^2}\int_0^1 \int_0^1 \min(x, y)dxdy$$

$$\min(x, y) = \frac{x + y}{2} - |x - y|$$

$$\int_0^1 \int_0^1 \big(\frac{x + y}{2} - |x - y|\big)dxdy = \frac{1}{3}$$

Cross Validation

7) [c] or [d]

PLA versus SVM

8) [c] or [d]

9) [c] or [d]

10) [c] or [d]