Quiz 1, 10 minutes

Question 1: Compute the 2-nn estimator at $x_0 = 2$ based on the data (x_i, y_i) taking values $(0, 2), (2, 5), (3, 5), (5, 2), (6, 2)$.

Question 2: The table below shows 10-fold cv errors and $10^{-1/2}\widehat{se}$ for different k. Which value of k would be selected based on 10-fold cv (**not** using the 1-se rule)?

k	1	2	3	4	5
cv	5	2	3	2.5	5
$10^{-1/2}\widehat{se}$	1	1.5	1.5	2	3

Question 3: Assume you have a sample of size $n = 20$ generated from the model $y_i = f(x_i) + \varepsilon_i$ where ε_i are i.i.d. with variance σ^2 and all x_i are different. Which of the following will always get smaller when k is increased from $k = 5$ to $k = 10$?

1. The squared bias of k-nn regression at a fixed x_0.
2. The variance of k-nn regression at a fixed x_0.
3. The MSE of k-nn regression at a fixed x_0.

Question 4: Which of the following statements are *always* true for k-nn regression with a sample of size $n = 50$?

1. 7-nn regression gives a constant $\widehat{f}$ (i.e. $\widehat{f}(x_0)$ is independent of x_0).
2. 50-nn regression gives a test error of 0.
3. 50-nn regression leads to a training error of 0.