z5286124 Jinghan Wang

(a).

(i). Suppose $E[T_1]$ is the Expected Value of T_1.

$E[T_1] = E[a\bar{X} + (1-a)c_* S] = aE[\bar{X}] + (1-a)E[c_* S]$.

As given by (F-2), $E[\bar{X}] = \gamma$, $E[c_* S] = \gamma$

$E[T_1] = a\gamma + (1-a)\gamma$

$= a\gamma + \gamma - a\gamma$

$= \gamma$

Therefore, T_1 is an unbiased estimator γ for any choice of the constant a.

(ii).

$MSE(T) = Var(T) + (E[T] - \theta)^2$

According to (i). $E[T_1] - r = 0$. $MSE(T) = Var(T)$

As $T_1 = a\bar{x} + (1-a) C_* S$,

$var(T_1) = a^2 Var(\bar{x}) + (1-a)^2 Var(C_* S) + 2a(1-a) Cov(\bar{x}, C_* S)$

As Given by (F1), $\bar{x}$, S^2 are independent $Cov(\bar{x}, C_* S) = 0$

$$var(T_1) = a^2 Var(\bar{x}) + (1-a)^2 Var(C_* S)$$

According to the $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$,

$$Var(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2} n\gamma^2 = \frac{\gamma^2}{n}.$$

According to the fundamental property of variance,

$Var(cX) = c^2 Var(X)$ and topic. $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$$Var(C_* S) = C_*^2 Var(S) = C_*^2 \frac{2\gamma^4}{n-1}$$

Therefore,

$$Var(T_1) = a^2 \frac{\gamma^2}{n} + (1-a)^2 C_*^2 \frac{2\gamma^4}{n-1}$$

$$\frac{\partial Var(T_1)}{\partial a} = \frac{2\gamma^2}{n} a - 2C_*^2 \frac{2\gamma^4}{n-1}(1-a) = 0$$

$$a = \frac{2 C_*^2 \gamma^2 n}{2 C_*^2 \gamma^2 n + n - 1}$$

$$T_i^* = \frac{2 C_*^2 \gamma^2 n}{2 C_*^2 \gamma^2 n + n - 1} \bar{X} + \left(1 - \frac{2 C_*^2 \gamma^2 n}{2 C_*^2 \gamma^2 n + n - 1}\right) C_* S$$

(iii).

$$MSE(T_2) = Var(T_2) + Bias^2(T_2)$$

According to (ii) prove,

$$Var(\bar{x}) = \frac{\gamma^2}{n} \quad Var(c_* S) = c_*^2 \frac{2\gamma^4}{n-1}$$

$$Var(T_2) = a_1^2 \frac{\gamma^2}{n} + a_2^2 \, c_*^2 \frac{2\gamma^4}{n-1}$$

According to (ii) prove and topic

$$E(T_2) = a_1 E(\bar{X}) + a_2 E(c_* S) = a_1\gamma + a_2\gamma$$

For the T_2 to be unbiased, $E(T_2) = \gamma$. Therefore, $a_1 + a_2 = 1$

$$Bias(T_2) = E(T_2) - \gamma = (a_1 + a_2 - 1)\gamma$$

$$MSE(T_2) = a_1^2 \frac{\gamma^2}{n} + a_2^2 \, c_*^2 \frac{2\gamma^4}{n-1} + (a_1 + a_2 - 1)^2 \gamma^2$$

Get the derivatives of $MSE(T_2)$ with a_1 and a_2.

$$\frac{\partial MSE(T_2)}{\partial a_1} = \frac{2a_1\gamma^2}{n} + \gamma^2(2a_1 + 2a_2 - 2)$$

$$\frac{\partial MSE(T_2)}{\partial a_2} = 2a_2 \cdot c_*^2 \frac{2\gamma^4}{n-1} + \gamma^2(2a_1 + 2a_2 - 2)$$

According to above, $a_1 + a_2 = 1$

$$a_1^* = \frac{n \cdot C_*^2 \, \frac{2\gamma^4}{n-1}}{\gamma^2 + n \cdot C_*^2 \, \frac{2\gamma^4}{n-1} + C_*^2 \, \frac{2\gamma^4}{n-1}}$$

$$a_2^* = \frac{\gamma^2}{\gamma^2 + n \cdot C_*^2 \, \frac{2\gamma^4}{n-1} + C_*^2 \, \frac{2\gamma^4}{n-1}}$$

$$T_2^* = \frac{n \cdot C_*^2 \, \frac{2\gamma^4}{n-1}}{\gamma^2 + n \cdot C_*^2 \, \frac{2\gamma^4}{n-1} + C_*^2 \, \frac{2\gamma^4}{n-1}} \bar{X} + \frac{\gamma^2}{\gamma^2 + n \cdot C_*^2 \, \frac{2\gamma^4}{n-1} + C_*^2 \, \frac{2\gamma^4}{n-1}} (c_* S)$$

(iv). According to (ii) and (iii). Let $C_*^2 \frac{2\gamma^4}{n-1} = Var(c_* S)$ in counting.

$$MSE(T_1^*) = \frac{\gamma^2 n Var(c_* S)}{(\gamma^2 + n Var(c_* S))^2} + Var(c_* S)\left(1 - \frac{n Var^2(c_* S)}{\gamma^2 + n Var(c_* S)}\right)^2$$

$$MS\hat{E}(T_2^*) = \frac{\gamma^4 Var(c_* S)}{(\gamma^2 + n Var(c_* S) + Var(c_* S))^2} +$$

$$\frac{\gamma^2 n Var^2(c_* S)}{(\gamma^2 + n Var(c_* S) + Var(c_* S))^2} +$$

$$\gamma^2\left(\frac{\gamma^2 + n Var(c_* S)}{\gamma^2 + n Var(c_* S) + Var(c_* S)} - 1\right)^2$$

Simplify both of them, we can get

$$MSE(T_1^*) - MS\hat{E}(T_2^*)$$

$$= \frac{\gamma^2 Var(c_* S)}{\gamma^2 + n Var(c_* S)} - \frac{\gamma^2 Var(c_* S)}{\gamma^2 + n Var(c_* S) + Var(c_* S)}$$

The numerator is same but $\gamma^2 + n Var(c_* S) \le \gamma^2 + n Var(c_* S) + Var(c_* S)$,

Therefore, $MSE(T_1^*) - MS\hat{E}(T_2^*) \geq 0$

$$MS\hat{E}(T_2^*) \le MSE(T_1^*)$$

(v). According to the topic.

- $T_2^* \geq 0$: $V_+ = T_2^*$
- $T_2^* < 0$: $V_+ = 0$

For $MSE(T) = E[(T-\gamma)^2]$

$$MSE(T_2^*) = Var(T_2^*) + Bias^2(T_2^*)$$

if $T_2^* \geq 0$ $\quad MSE(V_+) = MSE(T_2^*)$

if $T_2^* < 0$ $\quad MSE(V_+) = E[(0-\gamma)^2] = \gamma^2$

$$T_2^* - \gamma < -\gamma$$

$$(T_2^* - \gamma)^2 < \gamma^2$$

$$MSE(T_2^*) > MSE(V_+)$$

Therefore,

$$MSE(V_+) \leq MSE(T_2^*)$$

(b)

(i).

According to the topic, $y = f(x) + \epsilon$. Suppose $t_i = f(z_i) + \epsilon_i$ is the response value for z_i. we can get

$$\hat{m}(x_0) = \frac{1}{k} \sum_{i \in N_k(x_0)} y_i = \frac{1}{k} \sum_{i=1}^{k} t_i$$

$$E[\hat{m}(x_0)] = \frac{1}{k} \sum_{i=1}^{k} E[t_i]$$

According to the topic, $\epsilon_i = 0$. Therefore we can simplify $E[t_i]$

$$E[t_i] = E[f(z_i) + \epsilon_i] = f(z_i) + E[\epsilon_i] = f(z_i)$$

$$E[\hat{m}(x_0)] = \frac{1}{k} \sum_{i=1}^{k} f(z_i)$$

According to the bias, we can get

$$\text{bias}(\hat{m}(x_0)) = \frac{1}{k} \sum_{i=1}^{k} f(z_i) - f(x_0)$$

$$[\text{bias}(\hat{m}(x_0))]^2 = \left(f(x_0) - \frac{1}{k} \sum_{i=1}^{k} f(z_i) \right)^2$$

(ii).

According to i) prove. we can get. $E(\hat{m}(x_0)) = \frac{1}{k}\sum_{i=1}^{k} f(z_i)$

According to the (i) suppose, $\hat{m}(x_0) = \frac{1}{k}\sum_{i=1}^{k} t_i$, $t_i = f(z_i) + \epsilon_i$

$$\hat{m}(x_0)^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} t_i t_j$$

$$E(\hat{m}(x_0)^2) = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} E(t_i t_j)$$

- For $i \neq j$,

$$E[t_i t_j] = E[(f(z_i)+\epsilon_i)(f(z_j)+\epsilon_j)] = f(z_i)f(z_j) + E[\epsilon_i \epsilon_j]$$

Since ϵ_i, ϵ_j is independent according to the topic $E[\epsilon_i, \epsilon_j] = 0$

$$E[t_i t_j] = f(z_i)f(z_j)$$

- For $i = j$.

$$E[t_i t_j] = E[f(z_i) + \epsilon_i]^2 = f(z_i)^2 + 2f(z_i)E[\epsilon_i] + E[\epsilon_i]^2.$$

According to the topic, $E[\epsilon_i] = 0$ and $E[\epsilon_i^2] = \sigma^2$

$$E[t_i t_j] = 2f(z_i)^2 + \sigma^2$$

$$E[\hat{m}(x_0)^2] = \frac{1}{k^2}\left(k\sum_{i=1}^{k}(f(z_i))^2 + \sigma^2) + k(k-1)\sum_{i\neq j} f(z_i)f(z_j)\right)$$

$$Var(\hat{m}(x_0)) = E[\hat{m}(x_0)^2] - E[\hat{m}(x_0)]^2$$

$$= \frac{1}{k^2}\left(k\sum_{i=1}^{k}(f(z_i))^2 + \sigma^2) + k(k-1)\sum_{i\neq j} f(z_i)f(z_j)\right) - \frac{1}{k}\sum_{i=1}^{k} f(z_i)$$

(iii) .

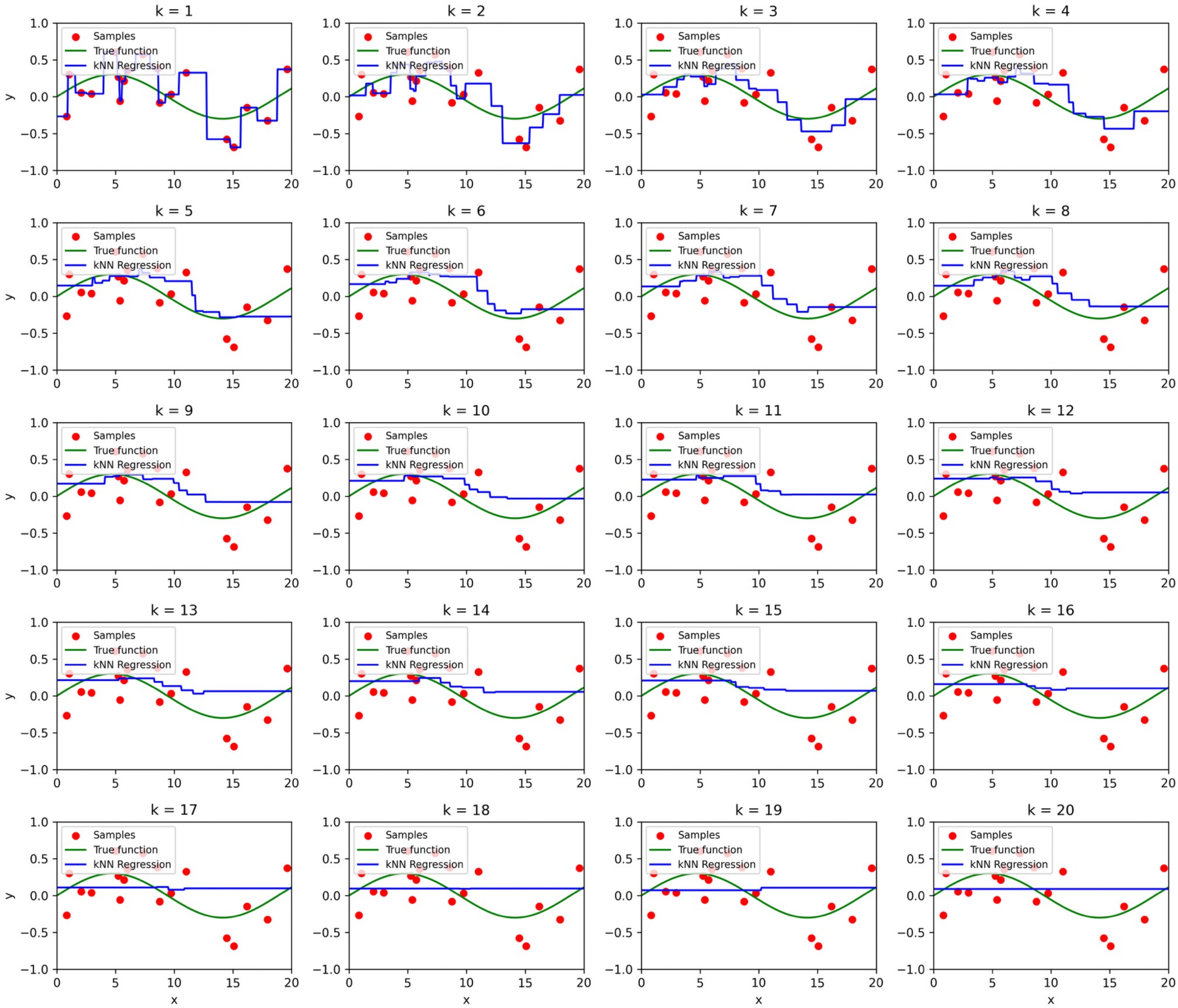
$$MSE(\hat{m}(x_0))$$

$$= \left(f(x_0) - \frac{1}{k}\sum_{i=1}^{k} f(z_i)\right)^2 + \frac{1}{k^2}\left(k\sum_{i=1}^{k}\left(f(z_i))^2 + \sigma^2\right) + k(k-1)\sum_{i \neq j} f(z_i)f(z_j)\right)$$

$$- \frac{1}{k}\sum_{i=1}^{k} f(z_i)$$

kNN regression performance is influenced by the choice of k and the nature of the data. It's essential to consider the trade-off between bias and variance when using kNN.

(iv).


k = 1
k = 2
k = 3
k = 4
k = 5
k = 6
k = 7
k = 8
k = 9
k = 10
k = 11
k = 12
k = 13
k = 14
k = 15
k = 16
k = 17
k = 18
k = 19
k = 20
Samples
True function
kNN Regression
x
y

```
import numpy as np
import matplotlib.pyplot as plt


f = lambda x: 0.3 * np.sin(x / 3)
sigma = 0.3
n = 20


x_samples = np.sort(np.random.uniform( low: 0,  high: 20, n))
y_samples = f(x_samples) + np.random.normal( loc: 0, sigma, n)


def kNN_regression(x, x_samples, y_samples, k):
    distances = np.abs(x_samples - x)
    k_nearest_indices = np.argsort(distances)[:k]

    return np.mean(y_samples[k_nearest_indices])


def q1_b_iv():
    x_values = np.linspace( start: 0,  stop: 20,  num: 400)
    k_values = list(range(1, 21))

    fig, axes = plt.subplots( nrows: 5,  ncols: 4, figsize=(14, 12))
    for ax, k in zip(axes.ravel(), k_values):
        y_pred = [kNN_regression(x, x_samples, y_samples, k) for x in x_values]

        ax.scatter(x_samples, y_samples, color='red', label='Samples', s=30)
        ax.plot(x_values, f(x_values), 'g-', label='True function')
        ax.plot(x_values, y_pred, 'b-', label='kNN Regression')
        ax.set_title(f"k = {k}")
        ax.set_xlim(0, 20)
        ax.set_ylim(-1, 1)
        if k in [1, 5, 9, 13, 17]:
            ax.set_ylabel('y')
        if k > 16:
            ax.set_xlabel('x')
        ax.legend(loc='upper left', fontsize='small')

    plt.tight_layout()
    plt.savefig( *args: 'q1_b_iv.png', dpi=300)
    plt.show()


if __name__ == '__main__':
    q1_b_iv()
```

(v).

When k is very small (1NN), the bias can be high because the prediction is based entirely on the most recent data point. The variance will be high, and the prediction is very sensitive to noise in a single data point.

When k is very large (k -> n), the bias increases because the prediction becomes an average of all the data and may deviate from the true value. The variance decreases because the prediction is no longer affected by noise at individual points.

Choosing a small k may result in high bias and high variance. while a large k may result in low variance and increased bias. A balance needs to be found between bias and variance.