Question 1. (20 points)

1) Calculate by hand the empirical variance for the following datasets:

(a) [4, 7, 9, 11, 15]

(b) [18, 22, 25, 30, 35, 40]

2) Once you have manually calculated the variance, confirm your result with Numpy.

- Note that the documentation for numpy variance is at: https://numpy.org/doc/stable/reference/generated/numpy.var.html
- Read carefully the ddof option it determines if you are using empirical or population variance.

3) Calculate the population variance for the following datasets:

(a) [12, 14, 16, 18, 20]

(b) [28, 32, 36, 40, 44, 48]

4) Once you have manually calculated the population variance, confirm your result with Numpy.
5) Comparing Variances

(a) Explain the key differences between empirical and population variance calculations.

(b) Provide an example where you should use population variance over empirical variance, and vice versa.

Answer

1) Empirical Variance Calculations

(a) Dataset: [4, 7, 9, 11, 15]

$$\bar{x} = \frac{4 + 7 + 9 + 11 + 15}{5} = 9.2$$

$$\begin{aligned}
(4 - 9.2)^2 &= 27.04 \\
(7 - 9.2)^2 &= 4.84 \\
(9 - 9.2)^2 &= 0.04 \\
(11 - 9.2)^2 &= 3.24 \\
(15 - 9.2)^2 &= 33.64
\end{aligned}$$

$$27.04 + 4.84 + 0.04 + 3.24 + 33.64 = 68.8$$

$$s^2 = \frac{68.8}{4} = 17.2$$

Empirical variance for dataset (a) = 17.2

(b) Dataset: [18, 22, 25, 30, 35, 40]

$$\bar{x} = \frac{18 + 22 + 25 + 30 + 35 + 40}{6} = 28.3333$$

$$(18 - 28.3333)^2 = 106.7778$$
$$(22 - 28.3333)^2 = 40.1111$$
$$(25 - 28.3333)^2 = 11.1111$$
$$(30 - 28.3333)^2 = 2.7778$$
$$(35 - 28.3333)^2 = 44.4444$$
$$(40 - 28.3333)^2 = 136.1111$$

$$106.7778 + 40.1111 + 11.1111 + 2.7778 + 44.4444 + 136.1111 = 341.3333$$

$$s^2 = \frac{341.3333}{5} = 68.2667$$

Empirical variance for dataset (b) = 68.2667

3) Population Variance Calculations

(a) Dataset: [12, 14, 16, 18, 20]

$$\mu = \frac{12 + 14 + 16 + 18 + 20}{5} = 16$$

$$(12 - 16)^2 = 16$$
$$(14 - 16)^2 = 4$$
$$(16 - 16)^2 = 0$$
$$(18 - 16)^2 = 4$$
$$(20 - 16)^2 = 16$$

$$16 + 4 + 0 + 4 + 16 = 40$$

$$\sigma^2 = \frac{40}{5} = 8$$

Population variance for dataset (a) = 8

(b) Dataset: [28, 32, 36, 40, 44, 48]

$$\mu = \frac{28 + 32 + 36 + 40 + 44 + 48}{6} = 38$$

$$\begin{aligned}
(28 - 38)^2 &= 100 \\
(32 - 38)^2 &= 36 \\
(36 - 38)^2 &= 4 \\
(40 - 38)^2 &= 4 \\
(44 - 38)^2 &= 36 \\
(48 - 38)^2 &= 100
\end{aligned}$$

$$100 + 36 + 4 + 4 + 36 + 100 = 280$$

$$\sigma^2 = \frac{280}{6} = 46.6667$$

Population variance for dataset (b) = 46.6667

5) Comparing Variances

(a) Key differences between empirical and population variance calculations:

1. **Denominator**: Empirical variance uses (n-1), population variance uses n.
2. **Bias**: Empirical variance is an unbiased estimator, population variance assumes entire population.
3. **Sample size**: Empirical for samples, population for entire population.
4. **Notation**: Empirical s^2, population σ^2.
5. **Estimation**: Empirical estimates true variance, population is actual variance.

(b) Examples of when to use each type of variance:

Population Variance: Use for entire, finite population. Example: variance of test scores for all students in a specific class.

Empirical Variance: Use for sample of larger population. Example: estimating variance of heights in a country based on 1000 people sample.

Question 2. (20 points)

Given

$$\min_w \sum_i (w^\top \phi(x_i) - y_i)^2 + \lambda \|w\|_2^2.$$

1)

$$\frac{\partial J}{\partial w} = 2X^T(Xw - Y) + 2\lambda w$$

2) Given:

$$\lambda = 2, \quad \eta = 0.5$$

$$X = \begin{bmatrix} 1 & -2 & 0 & 1 \\ -2 & -1 & 1 & 2 \\ 1 & 2 & -1 & 1 \end{bmatrix}, \quad w_0 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad Y = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$$

First Iteration:

$$Xw_0 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

$$Xw_0 - Y = \begin{bmatrix} -2 \\ -2 \\ 0 \end{bmatrix}$$

$$X^T(Xw_0 - Y) = \begin{bmatrix} 2 \\ 6 \\ -2 \\ -2 \end{bmatrix}$$

$$\nabla J(w_0) = \begin{bmatrix} 0 \\ 12 \\ 0 \\ 0 \end{bmatrix}$$

$$w_1 = \begin{bmatrix} -1 \\ -6 \\ 1 \\ 1 \end{bmatrix}$$

Second Iteration:

$$Xw_1 = \begin{bmatrix} 12 \\ -7 \\ -13 \end{bmatrix}$$

$$Xw_1 - Y = \begin{bmatrix} 10 \\ -10 \\ -12 \end{bmatrix}$$

$$X^T(Xw_1 - Y) = \begin{bmatrix} 32 \\ -6 \\ 2 \\ -12 \end{bmatrix}$$

$$\nabla J(w_1) = \begin{bmatrix} 60 \\ -36 \\ 8 \\ -20 \end{bmatrix}$$

$$w_2 = \begin{bmatrix} -31 \\ 12 \\ -3 \\ 11 \end{bmatrix}$$