

L11 PROBLEM 2 (7/7 points)

Assume you have the following set: $[-6, -6, -4, -4, 2, 2, 2]$. Consider the following two clusters:

$C1 = [2, 2, -6, -6]$ and $C2 = [-4, -4, 2]$.

1. What is the variance of C1? Use the variance as defined in the video, where you do not divide by the size.

Answer: 64

When calculating this "variation on variance" when comparing clusters, the instructor decided it would be better to not divide by the length of the cluster. If you divide by the length of the cluster, you will be giving the same importance to a large cluster as a small one. However, a large cluster with a small variance should be more important because that would be harder or less likely to achieve.

2. What is the variance of C2? Use the variance as defined in the video, where you do not divide by the size.

Answer: 24

3. What is the badness of the set of clusters C1 and C2?

Answer: 88

EXPLANATION

The mean of C1 is $(2+2-6-6)/4 = -2$. The variance is $(-2-2)^2 + (-2-2)^2 + (-2+6)^2 + (-2+6)^2 = 4 \cdot 16 = 64$.

The mean of C2 is $(2-4-4)/3 = -2$. The variance is $(-2-2)^2 + (-2+4)^2 + (-2+4)^2 = 16 + 4 + 4 = 24$.

The badness is the sum of $\text{var}(C1)$ and $\text{var}(C2)$.

4. Now consider the following two clusters: $C1 = [2, 2, 2]$ and $C2 = [-6, -6, -4, -4]$.

What is the variance of C1?

0.0

Answer: 0

5. What is the variance of C2?

4.0

4.0

Answer: 4

6. What is the badness of the set of clusters C1 and C2?

4.0

4.0

Answer: 4

7. Are there other cluster options for this data that would lead to a smaller badness value? Do not consider the case of each data point being a separate cluster.

☒ Yes



☐ No

EXPLANATION

Having clusters $C1 = [2, 2, 2]$, $C2 = [-6, -6]$, and $C3 = [-4, -4]$ will lead to a badness value of 0.

Check

Hide Answer

Show Discussion



EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities,

About edX

About

News

Contact


FAQ

edX Blog

Follow Us

 Twitter

 Facebook

 Meetup

 LinkedIn

law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

© 2014 edX, some rights reserved.

[Terms of Service and Honor Code](#)

[Privacy Policy \(Revised 4/16/2014\)](#)

[Donate to edX](#)

[Jobs at edX](#)



[Google+](#)