Quiz 9 Solution

Q1

Offline Algorithm (0.5pt): Given the **entire input**, then compute some of function of it.
Online Algorithm (0.5pt): Given the **input one piece at a time(streaming)**, and need to make **irrevocable decision** along the way.
Offline Algorithm Examples (0.5pt): K-Means, BFR, CURE…
Online Algorithm Examples (0.5pt): Blooming filter, DGIM…

Q2

a) (1pt) The **Mahalanobis distance** is less than a threshold, then the point is close enough to the cluster.

$$\sqrt{\sum_{i=1}^{d}\left(\frac{p_i - c_i}{\sigma_i}\right)^2}$$

b) (1pt) In addition to **Mahalanobis distance**. Compute the **variance** of the **combined mini-cluster**. Combine if the combined variance is below some threshold.
Or: Treat dimensions differently, consider density.

f) (2pt)

1. Move the representative points say 20% distance toward the centroid.
2. Replace each point p in the “closest cluster” (Find the closest representative to p and assign it to representative’s cluster.

Q3 B/C(1pt) D (2pt)