P2

Wednesday, March 6, 2024 3:24 PM

P2 :

Proof: Number that minimizes mean squared error is the mean of the dataset.

$$f(c) = \frac{1}{N}\left[(x_1 - c)^2 + (x_2 - c)^2 + \ldots + (x_N - c)^2\right]$$

$$f'(c) = \left[\frac{d}{dc}\left[(x_1 - c)^2\right] + \frac{d}{dc}\left[(x_2 - c)^2\right] \ldots\right)\frac{1}{N}$$

$$\downarrow$$

$$\frac{d}{dc}\left[x_1^2 - 2x_1c + c^2\right]$$

$$= \frac{1}{N}\left(\left[2c - 2x_1\right] + \left[2c - 2x_2\right] + \ldots \left[2c - 2x_n\right]\right)$$

$$f'(c) = \frac{1}{N}\left(2Nc - 2\left[x_1 + x_2 + \ldots + x_N\right]\right)$$

$$f'(c) = 2c - \frac{2}{N}\left[x_1 + x_2 + \ldots + x_N\right]$$

$$f'(c) = 2\left[c - \frac{1}{N}[x_1 + x_2 + \ldots + x_N]\right]$$

$$f'(c) = 0 = 2\left[c - \frac{1}{N}[x_1 + x_2 + \ldots + x_N]\right]$$

$$0 = c - \frac{1}{N}[x_1 + x_2 + \ldots + x_N]$$

$$\boxed{c = \frac{1}{N}[x_1 + x_2 + \ldots + x_N]}$$

$$c = \frac{1}{4}\left(175 + 172 + 180 + 185\right)$$

$$\boxed{c = 178}$$