For the Gaussian distribution μ & σ are represented $\theta = \{\mu, \sigma\}$.

Now, if $x_1, x_2, \cdots x_n$ are different points we need to find parameter (estimate) for which the above points are most probable to be found under the Gaussian curve.

If we treat individual data point as single event and its continuous data the PDF (density function) is joint probability as below—

$$f(x_1, x_2 \cdots x_n \mid \theta) = f(x_1 \mid \theta) \cdot f(x_2 \mid \theta) \cdots f(x_n \mid \theta)$$

$$= \prod_{i}^{n} f(x_i \mid \theta)$$

Goal is to maximize the above for a θ

or, $$\hat{\theta}_{MLE} = \text{argmax}_\theta \prod_{i=1}^{n} f(x_i \mid \theta)$$

Since, we are looking for max, we can take derivative and equate to 0.

$$\text{argmax}_\theta \prod_{i=1}^{n} f(x_i \mid \theta) \rightarrow \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i \mid \theta) = 0$$

Now, applying monotonic logarithmic we can rewrite —

$$\frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i \mid \theta) \sim \frac{\partial}{\partial \theta} \ln\left(\prod_{i=1}^{n} f(x_i \mid \theta)\right)$$

$$= \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \ln\langle f(x_i \mid \theta)\rangle$$

$$= \sum_{i=1}^{n} \ln\langle f(x_i \mid \theta')\rangle = 0$$

Using gradient notation -

0 $\rightarrow \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln \langle f(x_i|\theta) \rangle = \sum_{i=1}^{n} \nabla_{\mu,\sigma} \ln \langle f(x_i|\mu,\sigma) \rangle = 0$

1 Now taking a gradient with respect to μ

$$\sum_{i=1}^{n} \nabla_{\mu} \ln \left\langle \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \right\rangle = 0$$

or
$$\sum_{i=1}^{n} \nabla_{\mu} \ln \left\langle \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \right\rangle$$

$$= \sum_{i=1}^{n} \nabla_{\mu} - \frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$= -\frac{1}{2\sigma^2} \sum \nabla_{\mu} (x_i-\mu)^2$$

$$= -\frac{1}{2\sigma^2} \sum_{i=1}^{n} -2(x_i-\mu) = \boxed{\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i-\mu)}$$

Setting this last term equal to zero, we get the solution for μ as follows:

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i-\mu) = 0 \rightarrow \boxed{\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

2 Similarly if we take gradient of function σ we get -

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i-\mu)^2 = 0$$

or $\boxed{\hat{\sigma}_{MLE} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (x_i-\mu)^2}}$