MLE : Before any actual Bayesian:

Suppose $\vec{x} \sim X \sim \mathcal{N}(\vec{\mu}, \Sigma)$.

$$\mathbb{P}(X \mid \theta) = \sqrt{(2\pi)^D |\Sigma|} \cdot e^{-\frac{1}{2}(\vec{x}-\vec{\mu}) \Sigma^{-1} (\vec{x}-\vec{\mu})^T}$$

How to estimate θ, assuming $X = (a_i)_{i=1..N}$, the data, follows $\mathcal{N}(\mu, \Sigma)$?

We assume θ is some value, and want the actual data X to be "realistic", or to have a large likelihood, i.e find the θ / proba of observing X is maximal:

$$\mathbb{P}(X = (\vec{x}_i) \mid \theta) = \underbrace{\mathcal{L}(\theta)}_{\text{likelihood of the data}}$$

Let's assume the $\vec{x}_i$ are i.i.d:

$$\mathcal{L}(\theta) = \mathbb{P}(X = (\vec{x}_i) \mid \theta) = \prod_{i=1}^{N} \mathbb{P}(X_i = \vec{x}_i \mid \theta)$$

$$\theta^* = \operatorname{argmax}_\theta (\mathcal{L}(\theta)) = \operatorname{argmax}_\theta \underbrace{\log}_{\text{monotonous}, \nearrow}(\mathcal{L}(\theta))$$

$$= \operatorname{argmax}_\theta \left[-N \cdot \log(\sqrt{2\pi}\,\sigma) - \frac{1}{2} \sum_i^N (x_i - \mu)^2 / \sigma^2 \right] \quad \left(\begin{matrix}1D \\ \text{case}\end{matrix}\right)$$

$$0 = \frac{\delta \ell}{\delta \mu} = 0 - \frac{1}{\cancel{2}} \sum_i \cancel{2} \frac{(x - \mu)}{\sigma^2} \iff 0 = \sum_i^N (x_i - \mu)$$

$$\Rightarrow \boxed{\hat{\mu} = \frac{1}{N} \sum (x_i)}$$

$$0 = \frac{\delta \ell}{\delta \sigma} \iff 0 = -N \frac{1}{\sigma} - \frac{1}{2} \sum \frac{(x_i - \mu)^2}{\sigma^2} \cdot (-2) \qquad \downarrow \times \sigma^3 \qquad (\sigma > 0)$$

$$\iff N\sigma^2 = \sum_i (x_i - \mu)^2$$

$$\iff \boxed{\hat{\sigma}^2 = \frac{1}{N} (x_i - \hat{\mu})^2}$$

We find the intuitive, frequentist-style estimates, $\hat{\mu}$ and $\hat{\sigma}$, but this method is general