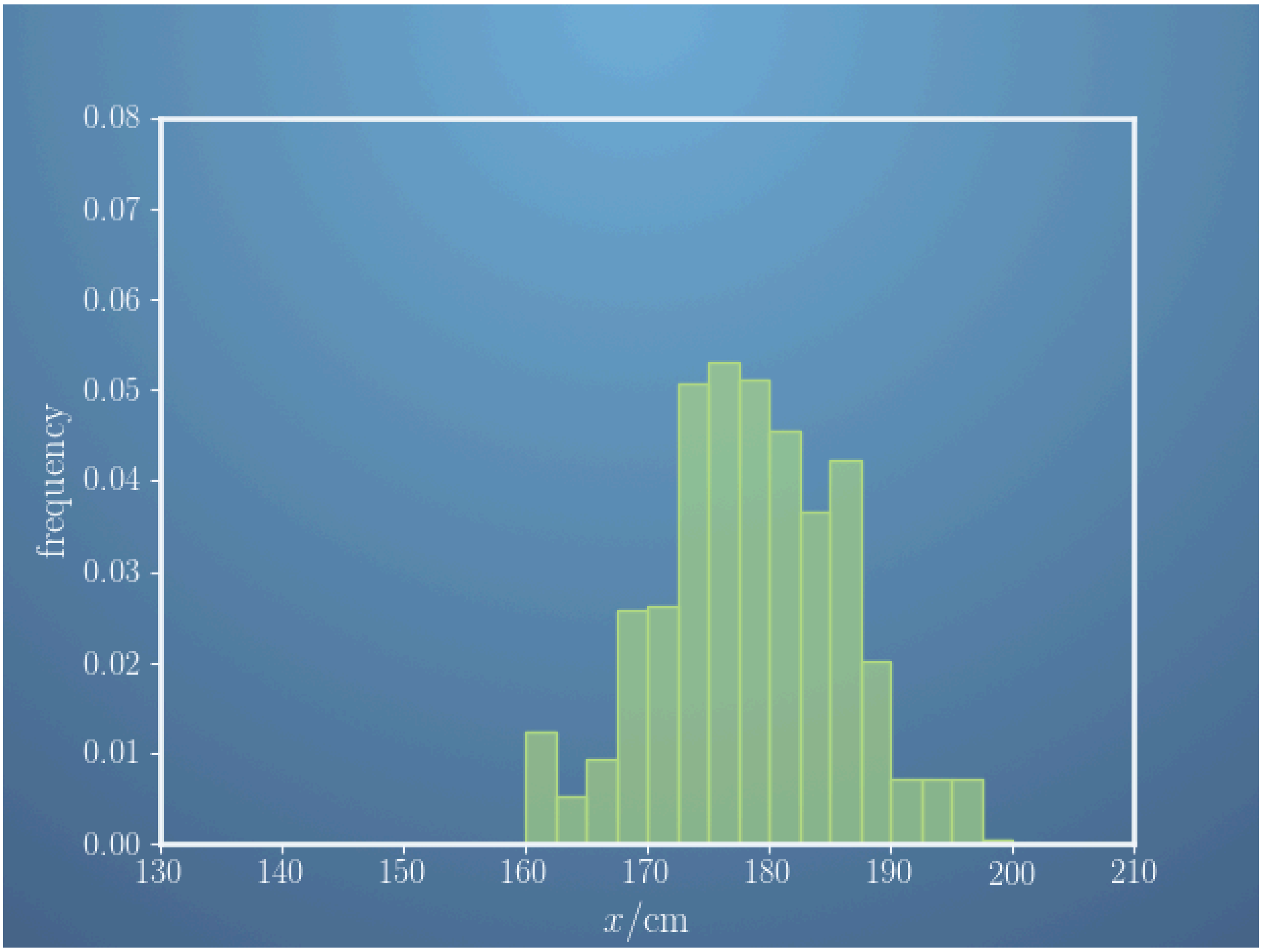
Quiz 1: Vectors in Machine Learning (MML Course)

Goal: This quiz reinforces the core idea that Vectors are a natural language for describing both **data** and **model parameters** in machine learning.

---

1. Data as a Vector



- **Concept:** A real-world data distribution, like a histogram of heights, can be represented as a vector.
- **Method:**
  1. Divide the data range (e.g., heights from 150cm to 210cm) into small, equal intervals or **"bins"** (e.g., 2.5cm wide).
  2. Count the **frequency** (how many data points) fall into each bin.
  3. Store these frequency counts in an ordered list. This list is a **vector**.
- **Example:** The frequency vector `f` could be:

$$\mathbf{f} = \begin{bmatrix} f_{150.0-152.5} \\ f_{152.5-155.0} \\ \vdots \end{bmatrix}$$

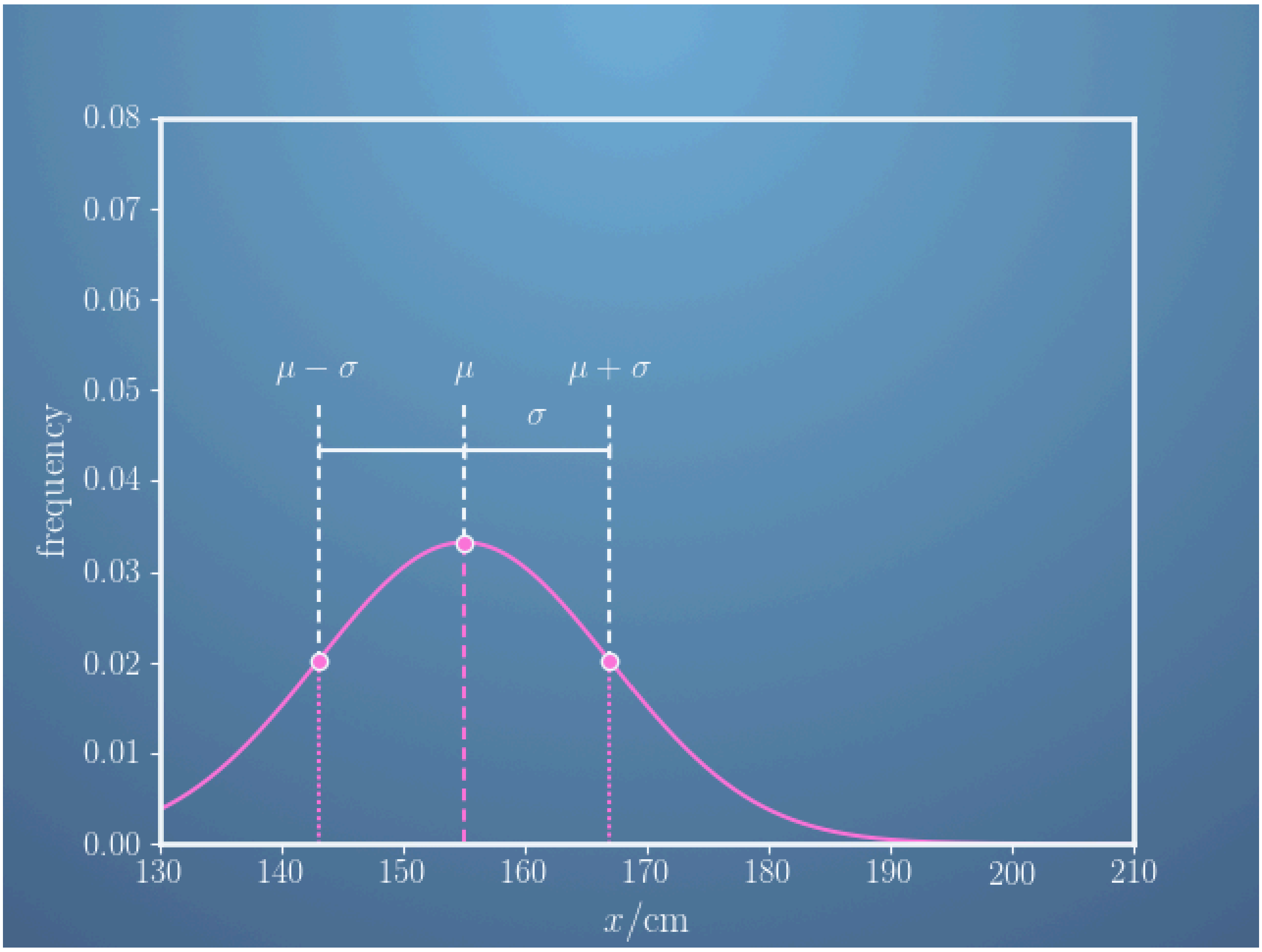
Each component of the vector corresponds to the height of a bar in the histogram.

- **Key Takeaways from Q1:**
  - A sufficiently large sample should be broadly representative of the true population.
  - The vector representation contains a specific number of elements (bins), many of which can have non-zero values.

---

2. Model Parameters as a Vector

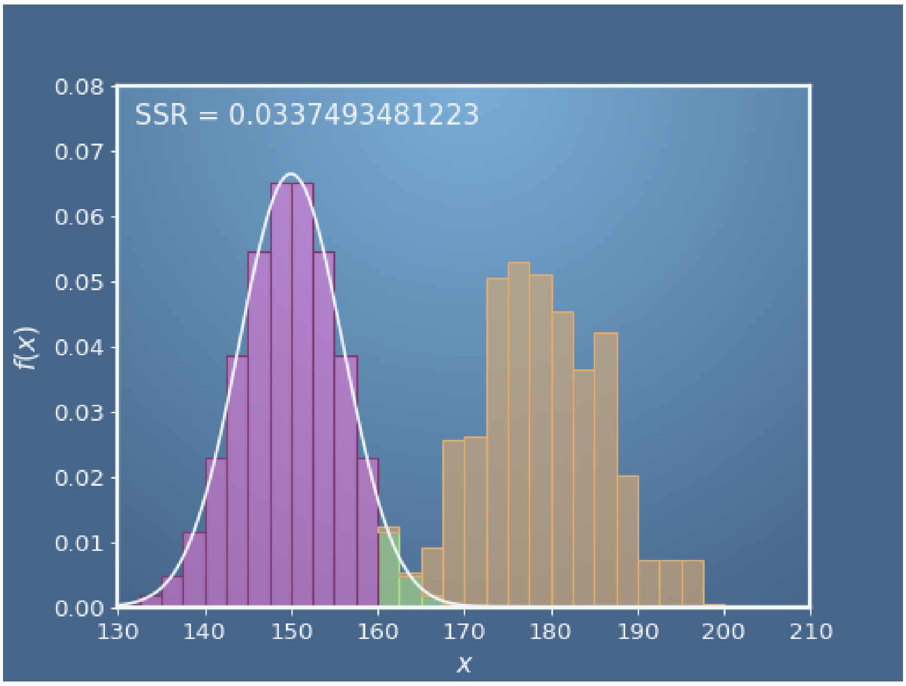
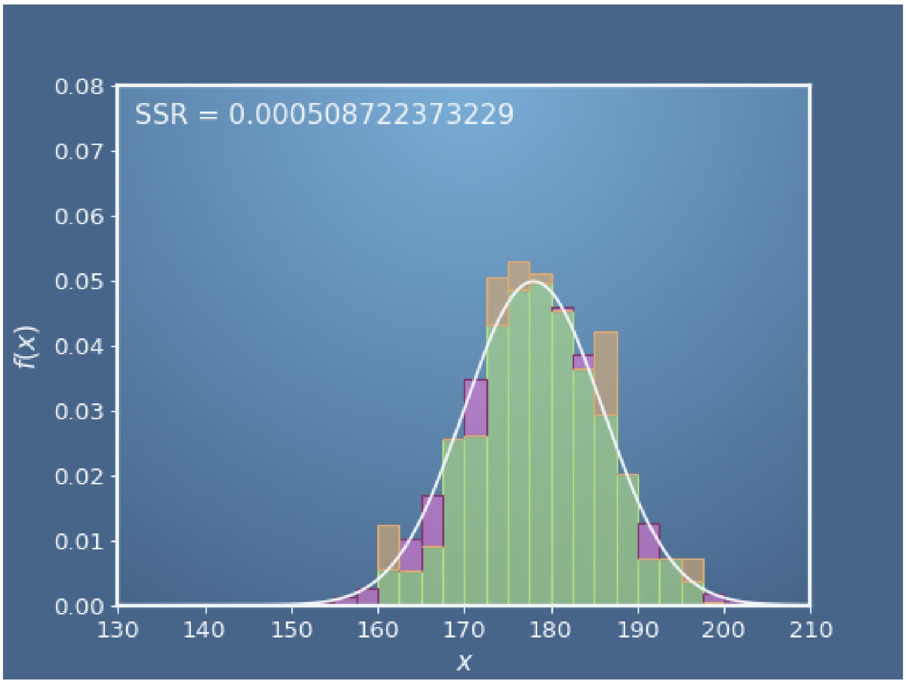
- **Concept:** The "knobs" or settings that control a mathematical model can also be grouped into a vector.



- **Example (Normal Distribution):**
  - The model `g(x)` is controlled by the mean `μ` and the standard deviation `σ`.
  - We can define a **parameter vector** `p`:

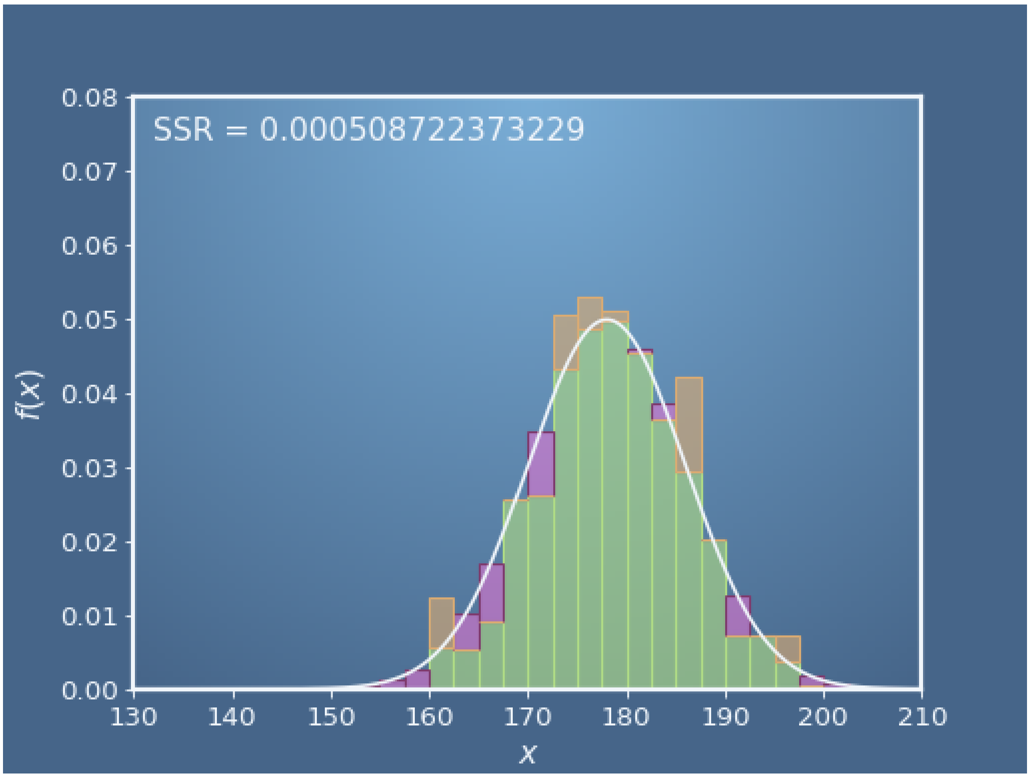
$$\mathbf{p} = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$$

- **Key Takeaways from Q2 & Q3:**
  - Different parameter vectors `p` produce different curves (different models).
  - We can visually inspect a data distribution to estimate the best-fitting parameter vector. For the given data, the center (`μ`) is around 178cm and the width (`σ`) is around 8cm.

> # Play with values of μ and σ to find the best fit.
> μ = 150 ; σ = 6
> p = [μ, σ]
>
> 
>
>
> # Play with values of μ and σ to find the best fit.
> μ = 178 ; σ = 8
> p = [μ, σ]
>
> 
>

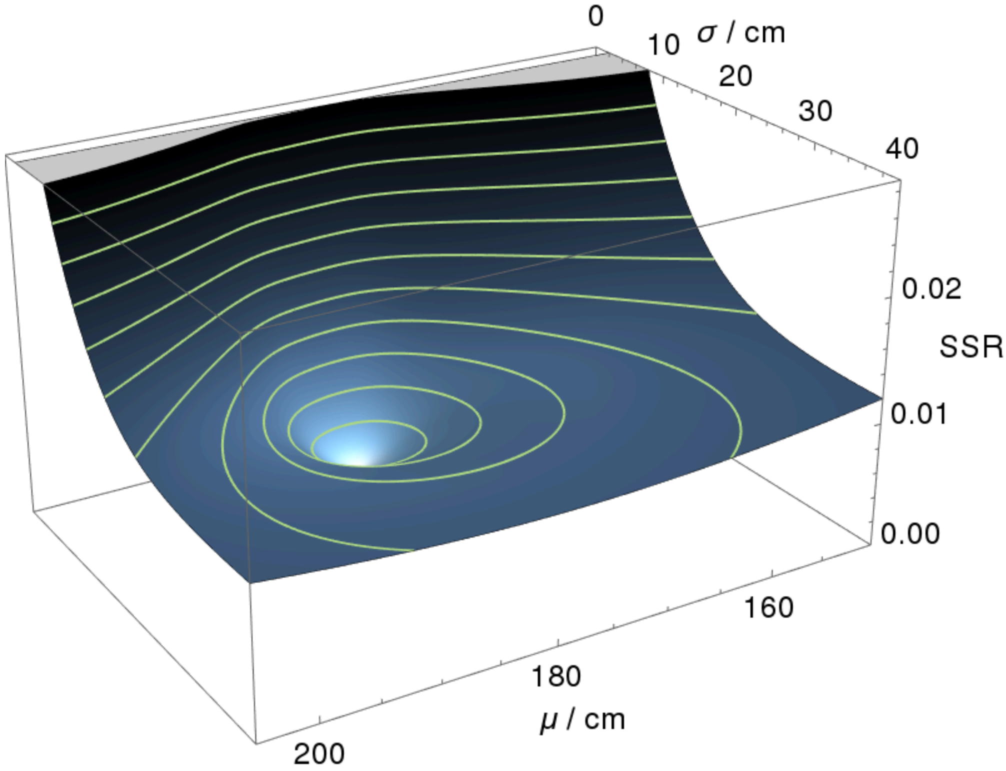
3. The Goal of "Fitting": Minimizing Residuals

- **Concept:** A good model is one that fits the measured data well. We need to quantify this.
- **Residuals:** The difference between the measured data (`f`) and the model's prediction (`g_p`) for each bin.
  - `Residual_vector = f - g_p`
- **Improving the Fit (Q4):**
  - If the model's curve is shifted to the left of the data's peak, we need to **increase the mean** `μ` .
  - If the model's curve is too narrow or wide compared to the data, we need to adjust the **standard deviation** `σ` .
- **Sum of Squared Residuals (SSR):** A single number to score the "badness" of a fit.
  - **Formula:** `SSR(p) = |f - g_p|²`
    (Calculate all residuals, square them, and add them all together).
  - **Goal:** Find the parameter vector `p` that **minimizes the SSR**.
- **Key Takeaway from Q5:** Through experimentation (or calculation), we can find a specific `p` (like `[178, 8]`) that results in a very low SSR, indicating a good fit.

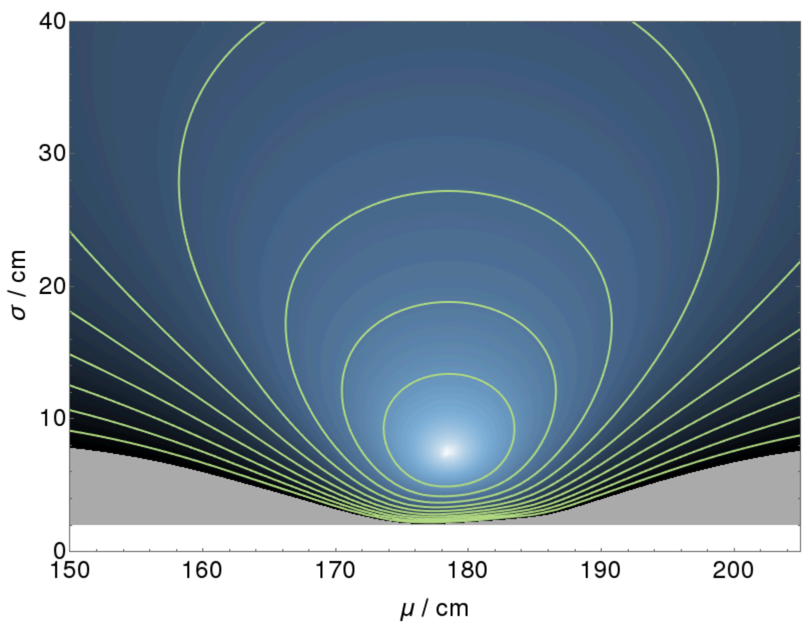


4. The Parameter Landscape and Optimization

- **Concept:** The SSR can be visualized as a **3D surface** or **landscape** over the **parameter space** (`μ` and `σ`).

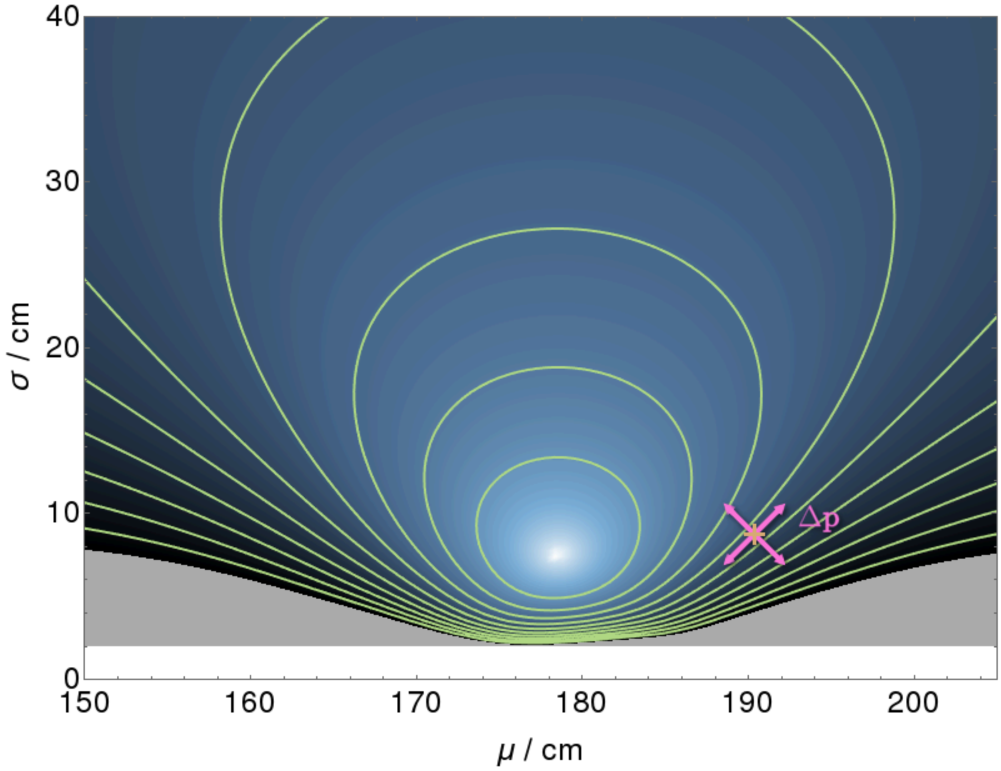


- **Global Minimum:** The lowest point in this landscape corresponds to the parameter vector `p` with the minimum SSR—the **best possible fit**.
- **Contour Map:** A top-down view of this landscape, where each line represents a constant SSR value.



- **Optimization as Navigation:**
  - Our goal is to find the lowest point in this landscape.
  - Moving **along a contour line** does not change the SSR (the fit quality remains the same).
  - Moving **perpendicular (at right angles) to contour lines** produces the greatest change in SSR. This is the most efficient way to go "downhill".
- **The Role of a New Vector, `Δp` (Q7):**
  - `Δp` represents a **"step"** or a change in our parameters: `[Δμ, Δσ]`.

- A "good" `Δp` is one that moves us from our current position `p` to a new position `p' = p + Δp` that has a lower SSR.
- The **best** `Δp` is the vector that points in the **steepest downhill direction** on the contour map (perpendicular to the contour lines, pointing inward).



Tags: #mml-specialization #linear-algebra #quiz-review #vectors #parameter-space #optimization #ssr