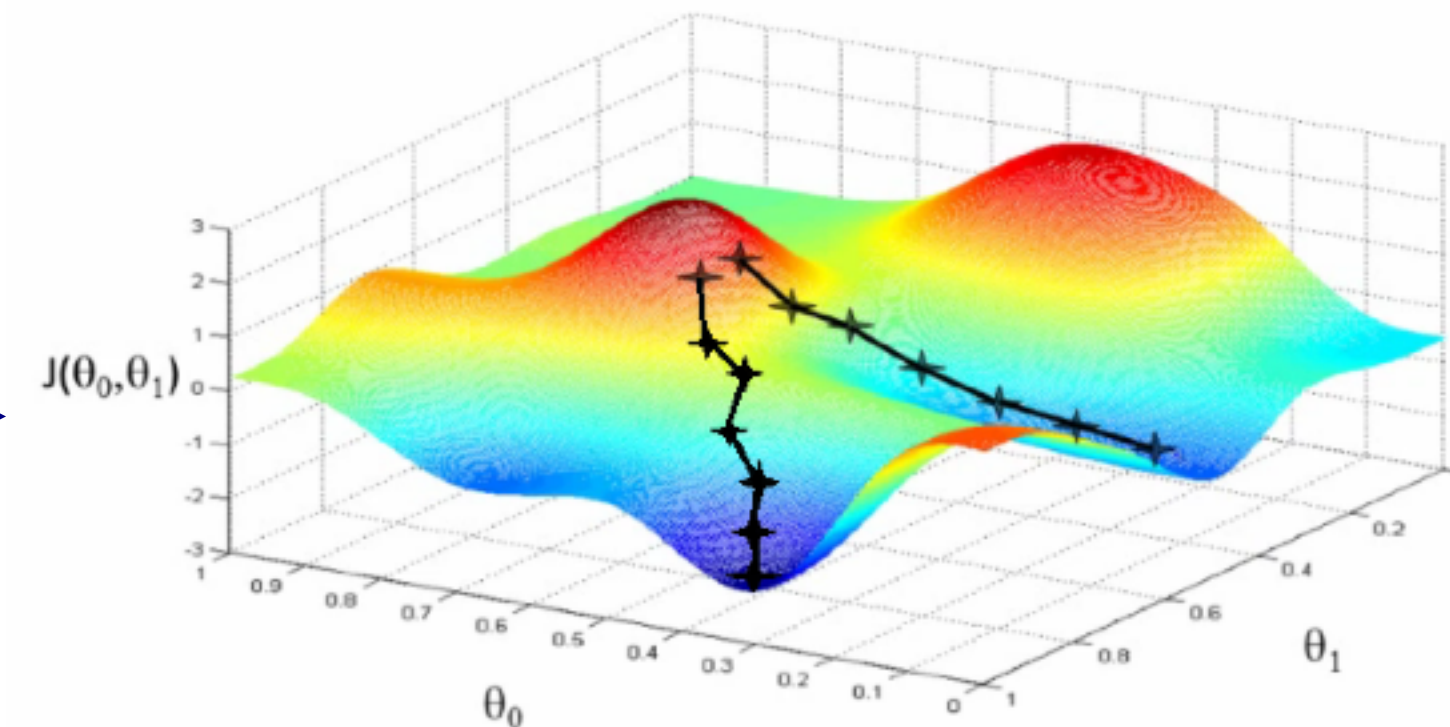# Stochastic Gradient Descent (SGD)

Dataset D

Trained Model

Repeat ..

| Select a random subset (batch) from D | → | Compute gradients $\nabla L(\theta)$ on the batch | → | Update $\theta' = \theta - \eta * \nabla L(\theta)$ |

.. until model convergence (i.e., many epochs)

Differential Privacy (Weaker notion)

X: The data universe.

$D \subset X$: The dataset (one element per person)