











"The constant value that minimizes MSE has a slope (derivative) of zero and is equal to the mean of the actual values."

And that's exactly why **gradient boosting starts from the mean** — it’s the best possible constant prediction to build on.

**Advantages of Gradient Boosting:**

* High accuracy, especially for complex datasets.
* Can be used for regression and classification tasks.
* Can handle non-linear relationships well due to decision trees being the weak learners.

**When to Use Gradient Boosting:**

* When you have structured data (tabular) with complex relationships.
* When you're looking for high accuracy in predictive modeling.
* It's very powerful but can overfit if not properly tuned, especially with too many trees.