Intuition of Gradient Boosting

Regression vs. Classification

Regression	**Classification**
Loss: $L = ½(y – y\_hat)^2$ Gradient (negative): **–dL/dy_hat = y – y_hat** → Each tree fits the **residual (numeric error)**. Intuition: The model corrects the *difference* between predicted and true values.	Loss: $L = –[y \log(p) + (1 – y) \log(1 – p)]$ Gradient (negative): **–dL/dF(x) = y – p** → Each tree fits the **probability residual** (difference between true label and predicted probability). Intuition: The model corrects how much the predicted probability *over- or underestimates* the truth.

Key Takeaway

Both regression and classification use the same idea:

■ Each new tree learns to fix what the previous trees got wrong.

The only difference is the space of residuals — numeric values vs. probabilities.