Gradient Descent: Intuition

Gradient descent is an optimization method that iteratively updates parameters to reduce a loss function.

The gradient points in the direction of steepest increase, so moving opposite the gradient typically decreases the loss.

Key idea: Take small steps downhill on the loss surface.

Softmax

Softmax converts a vector of scores into probabilities by exponentiating and normalizing.

Formula:

softmax(z_i) = exp(z_i) / sum_j exp(z_j)

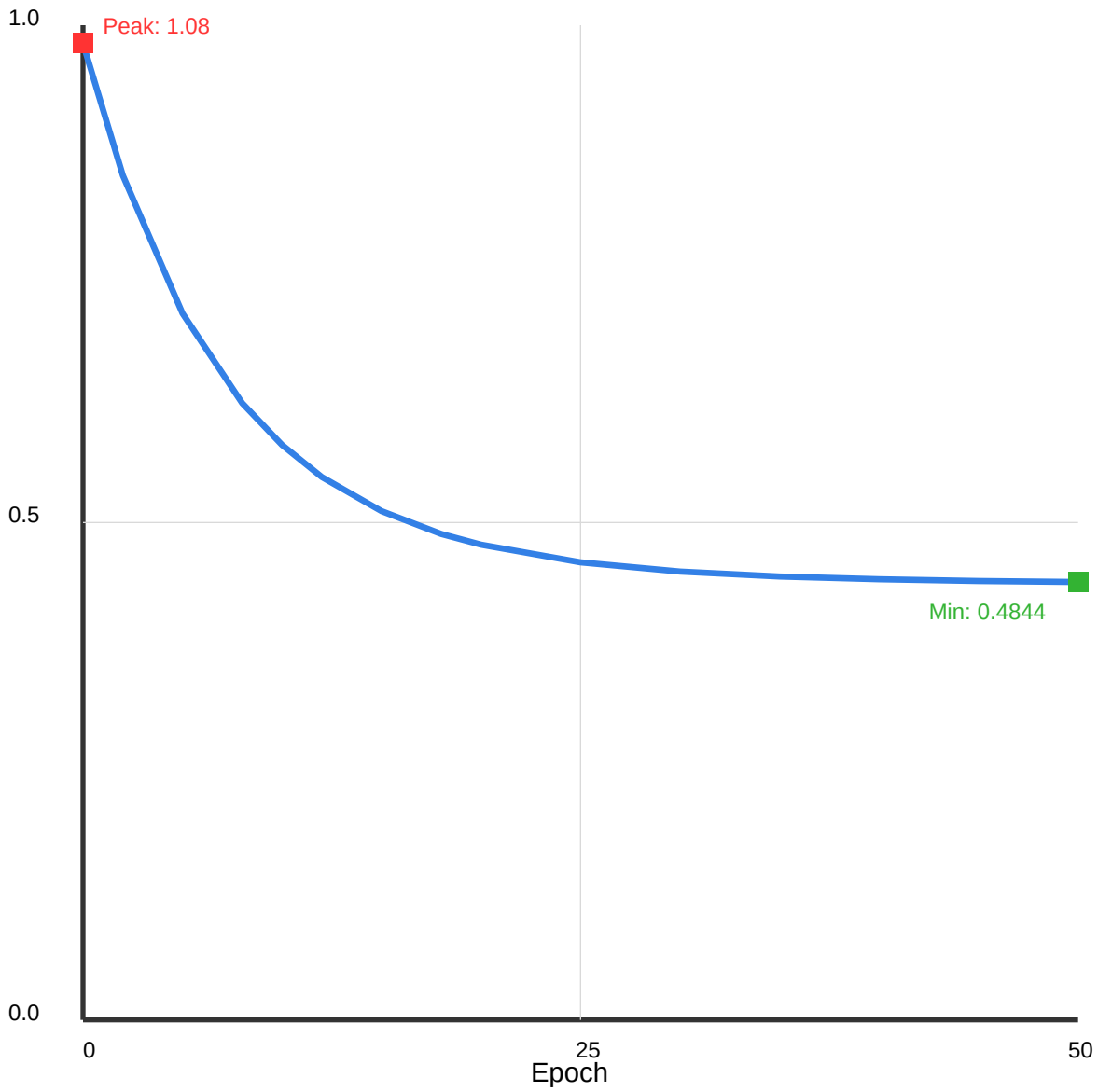
Symbols:

z_i = score (logit) for class i
exp = exponential function
sum_j = sum over all classes j

Example:

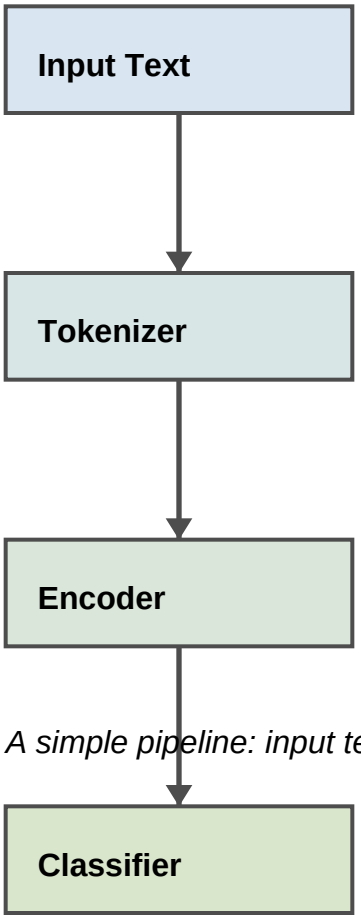
If scores are [2, 1, 0], exponentiate to [e^2, e^1, e^0]
and divide each by their sum.

Training Loss Curve



Loss decreases quickly early on and then flattens as training converges.

Model Pipeline



Raw text provided by the user.

Splits text into tokens usable by the model.

Transforms tokens into vector representations.

Outputs the predicted class probabilities.

A simple pipeline: input text is tokenized, encoded, and then classified.