input (hidden.dim)

↓ vocab size.

W2 = [100, 27] → o/p. 27 possible characters.

b2 = [27] → vocab size.

logits = h @ W2 + b2

hidden layer [32x100]

→ [32 x 27]

counts = logits.exp()

prob. = counts / counts.sum(1, keepDims = True)

↳ [32x27]

prob [torch.arange(32), Y] - probability of each of the 32 characters.

Negative Log likelyhood loss

loss = - prob [torch.arange(32), Y].log().mean()

→ F. cross_entropy (logits, Y).