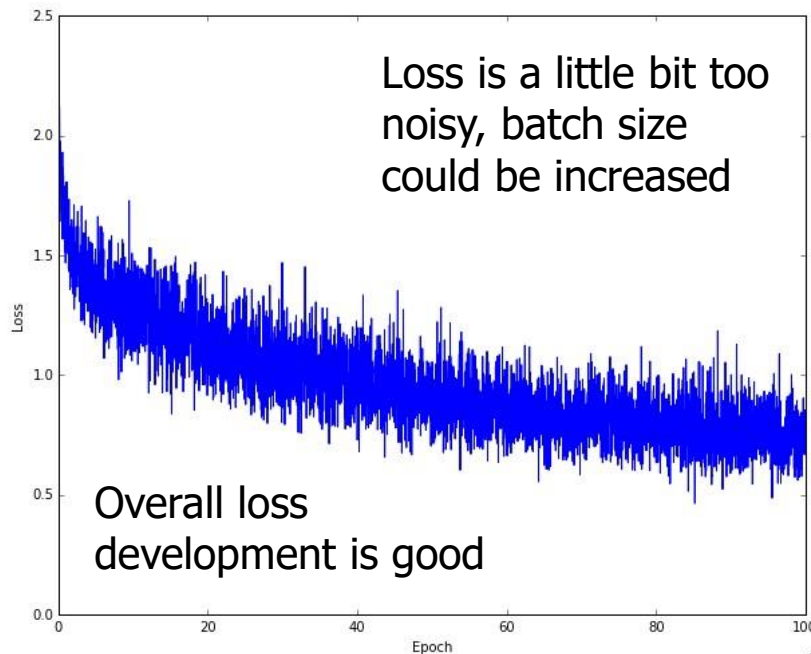## Why minibatches?

- Unbiased estimate of the gradient
- Computational effort

- Noise induced regularization for small batch sizes
  Note: This is usually not worth it

Neural Network Regularization – Batch methods

Which batch size should I go for?

- Hardware restrictions set upper limit
- Power-of-two batch sizes match physical processor and improve runtime
- Loss band should be smooth, implying even gradient estimates.

Loss is a little bit too noisy, batch size could be increased

Overall loss development is good

http://cs231n.github.io/neural-networks-3/#baby