**Lecture 7 poll**

**Slide 13**

Select all that are true

* The actual loss function we try to minimize requires batch updates
* Batch updates minimize the total loss over the entire training data
* Batch updates optimize the actual loss function
* Batch updates require processing the entire training data before we perform a single update

**Slide 50**

Select all that are true

* SGD is an online version of batch updates
* SGD can have oscillatory behavior if we do not randomize the order of the inputs
* SGD can converge faster than batch updates, but arrive at poorer optima
* SGD convergence to the global optimum can only be guaranteed if step sizes shrink across iterations, but sum to infinity in the limit

**Slide 80**

Select all that are true

* Minibatch descent is an online version of batch updates
* Minibatch descent is faster than SGD when the batch size is 1
* The variance of minibatch updates decreases with batch size
* Minibatch gradient approaches batch updates in variance, but SGD in efficiency when we use vector processing and large batches

**Slide 99**

Which of the following are true

* Vanilla SGD considers the long-term trends of gradients in update steps
* Momentum methods consider the long-term average of derivatives to make updates
* RMSprop only considers the second order moment of derivatives, but not their average trend, to make updates
* ADAM considers both the average trend and second moment of derivatives to make updates
* Trend-based optimizers like momentum, RMSprop and ADAM are important to smooth out the variance of SGD or minibatch updates