

## CS4850 HW8

- When training a deep neural network for image classification, our error function is a sum of individual errors for each image in the training set. There may be a million images in the training set. Instead of taking the derivative of a million-term summation, we can take the derivative of one term (or of the sum of a few random ones). We then use this cheaper derivative to update the weights. This method is much faster but why might it also give a better minimum?
- Select a value for  $k$  and create a set of vectors:

$$H = \{x \mid x = (x_1, x_2, \dots, x_k), x_i \in \{0, 1, \dots, k-1\}\},$$

where  $H$  is pairwise independent and  $|H| < k^k$ . A set of vectors is pairwise independent if, for any subset of two of its coordinates, all of the  $k^2$  possible pairs of values that could appear in those coordinates such as  $(0, 0), (0, 1), \dots, (1, 0), (1, 1), \dots$  occur the exact same number of times.