|  |  |
| --- | --- |
| If you have 10,000,000 examples, how would you split the train/dev/test set? | 98% train . 1% dev . 1% test |
| The dev and test set should: | Come from the same distribution |
| If your Neural Network model seems to have high variance, what of the following would be promising things to try? | - Add regularization  - Get more training data |
| You are working on an automated check-out kiosk for a supermarket, and are building a classifier for apples, bananas and oranges. Suppose your classifier obtains a training set error of 0.5%, and a dev set error of 7%. Which of the following are promising things to try to improve your classifier? (Check all that apply.) | * Increase the regularization parameter lambda * Get more training data |
| What is weight decay? | A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration. |
| What happens when you increase the regularization hyperparameter lambda? | Weights are pushed toward becoming smaller (closer to 0) |
| With the inverted dropout technique, at test time: | You do not apply dropout (do not randomly eliminate units) and do not keep the 1/keep\_prob factor in the calculations used in training |
| Increasing the parameter keep\_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply) | - Reducing the regularization effect  - Causing the neural network to end up with a lower training set error |
| Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) | * Data augmentation * L2 regularization * Dropout |
| Why do we normalize the inputs x*x*? | It makes the cost function faster to optimize |