When someone refers to a “machine learning pipeline,” he or she is referring to:

A PhotoOCR system.

A character recognition system.

A system with many stages / components, several of which may use machine learning.

An application in plumbing. (Haha.)

Suppose you are running a text detector using 20x20 image patches. You run the classifier on a 200x200 image and when using sliding window, you “step” the detector by 4 pixels each time. (For this problem assume you apply the algorithm at only one scale.) About how many times will you end up running your classifier on a single image? (Pick the closest answer.)

About 100 times.

About 400 times.

About 2,500 times.

Correct

About 40,000 times.

Suppose you are training a linear regression model with m examples by minimizing: $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2)$. Suppose you duplicate every example by making two identical copies of it. That is, where you previously had one example $(x^{(i)}, y^{(i)})$, you now have two copies of it, so you now have 2m examples. Is this likely to help?

Yes, because increasing the training set size will reduce variance.

Yes, so long as you are using a large number of features (a “low bias” learning algorithm).

No. You may end up with different parameters θ, but they are unlikely to do any better than the ones learned from the original training set.

No, and in fact you will end up with the same parameters θ as before you duplicated the data.

Correct

You’ve just joined a product group that has been developing a machine learning application for the last 12 months using 1,000 training examples. Suppose that by manually collecting and labeling examples, it takes you an average of 10 seconds to obtain one extra training example. Suppose you work 8 hours a day. How many days will it take you to get 10,000 examples? (Pick the closest answer.)

About 1 day.

About 3.5 days. Correct

About 28 days.

About 200 days.

Suppose you perform ceiling analysis on a pipelined machine learning system, and when we plug in the ground-truth labels for one of the components, the performance of the overall system improves very little. This probably means: (check all that apply)

We should dedicate significant effort to collecting more data for that component.

It is probably not worth dedicating engineering resources to improving that component of the system.

If that component is a classifier training using gradient descent, it is probably not worth running gradient descent for 10x as long to see if it converges to better classifier parameters.

Choosing more features for that component may help (reducing bias), and reducing the number of features for that component (reducing variance) is unlikely to do so.

BC
