

# Error analysis

$m_{cv} =$ ~~500~~<sup>5000</sup> examples in cross validation set.

Algorithm misclassifies ~~100~~<sup>1000</sup> of them.

Manually examine 100 examples and categorize them based on common traits.

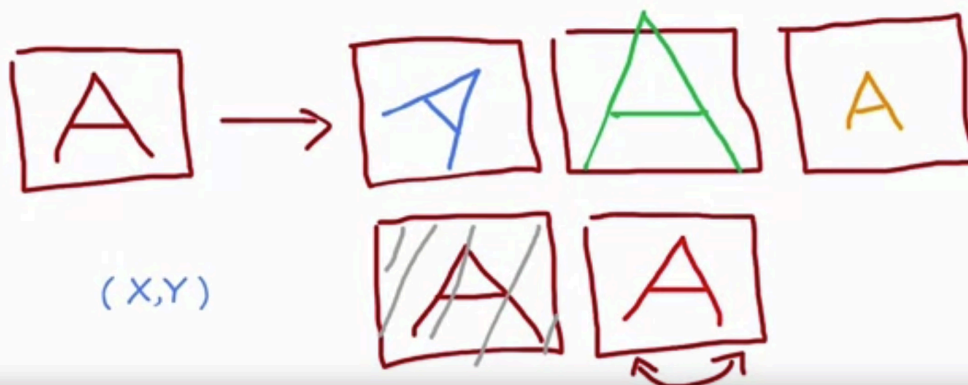
- Pharma: 21 → more data features
- Deliberate misspellings (w4tches, med1cine): 3
- Unusual email routing: 7
- Steal passwords (phishing): 18 → more data features
- Spam message in embedded image: 5

Which of these is a way to do error analysis?

- ☐ Collecting additional training data in order to help the algorithm do better.
- ☐ Calculating the training error  $J_{train}$
- ☒ Manually examine a sample of the training examples that the model misclassified in order to identify common traits and trends.
- ☐ Calculating the test error  $J_{test}$

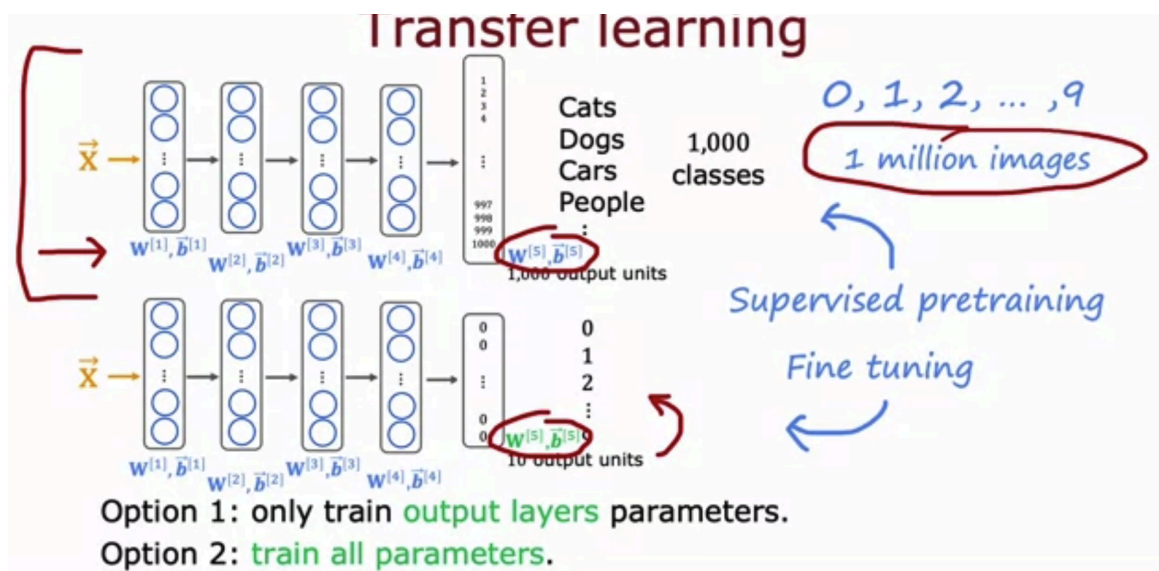
# Data augmentation

Augmentation: modifying an existing training example to create a new training example.



2. We sometimes take an existing training example and modify it (for example, by rotating an image slightly) to create a new example with the same label. What is this process called?

- ☐ Bias/variance analysis
- ☐ Machine learning diagnostic
- ☒ Data augmentation
- ☐ Error analysis



3.

What are two possible ways to perform transfer learning? Hint: two of the four choices are correct.

- ☒ You can choose to train just the output layers' parameters and leave the other parameters of the model fixed.
- ☒ Download a pre-trained model and use it for prediction without modifying or re-training it.
- ☐ Given a dataset, pre-train and then further fine tune a neural network on the same dataset.
- ☐ You can choose to train all parameters of the model, including the output layers, as well as the earlier layers.