CS230: Lecture 5
Case Study
Kian Katanforoosh
Goal: Determine which parts of a microscope image corresponds to which individual cells.

Data: Doctors have collected 100,000 images from microscopes and gave them to you. Images have been taken from three types of microscopes:

<table>
<thead>
<tr>
<th>Type</th>
<th>Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td>50,000</td>
</tr>
<tr>
<td>Type B</td>
<td>25,000</td>
</tr>
<tr>
<td>Type C</td>
<td>25,000</td>
</tr>
</tbody>
</table>

Question: The doctors who hired you would like to use your algorithm on images from microscope C. How you would split this dataset into train, dev and test sets?
**Question**: The doctors who hired you would like to use your algorithm on images from microscope C. How you would split this dataset into train, dev and test sets?

**Answer**: 
1) Split has to be roughly 90,5,5. Not 60,20,20.  
2) Distribution of dev and test set have to be the same (contain images from C).  
3) There should be C images in the training as well, more than in the test/dev set.

**Question**: Can you augment this dataset? If yes, give only 3 distinct methods you would use. If no, explain why (give only 2 reasons).

**Answer**: Many augmentation methods would work in this case:  
- cropping  
- adding random noise  
- changing contrast, blurring.
- flip  
- rotate
**First try**: You have coded your neural network (model M1) and have trained it for 1000 epochs. It doesn’t perform well.

**Transfer Learning**: One of your friends suggested to use transfer learning using another labeled dataset made of 1,000,000 microscope images for skin disease classification (very similar images).

A model (M2) has been trained on this dataset on a 10-class classification. Here is an example of input/output of the model M2.

**Question**: You perform transfer learning from M2 to M1, what are the new hyperparameters that you’ll have to tune?
**Question**: You perform transfer learning from M2 to M1, what are the new hyperparameters that you’ll have to tune?

\[ l = \text{number of layers transferred from M2} \]
\[ l_a = \text{number of new layers added to the new model’s head} \]
\[ l_f = \text{number of frozen layers} \]
**Question:** How can you correct your model and/or dataset to satisfy the doctors’ request?

**Answer:** Modify the dataset in order to label the boundaries between cells. On top of that, change the loss function to give more weight to boundaries or penalize false positives.

\[
L_{binary} = - \sum_{i=1}^{n_y} (3y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i))
\]

\[
L_{multi-class} = - \sum_{i=1}^{n_y} w_{y_i} y_i \cdot \log(\hat{y}_i)
\]

\[
w_{y_i} = \begin{cases} 
1 & \text{if } y_i = (1,0,0) \\
5 & \text{if } y_i = (0,1,0) \\
3 & \text{if } y_i = (0,0,1)
\end{cases}
\]
- C4M1 and C4M2 for next Tuesday 10/20 at 8:30 AM PDT
- TA Section on Friday 10/16
- Midterm Review Session: Tuesday, 10/20 12pm-1:20pm PDT via Zoom
- We will upload a number of practice midterms from past quarters to Piazza.