

A4

Questions worth 20 points each

1. Write pseudocode to generate a mixup sample of audio. Assume the following function signature:

```
def gen_mixup_sample(main: np.ndarray, mixin: np.ndarray,
                    weight:float) -> np.ndarray:
    """
    Generate a mixup example from a pair of audio examples
    sampled at the same Fs. Examples need not be the same
    length, and the shorter one is randomly placed within
    the longer one.
    The mixin example is attenuated by a factor of weight.
    :param main: samples from main example
    :param mixin: samples, example to be attenuated & added
    :param weight: scale factor between (0, 1]
    :return: example combining main and attenuated mixin
    """
```

Suppose that we call this for an example of class0 and class3 as follows:
`gen_mixup_sample(c0_example, c3_example, .2)`. What would be a logical training target for the example if there are 4 classes with one-hot encoding?

NOTE: We use a scalar for simplicity here, a more meaningful mixup might modify the mixin signal to have a specific signal-to-noise ratio.

2. Suppose we have 7 classifiers that are bootstrap aggregated. Under what circumstances would we expect the bootstrap aggregation to perform better than the individual classifiers and why?
3. Explain in your own words how L^p penalties move the result of optimization away from the parameters that minimize the loss function. Why is this desirable?
4. Provide an example of a situation where a gradient is very large, possibly producing large and potentially undesirable changes to the model parameters. What can be done to prevent very large changes to the model?

5. Suppose the first two gradients of an optimization produce the following mean gradient vectors with respect to the loss: $[.1 \ .3]^T$ and $[.15 \ -.1]^T$. Assuming $\rho = .7$, $\epsilon = .001$, compute the offsets that will be applied to the model using RMSprop for these two gradients.
6. Record yourself saying the word “eat” [It]. Use an audio viewer (e.g., wavesurfer, Adobe Audition, Audacity) to zoom in to the vowel portion. Create a screen shot that has a little over 3 cycles of your vowel. Show the screen shot and mark the start and end of the three cycles as well as its duration. Compute F0 from the duration of the three cycles. (Using multiple cycles reduces the measurement error.)
7. For the same recording, create a narrowband spectrum¹ of a section of the speech (in WaveSurfer, highlight the signal and right-click spectrum section). Find the peaks in the spectrum that correspond to F1 and F2. Take a screenshot and annotate the first two formants, noting their frequency. Then plot the peaks against the male or female data from the Peterson and Barney (1952) data shown on the following pages. You can print out or copy the appropriate figure below and hand annotate your vowel at the appropriate spot. Do you fall within the range of the speakers that Peterson and Barney measured?
8. Record yourself saying “boing” /bɔɪŋ/
 - a. Use wavesurfer’s speech analysis configuration to estimate formants and F0. Label the phonemes /b/, /ɔɪ/, /ŋ/, by drawing on a screenshot. Use both the distribution of energy in the spectrogram and listening to determine the boundaries.
 - b. What type of phoneme is /ɔɪ/? What characterizes this type of phoneme and how do you see it in the spectrogram?
 - c. How would one characterize the phoneme /b/? Apart from it being a consonant, what are the three things that make this a /b/ versus some other consonant?

REFERENCES

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* **24**, 175-184.

¹ We have not yet covered short and wide-band spectra and spectrograms. A narrow-band spectrum or spectrogram is one taken from a longer sample, usually at least 20 ms. Because the spacing between frequency samples is inversely related to the sample length, long-time durations result in good frequency resolution, or narrow spacing between the frequency bands; hence the term narrow-band spectrogram.



