Task 4: Discovering Future Solar Energy Materials

Pretraining feature representations

In lecture, you have learned about training neural network autoencoders. An autoencoder consists of an encoder f and a decoder g, which are jointly optimized to model the identity function:

$$\mathbf{x} \approx g\Big(f(\mathbf{x};\theta);\phi\Big) \tag{1}$$

Because the vector $\mathbf{z} = f(\mathbf{x};\theta)$ usually has a lower dimensionality than $\mathbf{x}$, the autoencoder learns an efficient lower dimensional *feature representation* for $\mathbf{x}$. This is an *unsupervised* technique because it allows learning from unlabelled data samples $\{\mathbf{x}_n\}$, e.g., images without their associated ground truth class label. The feature representation $f(\mathbf{x};\theta)$ may prove useful in related tasks, where we work with the same inputs $\mathbf{x}$ but need to predict a quantity y about $\mathbf{x}$.

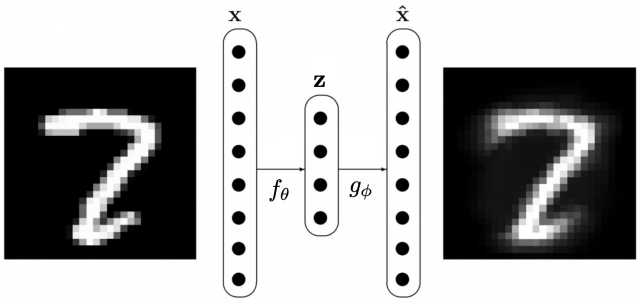

Figure 1: Unsupervised training of a feature representation $f(\mathbf{x};\theta)$

In this task, we not only have access to an unlabelled data set $\{\mathbf{x}_n\}$ that we can leverage, for instance, using an autoencoder. Instead, we have *labeled* data $\{(\mathbf{x}_n, y_n)\}$, and while the quantity y is different from our quantity of interest y^*, it is highly related. Thus, a good feature representation for predicting y is likely also useful for predicting y^*, and more so than using only unlabelled data $\{\mathbf{x}_n\}$. For example, we can train an encoder f and a final predictor h to approximate y:

$$y \approx h\Big(f(\mathbf{x};\theta);\phi\Big) \tag{2}$$

As with the autoencoder, we can use the *pretrained* feature extractor $f(\mathbf{x};\theta)$ in a related task.

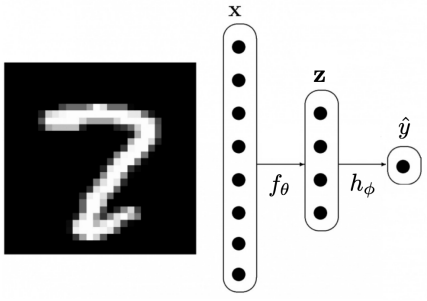

Figure 2: Supervised training of a feature representation $f(\mathbf{x};\theta)$