UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering
ECE 417 Multimedia Signal Processing

Lecture 22 Sample Problem Solutions

Problem 22.1

1. There's a range of acceptable answers. For example, $\vec{w} = [1, 1, \ldots, 1]^T$ and $b = -150$ solves the problem with zero error.

2. We need an $\vec{x}$ with the smallest possible pixel values, but such that $\vec{w}^T\vec{x} > 150$. The vector $\vec{x} = \left(\frac{1-b}{100}\right)\vec{w}$ works: the classifier considers this image to be part of $\hat{y} = 1$. A human observer would consider this to be more like class 0, because each pixel is relatively dark (just 1.51 intensity).

Problem 22.2

There are many possible solutions. One solution would create a new weight vector, W, and a new output, $\hat{v} = W\vec{y}$. Then we could train the matrices U, V, and W as follows:

1. Using the clean data samples $\vec{s}_i$, train U and V to minimize the cross-entropy between the network output $\vec{z}_i$ and the targets $\vec{\zeta}_i$:

$$E_{\text{primary}} = -\frac{1}{n}\sum_{i=1}^{n} H(\vec{\zeta}_i \| \vec{z}_i)$$
$$U \leftarrow U - \eta\frac{\partial}{\partial U}E_{\text{primary}}$$
$$V \leftarrow V - \eta\frac{\partial}{\partial V}E_{\text{primary}}$$

2. Generate a lot of noisy samples, $\vec{x}_i = \vec{s}_i + \vec{v}_i$, by randomly generating noise vectors and adding them to the clean training samples. Then train W to minimize

$$E_{\text{adversary}} = \frac{1}{2n}\sum_{i=1}^{n} \|\vec{v}_i - \hat{v}_i\|^2$$
$$W \leftarrow W - \eta\frac{\partial}{\partial W}E_{\text{adversary}}$$

3. Once U, V, and W have been pre-trained as described above, then you can re-train them simultaneously. U is trained to minimize $E_{\text{primary}} - E_{\text{adversary}}$, V to minimize E_{primary}, and W is trained to minimize $E_{\text{adversary}}$:

$$U \leftarrow U - \eta\frac{\partial}{\partial U}\left(E_{\text{primary}} - E_{\text{adversary}}\right)$$
$$V \leftarrow V - \eta\frac{\partial}{\partial V}E_{\text{primary}}$$
$$W \leftarrow W - \eta\frac{\partial}{\partial W}E_{\text{adversary}}$$