EE621 ADAPTATION AND LEARNING **Instructor: Ali H. Sayed**
Homework Assignment #4 (Neural Networks) **Due: May 20, 2019**

---

1) How would the listing of the stochastic-gradient backpropagation algorithm change if ℓ_1-regularization is used, i.e., if we employ the following regularized empirical risk:

$$R_{\rm emp}(W,\theta) \triangleq \sum_{\ell=1}^{L-1} \alpha\|{\rm vec}(W_\ell)\|_1 + \frac{1}{N}\sum_{n=0}^{N-1} \|\gamma_n - \widehat{\gamma}_n\|^2$$

where $\alpha > 0$ and the notation ${\rm vec}(A)$ refers to replacing a matrix A by a vector constructed by stacking its columns on top of each other.

2) Consider a feedforward neural network with a single output node; its output signals are denoted by $\{z(n), \widehat{\gamma}(n)\}$. Set all activation functions to the hyperbolic tangent function and consider the regularized logistic risk:

$$R_{\rm emp}(W,\theta) \triangleq \sum_{\ell=1}^{L-1} \rho\|W_\ell\|_{\rm F}^2 + \frac{1}{N}\sum_{n=0}^{N-1} \ln\left(1+e^{-\gamma(n)\widehat{\gamma}(n)}\right)$$

where $\gamma(n) = \pm 1$. Repeat the derivation that led to the backpropagation algorithm for this risk function and determine the necessary adjustments to the listing of the algorithm.

3) Let $x = {\rm col}\{h, z\}$ with binary entries $\{0, 1\}$, and define $\theta = {\rm col}\{\theta_b, \theta_h\}$. Consider a quadratic energy function of the form

$$E(h,z) \triangleq \theta^{\sf T} x - x^{\sf T}\left[\begin{array}{cc} A & \frac{1}{2}W^{\sf T} \\ \frac{1}{2}W & B \end{array}\right] x$$

where A and B are symmetric matrices with zero entries on their diagonals. Let $x(k)$ denote the $k-$th entry of x and let x_{-k} denote the column vector that excludes $x(k)$. Determine an expression for the conditional probability $\mathbb{P}(\boldsymbol{x}(k) = 1 | \boldsymbol{x}_{-k} = x_{-k})$.

4) Consider an input image of size $L \times L$ and a mask of size $K \times K$. Assume the mask strides horizontally by s_h steps and vertically by s_v steps. The size of the filtered image is denoted by $L' \times L'$. Find an expression for L' in terms of (L, K, s_h, s_v).

5) Consider a convolutional neural network consisting of C pairs of correlation and pooling layers. Assume we insert between each pair of correlation and pooling layers a new rectification layer, whose operation is as follows. For every map d in convolutional stage c, it replaces each entry in the feature vector $\mathcal{F}_n^{(c,d)}$ by its absolute value $|\mathcal{F}_n^{(c,d)}|$. This rectified map is then fed into the pooling layer. How should the listing of the stochastic-gradient backpropagation algorithm be adjusted in this case?

Computer project. Select a dataset of your choice and train a feedforward neural network with and without dropout. Use three strategies for initializing the weight vectors: random initial conditions, auto-encoding, and contrastive divergence. Compare the results.